République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahleb de Blida

Mémoire de fin d'études

Pour l'obtention du diplôme de master en informatique

Option : Sécurité des Systèmes d'Information

Thème

Mise en place d'une solution SIEM pour la Supervision d'un système d'Information

Organisme d'accueil : AL SALAM BANK

Réalisé par :

OUCHENE Hiba

Devant le Jury compose de :

Présidente Mme. N.BOUSTIA

Promotrice Mme. M.ARKAM

Examinateur M. A.KAMECHE

Année universitaire **2019/2020**

Résumé

La supervision de la sécurité des systèmes d'information présente un grand défi pour les grandes entreprises, elle est considérée comme un pilier essentiel pour assurer la sécurité de tous les composants du système d'information afin de détecter les failles et les violations de sécurité et y remédier à eux.

Parmi les parties analysées dans ces entreprises, on trouve ce qu'on appelle les fichiers Journaux, ces derniers sont construits et enregistrés avec un format spécial par de chaque équipement, système ou composant du SI en général. Ces fichiers aident à suivre les activités des utilisateurs dans le SI, ce qui permet de détecter et suivre les comportements suspects qui viole la politique de sécurité d'une façon ou d'une autre.

Le suivi et l'analyse de ces évènements est fait dans la plupart du temps par une équipe de sécurité. Cette tâche est une tâche difficile, du fait que le suivi des évènements doit être fait d'une manière enchainée sans négliger ou oublier aucun d'eux, surtout si cette surveillance concerne une activité critique.

Dans ce projet, on va proposer une solution pour l'analyse des enregistrements de sécurité du SI du « ALSSALEM Bank », et ceci en les récupérant à partir du serveur du stockage, les enrichir en utilisant différentes méthodes, cela sera suivi par leur indexation et stockage, afin qu'ils soient analysés par le système et contrôlés d'une manière simplifiée par l'équipe de sécurité.

Cette solution va permettre une analyse dynamique de ces événements selon des règles construites et stockées par cette équipe, ce qui va aider majoritairement dans la facilitation de la supervision de la sécurité du système en offrant la possibilité de la recherche rapide et la filtration de ces évènements. A la fin de chaque analyse, un ensemble des alertes peut être ajouté et affiché dans le tableau de bord de ce système.

La solution est mise en œuvre en la présentant à travers une interface d'une application web qui respecte les règles d'interface Homme-Machine (IHM) et les critères de sécurité des application web 2.0.

Mots clés : SIEM, fichier log, événement, supervision, SI, détection, analyse d'événements, indexation, alerte, tableau de bord, recherche et filtrage.

ملخص

من بين التحديات التي تواجه المؤسسات التي تمتلك نظاما معلوماتيا: الرقابة الأمنية، والتي تعتبر ركيزة أساسية من اجل ضمان سلامة وأمان كافة أجزاء النظام المعلوماتي وذلك من اجل الكشف عن الاختراقات والتجاوزات الأمنية والتصدي لها.

أحد الأجزاء التي يتم مراقبتها داخل هذه المؤسسات نجد ما يعرف بملف السجلات الذي هو عبارة عن ملف يتم تكوينه وتخزينه بطريقة خاصة من طرف كل جزء من نظام المعلومات كل على حدة. هذه الملفات تساعد على معرفة النشاطات الذي يقوم بها المستخدم داخل النظام المعلوماتي مما يوفر امكانية متابعتها واستخراج المشبوهة منها والتي تهدد أمنه بأي طريقة كانت.

غالبا ما تتم متابعة ومراقبة هذه الأحداث من طرف فريق أمني، لكن ذلك يعتبر أمرا صعبا لما فيه من متابعة متسلسلة لهذه الأحداث والحرص على الالمام بها دون الغفلة عن أي منها خاصة إذا كانت هذه المراقبة تخص مؤسسة كبيرة.

في هذه المشروع، سوف نقوم بتقديم حل من أجل مراقبة السجلات الأمنية للنظام المعلوماتي لبنك السلام وذلك عن طريق استرجاعها من خادم التخزين واثرائها بعدة طرق، ليتم بعد ذلك فهرستها وتخزينها من أجل ان يتم تحليلها ومراقبتها بشكل مبسط من طرف الفريق الأمني لهذه المؤسسة. هذا الحل سيوفر تحليلا ديناميكيا لهذه الأحداث حسب قواعد يتم تكوينها وتخزينها من طرف هذا الفريق والتي ستساعد وبشكل كبير في تسهيل المراقبة الأمنية للنظام. كما سنوفر إمكانية البحث السريع والتصفية لهذه الأحداث. ينتهي هذا التحليل عادة بتنبيهات معروضة على مستوى وحدة القيادة.

سيتم كخطوة نهائية عرض الحل عن طريق واجهة تطبيق ويب سهل الاستعمال وبتصميم يحترم قواعد انشاء واجهة الإنسان- آلة ويحترم معايير أمن تطبيقات الويب Web 2.0.

الكلمات المفتاحية: نظم إدارة المعلومات الأمنية، سجلات الأحداث، حدث أمني، مراقبة، امن، نظام معلومات، كشف، تحليل الحدث، فهرسة، تنبيه، لوحة القيادة، بحث وتصفية.

Abstract

The supervision of information systems security presents a big challenge for large companies, it is considered as an essential pillar to ensure the security of all components of the information system in order to detect security breaches and remedy to them.

Among the parts controlled in these companies, we can find what call Log files, these are built and saved with a special format by each equipment, system or component of the IS in general....

These files help the tracking of user's activities in the Information System, which can be useful to detect and investigate suspicious behavior that violates the security policy in a way or another.

The monitoring and the analyzing of these events is mostly done by a security team. This task is a difficult task, as the monitoring of events must be done in a sequential manner without neglecting or forgetting any event, especially if this monitoring concerns a critical activity.

In this project, we will propose a solution for the analysis of security logs of the IS of the "ALSSALEM Bank", and this is done by retrieving them from the storage server, enriching them using different methods, indexing them and storing them, so that they can be analyzed by the system and checked in a simplified interface by the security team.

This solution will allow a dynamic analysis of these events based on rules that are built and stored by the security team. This will mainly help in facilitating the process of supervision of the security of the system by offering the possibility of the rapid search and filtration on these events. At the end of each analysis, a set of alerts can be added and displayed on the dashboard of this system.

As conclusion step, the solution will be presented through an interface of a web application that respects the rules of Human-Machine interface (HMI) and the security criteria of web applications 2.0.

Keyword: SIEM, Log file, event, supervision, SI, detection, event analysis, indexing, alert, dashboard, search and filtering.

Remerciements

Je remercie tout d'abord Dieu le tout puissant de m'avoir donné la force, le courage, la volonté et la santé pour pouvoir accomplir ce travail.

Je tiens à exprimer toute ma reconnaissance à ma promotrice madame « ARKAM Meriem » pour sa grande participation, sa disponibilité et pour avoir accepté de diriger ce travail. J'aimerais aussi la remercier pour ses soutiens scientifiques et moraux et ses précieux conseils qui m'ont permis de mener à bien ce travail.

Je remercie mes encadreurs « BOUROUIS Adel » et « SLIMANI Nour Elhouda Khadidja », pour la confiance qu'ils m'ont accordé en proposant ce travail, leurs encadrements et le temps qu'ils m'ont consacré durant la réalisation de ce projet.

Je remercie également les membres du jury pour m'avoir fait l'honneur d'accepter d'examiner mon travail.

Je souhaite formuler mes remerciements les plus affectueux à ma famille et surtout mes très chers parents et frères, qui ont toujours été là pour moi, leur confiance, leur amour me porte et me guide tous les jours. Sans oublier mes amies qui m'ont toujours encouragée au cours de la réalisation de mon travail.

Enfin, j'adresse mes sincères remerciements et ma gratitude la plus profonde à tous ceux qui m'ont apporté leurs aides et qui ont contribué à l'élaboration de ce mémoire. Merci à tous et à toutes.

DÉDICACES

Je dédie ce modeste travail et ma profonde gratitude à:

ma mère et mon père; pour l'éducation qu'ils m'ont prodigué avec tous les moyens et au prix de toutes les sacrifices qu'ils ont consentis à mon égard, pour le sens du devoir qu'ils mon enseigné depuis mon enfance.

A mes chères frères,

A mes chères amies,

A mes collègues

A toute personne qui occupe une place dans mon coeur et a toutes les personnes qui m'ont aidé à achever ce niveau...

Contents

List of Tables

List of Figures

Liste des abréviations

AWS: Amazon Web Services.
BPMN: Business Process Model and Notation.
CLF: COMMON Log Format.
CSS: Cascading Style Sheet.
CSV: Comma Separated Values.
DNS : Domain Name System.
ELF: Extended Log Format.
ELK: Elasticsearch Logstash Kibana.
ESM: Enterprise Security Manager.
HTML: Hyper Text Markup Language.
IDS: Intelligence artificielle.
IDS: Intrusion Detection System.
IPS: Intrusion Prevention System.
IT: Information technology.
JSON: JavaScript Object Notation.
MDA: Model Driven Architecture.
MVC: Modèle-Vue-Contrôleur.
NCSA: National Center for Supercomputing Applications.
OSSIM : Open Source Security Information and Event Management.
RBAC: Role Based Access Control.
SFTP: Secure File Transfer Protocol.
SIEM: Security Information and Event Management.
SEM: Security Event Management.
SIM: Security Information Management.
SI: Système d'information.
SoaML: Service Oriented Architecture Modeling Language.
TCAC: Taux de Croissance Annuel Composé.
TOGAF: The Open Group Achitecture Framework.
UML: Unified Modeling Language (Langage de modélisation unifié).
W3C: World Wide Web Consortium.
XMI: XML Metadata Interchange.
XML: eXtensible Markup Language.

Glossaire

Terme	Définition
Apprentissage automatique	l'apprentissage automatique est un champ de l'intelligence artificielle. Il permet à la machine grâce à l'utilisation massive de données et d'algorithmes d'apprentissage, d'analyser, résoudre des problèmes par elle-même et mettre en œuvre les solutions sans être explicitement programmés pour chacune.
Intelligence artificielle (IA)	Est un ensemble de techniques permettant à un ordinateur d'imiter une forme d'intelligence réelle. En d'autres termes, il permettre à une machine de traiter des tâches normalement associées à l'être humain afin de la rendre intelligente.
Big Data	Le terme Big Data signifie mégadonnées, grosses données ou encore données massives. Il décrit un ensemble de très gros volumes de données qu'aucun outil classique de gestion de base de données ou de gestion de l'information ne peut vraiment les traiter. Autrement dit, il s'agit d'un concept permettant de capter, traiter, rechercher, partager, stocker et analyser une grande quantité de données de manière très rapide.
Conformité	signifie la conformité à la réglementation; Est la mise en œuvre de spécifique principes selon lesquels l'entreprise se conforme aux lois et réglementations applicables dans l'environnement des affaires.
Active Directory	Active Directory (AD) est un service d'annuaire pour les environnements Windows Server. Il s'agit d'une base de données distribuée et hiérarchisée qui partage des informations d'infrastructure permettant de localiser, sécuriser, gérer et organiser les ressources informatiques et réseau, notamment les fichiers, utilisateurs, groupes, périphériques et appareils réseau.
API-REST	Une API compatible REST(REpresentational State Transfer), ou « RESTful », est une interface de programmation d'application qui fait appel à des requêtes HTTP pour obtenir (GET), placer (PUT), publier (POST) et supprimer (DELETE) des données.
MVC	MVC (Modèle/Vue/Contrôleur) est un modèle de conception pour organiser une interface graphique d'un programme, très répandu pour réaliser des applications web. Il est utilisé pour séparer l'interface utilisateur (vue), les données (modèle) et la logique d'application (contrôleur), afin de mettre à l'échelle la complexité de l'application et simplifier le travail.
RBAC	RBAC (Role-Based Access Control) est un modèle de contrôle d'accès dans lequel chaque droit d'accès est basé sur le rôle auquel l'utilisateur est associé dans une organisation. Un rôle associe à un sujet des autorisations d'accès sur un ensemble d'objets.
SFTP	SFTP pour « Secure File Transfer Protocol » aussi connu sous le nom « SSH File Transfer Protocol », est un protocole réseau de transfert de fichiers sécurisé, fiable, avec une configuration plus facile. Il fonctionne sur le protocole SSH. Il protège l'intégrité des données à l'aide de fonctions cryptographique, et identifie automatiquement le serveur et l'utilisateur.

Terme	Définition
DNS (Domain Name System)	Est un service qui permet de traduire une adresse humainement compréhensible (le nom de domaine), en une adresse IP au format numérique compréhensible par l'ordinateur, de manière unique. Autrement dit, le serveur DNS agit comme un annuaire que consulte un ordinateur au moment d'accéder à un autre ordinateur via un réseau. Il a été mis en place pour faciliter la recherche d'un site donné sur internet, en tapant seulement son adresse sans avoir besoin de connaître son adresse IP exacte.
Cloud	Un ensemble de serveurs situés à distance et accessibles à partir de n'importe quel appareil et à n'importe quel moment via une connexion internet sécurisée et protégée. Ces serveurs contiennent des données de stockage ou des logiciels qui sont habituellement stockés sur l'ordinateur d'un utilisateur, voire sur des serveurs installés en réseau local au sein d'une entreprise. Il représente la nouvelle tendance pour stocker les données.
Systèmes d'information	Sont des combinaisons de matériel, de logiciels et de réseaux de télécommunications que les gens construisent et utilisent pour collecter, créer et distribuer des données utiles, généralement dans des contextes organisationnels.
Web 2.0	Est un désignant les sites web moderne sur lesquels il y a une interface à partir de laquelle les membres peuvent interagir entre eux et participer à la création ou la modification du contenu(textes, photos, vidéos ...).

Introduction Générale

Contexte de travail

Avec le développement rapide des différentes technologies et des réseaux de communication, la sécurité informatique est devenue un enjeu majeur pour les entreprises. Étant donné que le système d'information a une importance capitale, la protection de ce dernier est une fonction nécessaire. En général, la sécurité des systèmes d'information consiste à protéger les ressources d'un organisme pour qu'ils soient utilisées dans le cadre prévu. Mais aujourd'hui avec les problèmes qui apparaissent, les utilisateurs malveillants peuvent accéder à ces ressources, et lancer des attaques de divers sources. Afin d'assurer la sécurité de ces systèmes et réduire ces problèmes, il est nécessaire pour une entreprise de développer des outils pour la surveillance proactive. Le SIEM pour Security Information Event Management est un excellent outil de surveillance qui permet de répondre aux problèmes liés à la sécurité des systèmes d'information.

Problématique

Une solution SIEM complète comprend la possibilité de collecter des informations à partir de diverses sources de données, de conserver les informations pendant une longue période,d'enrichir les différents événements, de corréler entre ces événements, de créer des règles de corrélation ou des alertes, d'analyser les données et les surveiller à l'aide de la visualisation et des tableaux de bord.
L'un des problèmes liés à la mise en œuvre d'une solution SIEM dans les entreprises est de choisir le bon SIEM en termes de prix, de fonctionnement et de compatibilité avec les formats d'enregistrement de l'entreprise.
La plupart des SIEM open source sont limités et sont beaucoup moins performants que les solutions SIEM complètes. De plus, ils nécessitent beaucoup d'expérience et et un correcte déploiement.
Les solutions SIEM commerciales nécessitent un coût de déploiement et de maintenance élevé. En plus du coût élevé, leur code est closed-source, il ne peut donc pas être compris ou modifié.

Objectifs du travail

L'objectif de ce travail est de mettre en place un système qui aide à analyser les différents événements réalisés par les utilisateurs d'un système particulier et éventuellement notifier à l'administrateur toute

action judiciaire prédéterminée par l'administrateur parmi une liste de tous les événements pouvant survenir dans le système.

Le système doit permettre :

► Récupérer un fichier log à partir d'un partage réseau ou d'un dossier prédéfini(Path défini).

► Stocker et indexer les informations recueillies.

► Effectuer des recherches rapides et spécialisées sur les informations stockées.

► Analyser les événements en suivant des règles définies par l'utilisateur, ces règles permettent de distinguer les actions anormales qui sont effectuées.

Organisation du mémoire

Pour mener bien notre travail , le mémoire est divisé en trois chapitres comme suite:

- **Chapitre 1 :**C'est une introduction aux notions de log et de SIEM. On décrira les différentes produits SIEM open source et commerciaux, on va faire une petite comparaison entre les produits commerciaux et on terminera par une discussion sur les difficultés rencontrés au début de la phase de conception.
- **Chapitre 2 :**Ce chapitre va comporter les détails de conception et de l'architecture de notre système.
- **Chapitre 3 :**Dans ce chapitre, on va montrer les outils utilisés dans le développement de mon application, comment les utiliser, les étapes de développement, et on va conclure par des captures d'écran avec une discussion de quelques résultats de tests.

Et à la fin , on va terminera par une conclusion générale et perspective.

Chapter 1
Généralité

1.1 Introduction

Les premiers produits SIEM ont vu la lumière vers la fin du 20ème siècle, l'originale idée derrière les produits SIEM est d'avoir un produit de sécurité qui aide à surveiller la sécurité des organisations en utilisant un seul produit.[2]
Les matériels essentiels des produits SIEM sont les fichiers Logs, il existe plusieurs types de logs, depuis les logs d'authentifications, les logs d'accès, les logs de trafics,...
Chaque type de logs a son propre format, le rôle principal d'un SIEM et de pouvoir lire les différents types de logs de sécurité, de les normaliser, et extraire les informations pertinentes depuis eux afin de les présenter à l'utilisateur sous format de rapports, de Dashboards et d'alarmes.
Dans ce chapitre, on va parler sur logs et leurs différents formats et types en général, les missions et les caractéristiques d'un SIEM, on va citer ainsi des cas d'utilisation de lui, et on va finir par la méthodologie technique utilisée par le SIEM.

1.2 Log

Avant d'entamer notre chapitre sur les SIEMs, on doit d'abord parler sur les logs et leurs importances pour le domaine de technologie et la sécurité des systèmes d'information plus précisément. Et donc on doit savoir au début c'est quoi un log et quelle est son utilité ?

1.2.1 Généralités sur les logs

K. Kent et M. Souppaya définissent un log de cette manière : **«Un log est un enregistrement des événements survenant dans les systèmes et les réseaux d'une organisation. Les logs sont composés d'entrées de journal; chaque entrée contient des informations relatives à un événement spécifique qui s'est produit au sein d'un système ou d'un réseau. »** [3]
S. Al-Fedaghi et F. Mahdi à leur tour définissent les logs comme celui-là : **« Une séquence ordonnée d'occurrences contenant la preuve de l'exécution d'un processus par des utilisateurs, des systèmes ou d'autres entités. Diverses sources et entités du système envoient des messages concernant leurs processus (par exemple, qui, quelles opérations, heure, etc.), ces derniers sont conservés dans**

plusieurs journaux ou fichiers logs » [4]

A la base de ces définitions, on peut dire qu'un log est un ensemble d'informations enregistrées sous un format spécifique, et qui décrit des évènements qui ont passé sur un système, un service, une application ou sur le réseau. Des exemples des informations qui peuvent apparaitre dans un fichier log sont le temps de l'évènement, son adresse IP source, sa description et son un degré de criticité.

1.2.2 Format d'un log

Tout d'abord et avant parler des différends formats des logs, c'est quoi le format d'un log et pourquoi les logs doivent avoir un format spécifique?

Le format d'un log est la structure des différents champs dans chaque ligne du fichier log, et leur emplacement, ces derniers peuvent être placés dans un document CSV, JSON, XML ou simplement un texte brut.[5]

Afin que les logs soient utiles et exploitables, ils doivent suivre un format reconnu par les différentes solutions de SIEM, ces dernières possèdent des parseurs spécifiques pour chaque type de log. Dans le cas où le format de log envoyé n'est pas reconnu par la solution de SIEM utilisée, des parseurs personnalisées introduit par l'utilisateur peuvent être utilisé à cette fin.

Un exemple des plus reconnus formats de log sont le **« COMMON Log Format » CLF** du NCSA et le **« Extended Log Format » ELF** du W3C, des logs d'accès des serveurs WEB en général et Apache spécialement.

Le format d'un COMMON Log est comme suit :

"%h %l %u %t \" %r\" %>s %b"

les champs du log sont les suivants :

- **%h :** : L'adresse IP du client distant ou bien son nom DNS.
- **%l :** : Le nom du log distant, si le champ n'est pas présent il est remplacé par un " – " .
- **%u :** : Le nom de l'utilisateur distant, si le champ n'est pas présent, il est remplacé par un " – " .
- **%t :** : Le temps de l'évènement sous le format DD/Mon/YYYY:hh:mm:ss.
- **\" %r\" :** : L'URL de la requête envoyée.
- **%>s :** :Le code d'état retourné par le serveur.
- **%b :** :Le nombre de bytes envoyés.

Un exemple d'un commun log est le suivant :

```
123.456.71.150 - - [14/Feb/2020:22:31:16 -0800] "GET
/produit/test.html HTTP/1.1" 200 97
```

Figure 1.1: Exemple d'un commun log.

Dans cet exemple l'adresse IP du client est **123.456.71.150**, la date et le temps de réception de la requête est **14/Feb/2020:22:31:16**, la page demandé se trouve dans le répertoire **/produit/test.html**, le code d'état de la réponse du serveur est **200**, ce qui indique que la requête a été servie avec succès et le nombre des bytes envoyés était **97**.
Le format d'un EXTENDED Log se rassemble à celui du COMMON Log, mais en ajoutant quelques supplémentaires, comme il est montré ci-dessous :

```
"%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\""
```

Figure 1.2: Le format d'un EXTENDED Log.

Les champs ajoutés par rapport au COMMON Log sont :

- **\" % { Referrer } i \" :** L'URL de la requête sans inclure les paramètres, il est extrait depuis l'entête de la requête envoyée..
- **\" %{User-agent}i\" :** :Le navigateur utilisé par le client, ce paramètre est extrait ainsi depuis l'entête de la requête envoyé.

Un exemple d'un EXTENDED Log est le suivant :

```
123.456.71.150 - - [14/Feb/2020:22:31:16 -0800]
"GET/produit/test.html HTTP/1.1" 200 97
"https://www.exemple.com/" "Mozilla/56.08 (Win10)"
```

Figure 1.3: Un exemple d'un EXTENDED Log.

Et donc dans cet exemple le champ Referrer est **https://www.exemple.com/** et le navigateur utilisé par le client est **"Mozilla/56.08 (Win10)"**.

1.2.3 Catégorisation des logs

Il existe actuellement plusieurs manières de catégoriser les logs ; selon leur type, selon leur degré de criticité ou selon leur niveau de journalisation.

Les logs Windows par exemple sont divisés sur quatre catégories ; les logs système, les logs d'installation, les logs de sécurité et les logs d'application. Les évènements systèmes contiennent tous les évènements relatifs aux pilotes de périphériques et au système d'exploitation. Les évènements d'installation contiennent à leurs tours les logs relatives aux paramètres de configuration du système d'exploitation. Les évènements de sécurités contiennent les logs d'authentifications, les logs des règles d'audit ainsi ceux de l'accès aux ressources systèmes. Les évènements d'application contiennent les logs relatifs aux applications installées sur le système.[5]
Les logs sont catégorisés ainsi selon leur degré de criticité en commençant par DEBUG, INFO, WARN, ERROR jusqu'on arrive au niveau FATAL. Le niveau de journalisation DEBUG désigne les évènements d'information les plus utiles pour déboguer une application. Le niveau de journalisation INFO désigne les logs qui concernent la progression de fonctionnement de l'application. Le niveau de journalisation WARN désigne les situations potentiellement dangereuses. Le niveau ERROR contient les évènements des erreurs d'application mais qui se génèrent permettent à l'application de continuer à s'exécuter. Le niveau FATAL contient les évènements des erreurs fatals et qu'empêchent l'application à continuer de s'exécuter.

1.2.4 Utilité des logs

Originalement, les logs ont été créé pour un seul but ; de diagnostiquer les problèmes qui se passent sur un système ou une application. Aujourd'hui, ils sont devenus une source indispensable d'informations pour les organisations et les entreprises.
Les catégories de logs décrient dans la section précédente contient différents types d'informations, chaque type est plus utile dans une situation précise, comme la détection des attaques, l'optimisation des performances des réseaux et des systèmes, l'enregistrement d'actions des utilisateurs et l'investigation des incidents de sécurité.[3]

1.3 SIEM

Avant d'entamer notre sujet et parler des capacités et des missions du SIEM, nous devrons tout d'abord définir c'est quoi un SIEM, et que ce qu'on peut qualifier comme une bonne solution de SIEM.
M Rouse définie un SIEM comme ça « La gestion des informations et des événements de sécurité (SIEM) est une approche de la gestion de la sécurité qui combine les fonctions SIM (gestion des informations de sécurité) et SEM (gestion des événements de sécurité) en un seul système de gestion de la sécurité. Les principes sous-jacents de chaque système SIEM sont d'agréger les données pertinentes provenant de sources multiples, d'identifier les écarts par rapport à la norme et de prendre les mesures

appropriées.» [6]

Mary K. Pratt définie un SIEM à son tour comme ceci : « Un SIEM est un outil qui combine la gestion des événements de sécurité (SEM) - qui analyse les données des journaux et des événements en temps réel pour fournir une surveillance des menaces, la corrélation des événements et la réponse aux incidents - avec la gestion des informations de sécurité (SIM) qui collecte, analyse et établit des rapports sur les données des journaux. » [7]

En basant sur ces définitions, on peut dire qu'un SIEM -Security Information and Event Management- (qui se traduit en français Gestion des évènements et des informations de sécurité) est un outil qui combine entre la gestion des évènements dé sécurité et la gestion des informations de sécurité. Un SIEM doit donc avoir la capacité de gérer les évènements de sécurité, de les stocker ainsi de les analyser afin d'extraire les informations de sécurité depuis eux pour les utiliser dans la détection des menaces et des incidents de sécurité.

1.3.1 La différence entre SIM, SEM et SIEM

Un SIEM actuellement assurent les missions du SIM (Security Information Manager qui se traduit en français gestionnaire des informations de sécurité) et du SEM (Security Event Manager qui se traduit en français gestionnaire des évènements de sécurité). Les rôles du SEM et du SIM peuvent apparaitre chevaucher, mais chaque un d'eux a des caractéristiques bien distinguées:

Un SEM a la capacité d'analyser un très grand nombre des évènements et de détecter les menaces en temps réel, il responsable ainsi sur la corrélation des évènements en temps réel, et l'envoie des alarmes.

Un SIM au contraire, a comme rôle de garder la traçabilité et d'indexer et stocker un volume important des données brutes. Et donc il offre des capacités d'indexation et de recherche pour les analystes et les experts de l'investigation numérique.

Le tableau suivant résume la différence entre un SEM, un SIM et un SIEM :

	Gestion des informations de sécurité (SIM)	Gestion des événements de sécurité (SEM)	Gestion de l'information et des événements de sécurité (SIEM)
Aperçu	Collecte et analyse des données relatives à la sécurité à partir des logs.	Analyse, visualisation et réponse aux incidents en temps réel.	SIEM, comme son nom l'indique, combine les capacités SIM et SEM.
Caractéristiques	Facilité de déploiement. Capacités de gestion des journaux.	Plus complexe à déployer. Supérieur dans la surveillance en temps réel.	Plus complexe à déployer. Fonctionnalité complète.
Exemples d'outils	OSSIM ALIEN VAULT OSSIM	Sentinelle NetIQ NetIQ.	SolarWinds solarwinds

Figure 1.4: La différence entre SIEM, SIM et SEM

1.3.2 Les Missions d'un SIEM

Il existe actuellement dans le marché plusieurs solutions SIEM, et qui ont prouvé leurs valeurs, chaque solution a ces points forts et ces points faibles, cependant une solution SIEM mature doit pouvoir exécuter ces fonctionnalités basiques du SIEM ; la gestion des logs, la recherche et la création des rapports, la surveillance de sécurité en temps réal, la gestion des incidents, le renseignement sur les menaces et la surveillance de comportement des utilisateurs et des entités. [8]

a. La gestion de logs

Un SIEM doit avoir la capacité de lire différents formats de logs depuis plusieurs sources, de posséder des parsers pour les plus reconnus formats, ainsi de donner la capacité de créer des parsers personnalisés. La gestion des logs est donc une solution centralisée pour sauvegarder les logs reçus sur la même base de données et de les archiver après.

b. La recherche et la création des rapports

Un SIEM doit fournir une interface riche qui permet la recherche des évènements selon différents critères (temps de l'évènement, source de l'évènement, nom de l'hôte, adresse ip source...). Un SIEM doit ainsi la possibilité de générer des rapports périodiques ou sur demande en utilisant des filtres bien définis.

c. La surveillance en temps réel

Un SIEM doit donner la possibilité de générer des alarmes (sur l'interface, par email ou par SMS), ces derniers sont basés sur des filtres définis par l'utilisateur et ils ont comme but d'alerter l'utilisateur du

SIEM en temps réel sur des évènements critiques, comme la détection d'un malware dangereux dans le réseau, d'une attaque de type DDoS ou le bruteforce d'un compte d'utilisateur.

d. La gestion des incidents

Un SIEM doit offrir la possibilité de détection des incidents, ainsi un workflow automatisé pour gérer les incidents, ceci inclus : l'analyse des évènements, le signalement des faux positifs ou les incidents aux experts de sécurité.

e. Le renseignement sur les menaces

Une solution SIEM doit avoir la possibilité d'être intégrée avec des sources externes pour identifier les menaces internes et externes. Un exemple sur les services de renseignement sur les menaces qui peut fournir un SIEM est le contact des adresses malveillantes par une machine.

f. surveillance de comportement des entités et des utilisateurs

Un SIEM dout donner une vision totale sur les actions effectuées par les utilisateurs et les entités, d'être capable de détecter les anomalies et les violations, ces derniers sont basés sur des routines bien définies ; comme les heures de login de l'utilisateur, ces permissions et son adresse do connexion.

1.3.3 Le cycle de vie d'un log dans le SIEM

Lorsqu'un log est envoyé au SIEM, il va passer par plusieurs étapes avant qu'il prend sa forme finale, nous allons essayer de cette partie du mémoire de décrire chaque une de ces étapes :

Le Parsing :

La première que chaque log doit passer lorsqu'il est reçu par le SIEM est le parsing ; cette étape sert à lire les différents champs d'un log et les stockées dans la base de données du SIEM. Et donc on peut dire le parsing est le processus de transformer un log brute en un log format SIEM et qui prêt à être utilisé par lui.
Certains logs ont des parsers prêts à être utilisée, d'autres nécessitent un perser personnalisé qui peut être un parser Json, XML, ou simplement des expressions régulière dans le cas d'un log sous format texte.

L'agrégation :

L'agrégation des logs c'est processus de regrouper plusieurs logs dans le même log selon des critères bien définis, l'agrégation s'applique sur les logs du même type et elle a comme but de facilité les vues

du SIEM. L'agrégation peut se faire sur un ou plusieurs champs du log comme l'adresse IP sources, l'adresse IP destination, le succès ou l'échec de l'opération.
Pour illustrer l'utilité de l'agrégation on va donner l'exemple de l'échoue de la connectivité d'un utilisateur à un équipement 100 fois dans 10 minutes, le rôle d'agrégation est de regrouper ces 100 évènements en un seul évènement, et donner plus de visibilité à cet évènement en multipliant son poids et le nombre de répétition de l'évènement par 100.
L'agrégation a plusieurs avantages comme la réduction du nombre des logs dans le SIEM, l'accélération des opérations de recherche et la facilitation des taches de surveillance. Cependant, la mauvaise implémentation des règles d'agrégation peut conduire à la perte des informations importantes.

La corrélation :

La corrélation des logs est le processus de création d'un nouvel évènement à la base d'un ensemble d'évènements qui ont passé dans une période donnée de temps, ces évènements ne doivent pas être forcément du même type. Le but de corrélation est la construction des logs de sécurité du deuxième niveau en basant sur des règles bien définis.
Afin d'illustrer le fonctionnement des règles de corrélation, on va donner les exemples de la réception des plusieurs logs de trafic depuis la même adresse IP source, sur des adresses IP qui trouvent dans la même plage IP xx.xx.xx.xx/xx, sur plusieurs ports (80 : HTTP, 22 : SSH, 21 : FTP,...).
En implémentant une bonne règle de corrélation qui base sur le nombre de logs trafic depuis la même adresse IP vers différents adresses IP et plusieurs ports dans une petite période de temps, on peut créer un nouveau log de Scan massive qui a comme adresse l'adresse IP du Scanner et comme destination la plage IP xx.xx.xx.xx/xx

Les alertes et les notifications :

Le SIEM est considéré comme l'œil et l'oreille de l'entreprise sur les menaces et les risques sécurité, il doit dont avoir un système de notifications et des alertes en temps réel. L'envoie des notifications et la génération des alarmes doit se baser sur la détection des logs des évènements de sécurité dangereux, comme le cas d'un scan massive qui nous avons vu dans la partie précédente. Cela va permettre le monitoring et la réaction immédiate dans le cas d'un incident de sécurité.

1.3.4 Déclaration des incidents :

Après l'envoie des alarmes, la prochaine étape est la création des incidents, et leurs gestions. Et donc le SIEM va permettre de gérer les incidents, en incluant leurs déclarations, leurs assignations aux

analystes de sécurité, et finissants par leurs clôtures.

On peut dire que les étapes citées précédemment composent le cycle de vie d'un log dans le SIEM :

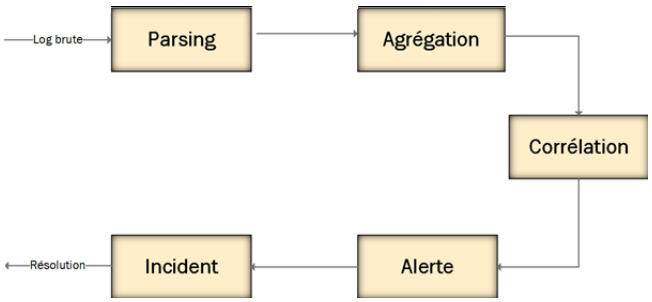


Figure 1.5: Cycle de vie d'un log dans le SIEM.

A noter qu'un log ne doit pas forcément passer par tous ces étapes, mais cette figure illustre l'ordre chronologique qu'on peut avoir depuis la réception d'un log brute, jusqu'au déclaration et résolution de l'incident.

1.3.5 Les produits SIEM Open Source existant

Les outils SIEM open source ouvrent littéralement leur conception au public. Cela permet aux de modifier et de partager plus librement le code de l'outil, offrant une personnalisation et une adaptabilité importantes. Dans cette section, une sélection des SIEMs open source est présentée.

Prelude OSS

Prelude OSS est la version open-source de Prelude SIEM. Il s'agit donc d'un SIEM open-source destiné à des environnements de tests. Il prend en charge plusieurs formats de journaux et peut s'intégrer avec d'autres outils de sécurité. Il offre également la normalisation des données d'événements dans une langue standard, ce qui peut aider à prendre en charge d'autres outils et solutions de sécurité. Prelude OSS bénéficie également d'un développement continu, ce qui lui permet de suivre les dernières informations sur les menaces.[9]

OSSIM

L'AlienVault OSSIM fourni par *AT&T* Cybersecurity est un outil SIEM open source basé sur la solution AlienVault USM. AlienVault OSSIM permet la surveillance des appareils et la collecte de journaux. Il fournit également la normalisation et la corrélation d'événements. intègre des produits open source pour fournir une plate-forme de base qui peut réaliser des fonctions de surveillance de la sécurité, collecter des journaux classifiés, identifier et résoudre les incidents de sécurité majeurs (priorité, identifier les journaux problématiques), et répondre aux exigences de surveillance de la sécurité et aux exigences d'audit et de conformité pour le stockage des journaux.

SIEMonster

SIEMonster est une solution logicielle de surveillance de la sécurité personnalisable et évolutive accessible aux petites et moyennes entreprises. L'abordabilité de SIEMonster permet de surveiller l'ensemble de réseaux un coût minimisé par rapport aux autres SIEM. SIEMonster propose désormais des options de corrélation de comportement humain pour enrichir les alertes et minimiser les faux positifs. Il fournit des informations sur les menaces en temps réel avec des flux commerciaux ou open source pour arrêter les attaques en temps réel. Grâce à l'apprentissage automatique, l'analyse du comportement basé sur l'homme montre que SIEMonster Deep Learning élimine automatiquement les attaques. SIEMonster peut être exécuter, sur site dans une VM, Bare Metal (Mac, Ubuntu, CentOS et Debian) ou l'un des fournisseurs Cloud tels qu'Amazon, GCP ou Azure.[10]

ELK

Il s'agit d'une combinaison de trois projets open source : Elasticsearch, Logstash et Kibana. La pile ELK utilise Elasticsearch pour la recherche, Logstash pour la collecte de données et Kibana pour la visualisation des données.Ces trois composants sont combinés pour former une solution complète de collecte, de traitement et d'analyse de journaux. Un nouveau projet FileBeat est ajouté à la solution, ce dernier est un outil léger de collecte et de traitement de journaux (Agent). Filebeat occupe moins de ressources et convient à la collecte de journaux sur divers serveurs et à leur transmission à Logstash. Et avec le nouveau projet (FileBeat), le nom du produit pourrait bientôt changer de ELK à l'un des noms suivants (BELK, BLEK, ELKB). [11]

Graylog

Graylog est l'une des principales solutions de gestion centralisée des journaux pour la capture, le stockage et l'analyse en temps réel. Spécialement conçue pour l'analytique de journaux moderne, Graylog élimine la complexité de l'exploration de données, des audits de conformité et de la recherche des menaces pour trouver facilement la signification des données et prendre des mesures plus rapidement. Elle est conçue pour les entreprises qui souhaitent collecter et normaliser des données de manière transparente et plus organisée à partir de n'importe quelle source de données et effectuer des analyses plus rapides à l'aide d'une solution abordable.[12]

1.3.6 Les produits SIEM commerciaux

Tout comme dans le monde open source, il existe également une abondance de produits SIEM commerciaux disponibles.Et afin de choisir les produits à présenter dans cette section, on va servir de Gartner.
Gartner est une entreprise américaine de conseil et de recherche dans le domaine des techniques avancées. Gartner mène des recherches, fournit des services de consultation, tient à jour différentes statistiques et maintient un service de nouvelles spécialisées.[13]
Gartner fournit chaque année des rapports sur différentes technologies et principes dont SIEM. La figure suivante montre le Gartner Magic Quadrant de Security Information and Event Management publier en février 2020:

Figure 1.6: Gartner Magic Quadrant de SIEM publier en février 2020 [1].

Cette édition 2020 distingue 7 « leaders » – tous d'origine américaine – ainsi désignés pour :

- leur aptitude à proposer des produits en accord avec les besoins généraux du marché.
- leur capacité à développer une base de clients et de revenus.

Selon Gartner, solutions respectives ne sont pas exemptes de défauts.

Splunk Enterprise SIEM

Est une solution SIEM offerte par l'entreprise américaine Splunk avec toutes les fonctionnalités SIEM basiques.

Cette solution fournit une surveillance des menaces améliorées telles que des tableaux de bord personnalisables, des investigations rapides, des analyses d'investigation,etc. Il peut être utilisé comme une bonne solution par les petites, moyennes et grandes entreprises. Un essai gratuit est disponible pour le produit mais avec une période d'essai diffère selon le produit(Splunk Entreprise 60 jours). Il fournit aussi une version gratuite de la plate-forme d'entreprise principale, ouverte au publique .[14]

Selon les avis des clients, c'est un outil coûteux et il est donc préférable pour les entreprises.

IBM Security QRadar

IBM Security QRadar est une plate-forme SIEM leader du marché, qui fournit une surveillance de la sécurité de l'ensemble de l'infrastructure de IT grâce à la collecte des log, à la corrélation des événements et à la détection des menaces. QRadar permet de hiérarchiser les alertes de sécurité à l'aide de bases de données de vulnérabilités et de renseignements sur les menaces, aussi à l'aide d'une solution intégrée de gestion des risques et prend en charge l'intégration avec les antivirus, IDS / IPS et les systèmes de contrôle d'accès. Elle propose une version d'essai de 14 jours et une version commercial avec un prix commence à 800$ par mois.[14]
Selon les avis des clients, il se concentre sur les incidents critiques.

Exabeam

Exabeam est le leader du marché de l'analyse du comportement de l'utilisateur et de l'entité (UEBA), Exabeam va au-delà des solutions SIEM classiques : en analysant des données importantes provenant de toutes sortes de sources, la plate-forme fait la connaissance des utilisateurs et de l'environnement et utilise l'apprentissage machine pour identifier les risques, en évaluant et en rendant visibles sur le portail les utilisateurs ayant un comportement déviant, et ce en temps réel. Les atteintes potentielles à la protection des données sont ainsi détectées et prévenues à un stade très précoce. Donc la plate-forme combine de manière unique un lac de données pour une collecte de données illimitée à un prix prévisible, un apprentissage automatique pour des analyses avancées et une réponse automatisée aux incidents dans un ensemble intégré de produits. Détecter la corrélation dans une grande quantité de données, c'est la puissance d'Exabeam.[15]

Securonix

Construit sur le Big Data, Securonix est la plateforme SIEM d'opérations de sécurité complète et de bout en bout de nouvelle génération, qui combine la gestion des journaux, l'analyse du comportement

des utilisateurs et des entités et la réponse aux incidents de sécurité. Il collecte d'énormes volumes de données en temps réel, utilise des algorithmes d'apprentissage automatique brevetés pour détecter les menaces avancées et fournit des capacités de réponse aux incidents de sécurité basées sur l'intelligence artificielle pour une correction rapide.[16]
Il sera fourni dans le cloud en tant que service. Il permet d'exporter les données visualisées dans des formats de données standard. [14]

Rapid7

C'est un outil SIEM puissant basé sur le cloud, dispose de fonctionnalités de recherche, de collecte et d'analyse de données et peut détecter un large éventail de menaces, y compris les informations d'identification volées, le phishing et les logiciels malveillants. Il peut utilise une technologie de déception avancée, une analyse du comportement des attaquants et des utilisateurs, une surveillance de l'intégrité des fichiers, une gestion centrale des journaux et d'autres fonctionnalités de découverte. Cela en fait un outil approprié pour analyser les différents points de terminaison et fournir une détection en temps réel des menaces de sécurité dans les petites, moyennes et grandes entreprises. La recherche dans les journaux, les points de terminaison et les données de comportement des utilisateurs fournissent des informations qui aident les équipes à prendre des décisions de sécurité rapides et intelligentes.[17]

LogRhythm

LogRhythm, qui est disponible sous forme de service cloud ou d'appliance sur site, dispose d'un large éventail de fonctionnalités supérieures allant de la corrélation de journaux à l'intelligence artificielle et à l'analyse comportementale. Il offre une plate-forme de renseignement de sécurité qui utilise l'intelligence artificielle pour analyser les journaux et le trafic dans les systèmes Windows et Linux. Il dispose d'options de stockage de données flexibles et constitue une bonne solution pour les flux de travail fragmentés en plus de fournir une détection des menaces segmentée, même dans des systèmes où il n'y a pas de données structurées, pas de visibilité centralisée ou d'automatisation. Adapté aux petites et moyennes entreprises, il permet de passer au crible les fenêtres ou d'autres journaux et de limiter facilement aux activités du réseau. Il est compatible avec une large gamme de types de journaux et de périphériques.[17]

RSA NetWitness

RSA NetWitness est l'une des outils SIEM les plus intermédiaires du marché. Est une plateforme qui rassemble des solutions évoluées de SIEM et de défense contre les menaces qui offrent une visibilité, des analyses et des capacités de réponse automatisées inégalées. Ces capacités combinées aident les équipes de sécurité à travailler de manière plus efficace et efficiente, en améliorant leurs compétences de recherche de menaces et en leur permettant d'enquêter et d'y répondre plus rapidement, sur toute l'infrastructure de leur organisation, que ce soit dans le cloud, sur site ou virtuelle.[18]

SolarWinds

Le SolarWinds Security Event Manager est un produit pour la gestion des informations de sécurité et des événements (SIEM).

Ce produit permet le collecte des journaux de diverses sources. Au fur et à mesure que les journaux sont collectés, SolarWinds Security Event Manager les analyse et les mettre dans un format commun et lisible, créant un emplacement central pour les analyste afin d'étudier facilement les menaces potentielles, de préparer les audits et de stocker les journaux.[19]

SolarWinds offre un essai gratuit entièrement fonctionnel pendant 30 jours.

Selon les critiques, SolarWinds ne dispose pas d'une suite de sécurité complète, mais offre de bonnes fonctionnalités et capacités de détection des menaces. [14]

McAfee ESM

McAfee Enterprise Security Manager (ESM) est considéré comme l'une des meilleures plates-formes SIEM en termes d'analyse. Les utilisateurs peuvent collecter divers journaux sur divers appareils via le système Active Directory. En termes de normalisation, le moteur de corrélation de McAfee peut facilement compiler différentes sources de données. Cela facilite la détection d'un incident de sécurité et l'hiérarchisation des menaces.[14]

AlienVault Unified Security Management (USM)

L'AlienVault OSSIM fourni par *AT&T* Cybersecurity est un outil SIEM open source basé sur la solution AlienVault USM. Semblable aux outils ci-dessus, AlienVault OSSIM combine plusieurs projets open source en un seul package. De plus, AlienVault OSSIM permet la surveillance des appareils et la collecte de journaux. Il fournit également la normalisation et la corrélation d'événements.

1.3.7 Tendances et défis sur le marché de la gestion d'événements et des informations de sécurité

Parmi les principaux facteurs responsables de la croissance du marché des SIEMs dans les prochaines années, citons le niveau croissant de l'évolution de la cybercriminalité, de la conformité et de les mandats réglementaires.[20]

Cependant, le coût élevé du déploiement et de l'évolutivité des logiciels de gestion des informations et des événements de sécurité peut réduire la croissance des revenus du marché des SIEMs[20]

MarketsandMarkets s'attend à ce que le marché total de logiciels de gestion des informations et des événements de sécurité devrait passer de 4,2 milliards de dollars(542.195 milliards de dinar algérien) en 2020 à 5,5 milliards de dollars (7,087.60 milliards de dinar algérien)en 2025, avec un TCAC de 5,5% au cours de la période de prévision.[21]

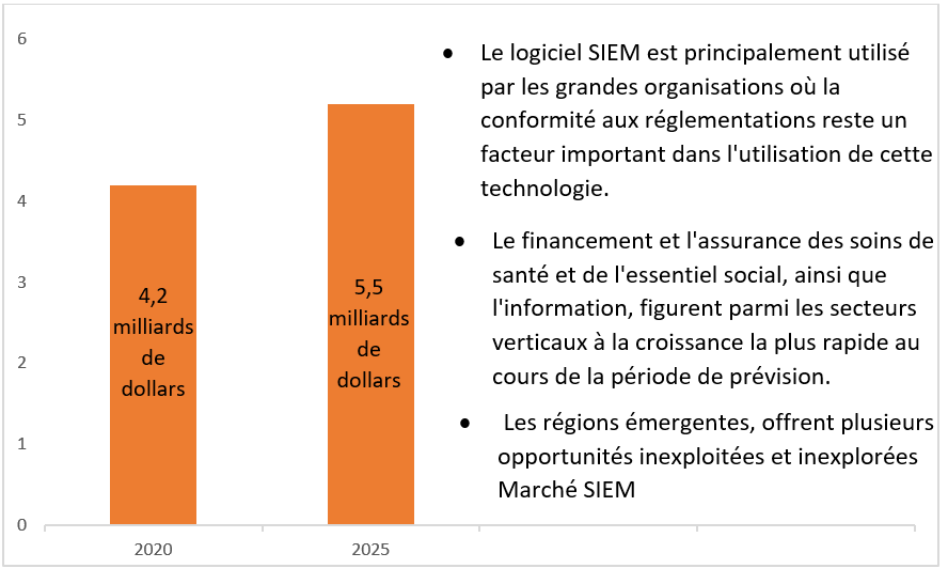


Figure 1.7: Taille du marché SIEM, 2020-2025 (en millions de dollars).

1.4 Comparaison des solutions disponibles sur le marché

Chaque solution SIEM disponible sur le marché diffère d'une autre sur plusieurs plans, car chaque fabricant vise une catégorie bien spécifique du marché.

SIEM	Type d'entreprise	Plateforme SE	Déploiement	Principales caractéristiques
SolarWinds	Petites, moyennes et grandes entreprises.	Windows, Linux, Mac, Solaris.	Sur site et dans le cloud	Interface de recherche, Détection des activités suspectes au moment de l'événement.
Datadog	Petites, moyennes et grandes entreprises.	Windows, Mac, Linux, Debian, Ubuntu, CentOS, RedHat.	Sur site et SaaS.	Détection des événements de sécurité en temps réel. Observation des traces, des logs et plus à partir d'un tableau de bord. Règles de détection préconfigurées prêtes à utiliser.
IBM Security QRadar	Moyennes et grandes entreprises.	Red Hat , Linux	Cloud, SaaS et sur site	Moteur de détection avancé Simplicité de la mise en place. Détection d'intrusion. Fonctions analytiques.
Splunk	Petites, moyennes et grandes entreprises.	Windows, Linux, Mac, Solaris.	Sur site et SaaS	Il combine l'analyse du réseau avec la gestion des logs. Excellent outil d'analyse. Analyse historique. Investigateur d'actifs.
McAfee ESM	Petites, moyennes et grandes entreprises.	Windows et Mac.	Sur site, cloud ou hybride	Consolidation des logs. Surveillance en direct. Collection via l'Active Directory.
RSA	Moyennes et grandes.	Red Hat Enterprise Linux	Cloud et sur site.	Surveillance de l'activité du réseau. Des graphes en direct. Outils analytiques

Figure 1.8: Comparatif des outils SIEM du marché.

1.5 Conclusion

Dans ce premier chapitre, on a commencé par définir les logs et expliquer leurs différents formats et types en général, puis on a exploré ce qu'est SIEM et quelles sont ses missions, et on a expliqué le cycle de vie d'un log dans le SIEM. Ensuite, on a parlé de certaines solutions SIEM open source et commerciales, et a fait une simple comparaison entre certaines d'entre elles.

Chapter 2
Conception et analyse des besoins

2.1 Introduction

Dans ce chapitre, on va présenter l'étude et la conception de notre système,et pour éclaircir cette conception, on va d'abord présenter les besoins de système, les différents diagrammes ainsi que l'architecture générale et expliquer chaque composant de cette architecture.

2.2 Présentation d'organisme d'accueil

Al Salam Bank-Algeria est agréée par la banque d'Algérie en septembre 2008. Elle débute son activi-té avec pour objectif principal d'offrir à sa clientèle des produits et services bancaires innovants.
Al Salam Bank-Algeria œuvre conformément à une stratégie claire visant à soutenir la croissance économique de l'ensemble des secteurs d'activités du pays, elle offre des services bancaires nova-teurs, aux fins de répondre aux attentes du marché, de la clientèle et des actionnaires. Banque alternative, Al Salam Bank-Algeria se caractérise par son engagement au respect des principes de la sharia dans toutes ses transactions.

2.3 Spécification des besoins de système

Dans cette section, on va identifier l'ensemble des besoins que notre solution doit fournir, les contraintes auxquelles est soumis la solution pour sa réalisation et son bon fonctionnement.
Cette identification sera répartis en identification des acteurs, exigences fonctionnelles et exigences non fonctionnelles.

2.3.1 Identification des acteurs

- **Administrateur principal:** C'est l'acteur principal qui a pour rôle de gérer et administrer toutes les opérations de système.
- **Administrateur:** Ayant pour mission principale d'examiner les alertes et les incidents et de diriger le processus de collection.

2.3.2 Exigences fonctionnelles

Dans cette section, on va présenter l'ensemble des besoins fonctionnels auxquels devrait répondre notre solution.

La solution à concevoir doit répondre aux exigences fonctionnelles qui se résument dans les points suivants :

► Le système doit permettre

- l'authentification des utilisateurs par un nom d'utilisateur et un mot de passe pour accéder aux différentes fonctionnalités.
- une collection dynamique de fichier log dans une heure spécifiée par l'utilisateur,en utilisant les informations stockées sur ce fichier.
- Stocker les informations de ce fichiers.
- Normaliser les événements de ce fichier.
- Stocker les événements dans la base de données.
- Stocker et indexer les événements normalisés de ce fichier.
- Analyser les événements normalisés en suivants des règles définit par l'utilisateur.
- stocker le résultat d'analyse dans la base de données.

► Le système doit permettre au Administrateur principal

- Gérer les utilisateurs.
- Gérer les groupes des permissions
- Gérer les tables de système.

► Le système doit permettre au Administrateur

- Gérer les entrées.
- Gérer les règles d'analyse.
- Effectuer des recherches et des filtres sur les événements.
- Modifier les paramètres de système.
- Gérer les tables de système, si autorisé.

2.3.3 Exigences non fonctionnelles

Il s'agit des besoins qui caractérisent le système. Ce sont des besoins en matière de performance, de type de matériel ou le type de conception. [22]
L'ensemble des extensions à réaliser doivent respecter les besoins suivants :

Type d'interface homme-machine :

Une interface graphique d'une application Web (interface web). Elle se manipule à l'aide d'un navigateur Web.

Intégrité de données :

L'intégrité des données est un processus qui garantit que les données sont exactes et cohérentes tout au long de leur cycle de vie, il comprend les points suivants:

- **Traçabilité :** le système doit conserver les traces des mouvements de l'information.
- **Validation de l'entrée:** le système doit vérifier et valider l'ensemble de données fournies par une source connue ou non connue (utilisateur , autre application, serveur ou toute autre source) pour assurer que l'entrée est correcte).
- **Validation des données :** les caractéristiques techniques et les attributs clés doivent être organiser dans le système pour certifier que les processus de données seront valides.

Confidentialité :

Les données ne doivent pas être modifiées que par des personnes autorisées et selon un processus défini.

2.4 Diagramme de cas d'utilisation(Use case)

Le diagramme de cas d'utilisation joue un rôle important lors du développement des logiciels. Il a été proposé en UML comme notation pour décrire les exigences et le comportement d'un système logiciel. Un cas d'utilisation spécifie une séquence d'actions, avec variantes éventuelles, réalisée par le système en interaction avec des acteurs de système[23].
Dans cette section, on va présenter le diagramme de cas d'utilisation global de notre application (fig 2.1). Par la suite on va détailler ce diagramme en montrant les descriptions textuelles et quelques diagrammes de cas d'utilisation principaux afin de mieux comprendre le fonctionnement de système.
La figure ci-dessous représente le diagramme de cas d'utilisation global de notre système.

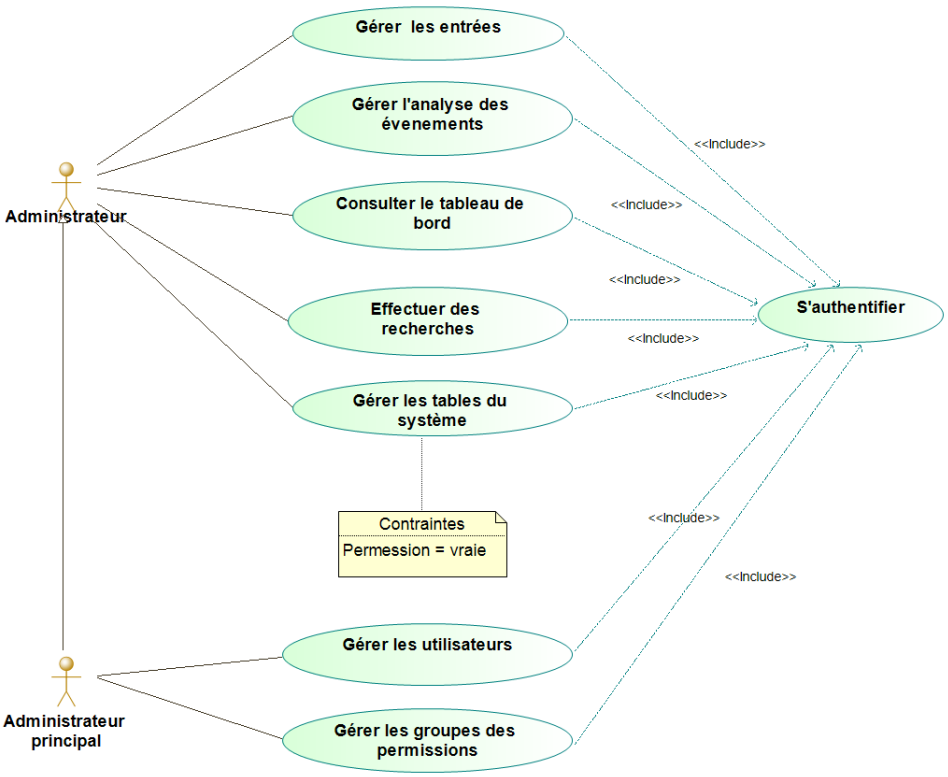


Figure 2.1: Diagramme de cas d'utilisation de système.

2.4.1 Cas d'utilisation «Gérer les entrées»

D'après la figure (fig 2.2), l'administrateur peut gérer les entrées de système en choisissant l'une des méthodes de téléchargement, soit le téléchargement à l'aide de protocole SFTP, à l'aide du chemin local, ou bien le téléchargement dynamique à l'aide de protocole SFTP ou de chemin local.

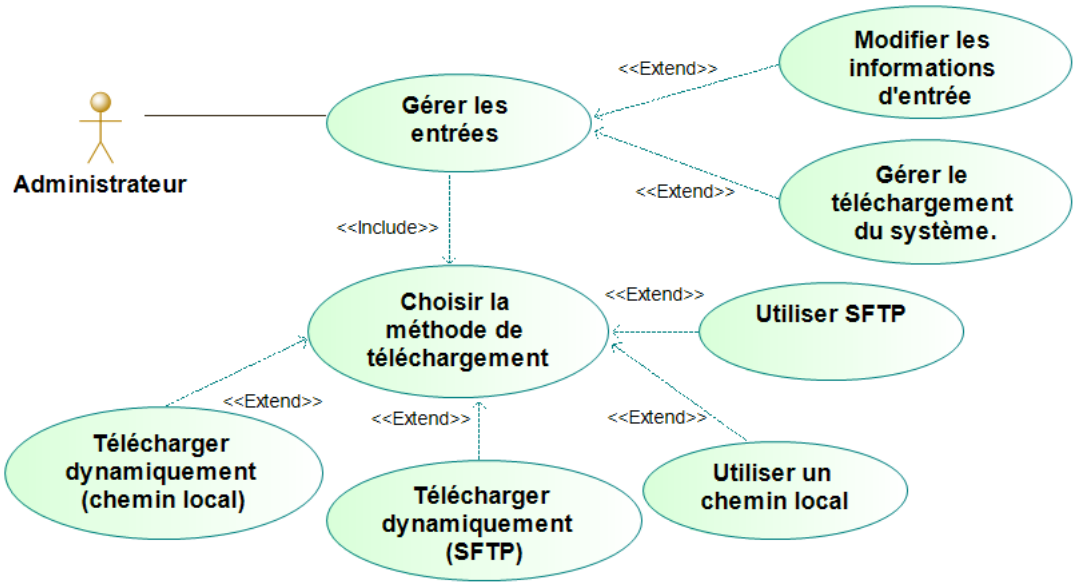


Figure 2.2: Cas d'utilisation détaillé «Gérer les entrées».

Cas d'utilisation «Gérer les entrées :Utiliser SFTP »

Partie 1 : Identification.	
Titre	Utiliser SFTP
Résumé	Ce cas d'utilisation montre comment l'administrateur peut gérer les fichiers comme entrée en utilisant le protocole SFTP;
Acteur principal	Administrateur
Partie 2 : Description des scénarios.	
Pré-conditions	**1-** Le serveur de stockage doit être à l'écoute d'une requête cliente éventuelle. **2-** Existence de toutes les informations nécessaires et valides pour établir une connexion avec le serveur de stockage. **3-** Existence de fichier dans le serveur de stockage. **4-** Les données de fichier doivent être valides.
Post-conditions	Les données du fichier seront bien enregistrées avec un retour de message de succès d'opération .
Scénario	**Scénarios d'exception :** Le contenu de fichier est invalide. **Scénario nominal :** L'acteur remplit le formulaire pour établir une connexion et récupérer le fichier. **Scénarios alternatifs :** Le serveur de stockage n'est pas à l'écoute.

Table 2.1: Description textuelle de cas d'utilisation «Gérer les entrées :utiliser SFTP»

Cas d'utilisation «Gérer les entrées : Utiliser un chemin local »

Partie 1 : Identification.	
Titre	Utiliser un chemin local.
Résumé	Ce cas d'utilisation montre comment l'administrateur peut gérer les fichiers comme entrée en utilisant le chemin local de fichier.
Acteur principal	Administrateur.
Partie 2 : Description des scénarios.	
Pré-conditions	**1-** Le fichier à télécharger doit être au format CSV. **2-** Les données de fichier doivent être valides.
Post-conditions	Les données de fichier seront bien enregistrées avec un retour de message de succès d'opération .
Scénario	**Scénarios d'exception :** Le contenu de fichier est invalide. **Scénario nominal :** L'acteur remplit les informations du serveur où le fichier est stocké, télécharge le fichier et envoie tout au traitement en cliquant sur un bouton. **Scénario alternatifs :** Le fichier n'est pas au format CSV.

Table 2.2: Description textuelle de cas d'utilisation «Gérer les entrées : Utiliser un chemin local»

Cas d'utilisation «Gérer les entrées : Télécharger dynamiquement (SFTP) »

Partie 1 : Identification.	
Titre	Télécharger dynamiquement (SFTP).
Résumé	Ce cas d'utilisation montre comment l'administrateur peut gérer les fichiers d'une manière dynamique en utilisant le protocole SFTP.
Acteur principal	Administrateur
Partie 2 : Description des scénarios.	
Pré-conditions	**1-** Le serveur de stockage doit être à l'écoute d'une requête cliente éventuelle. **2-** Les informations stockées dans la base de données pour établir une connexion au serveur de stockage doivent être valides. **3-** Existence de fichier dans le serveur de stockage. **4-** Les données de fichier doivent être valides.
Post-conditions	Les données de fichier seront bien enregistrées avec un retour de message de succès d'opération .
Scénario	**Scénarios d'exception :** Le contenu de fichier est invalide. **Scénario nominal :** L'acteur clique sur le bouton de téléchargement. **Scénario alternatifs :** Le serveur de stockage n'est pas à l'écoute.

Table 2.3: Description textuelle de cas d'utilisation «Gérer les entrées : Télécharger dynamiquement(SFTP)»

Cas d'utilisation «Gérer les entrées : Utiliser un chemin local »

Partie 1 : Identification.	
Titre	Gérer les entrées : Utiliser un chemin local.
Résumé	Ce cas d'utilisation montre comment l'administrateur peut gérer les fichiers d'une manière dynamique en utilisant le chemin local de fichier.
Acteur principal	Administrateur
Partie 2 : Description des scénarios.	
Pré-conditions	**1-** Les information sur le fichier doivent être valides. **2-** Les données de fichier doivent être valides.
Post-conditions	Les données de fichier seront bien enregistrées avec un retour de message de succès d'opération .
Scénario	**Scénarios d'exception :** Le contenu de fichier est invalide. **Scénario nominal :**L'acteur clique sur le bouton de «téléchargement dynamique à partir du chemin local».

Table 2.4: Description textuelle de cas d'utilisation «Gérer les entrées : Utiliser un chemin local»

2.4.2 Cas d'utilisation «Gérer les règles d'analyse»

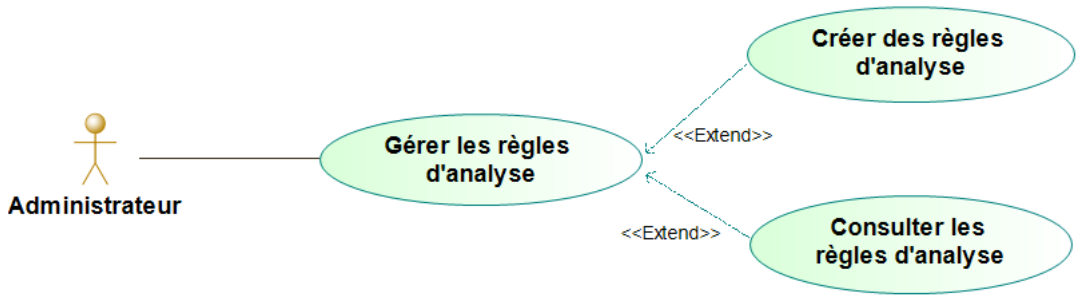


Figure 2.3: Cas d'utilisation détaillé «Gérer les règles d'analyse».

Cas d'utilisation «Gérer les règles d'analyse : Consulter les règles d'analyse»

Partie 1 : Identification.	
Titre	Gérer les règles d'analyse : Consulter les règles d'analyse.
Résumé	Permettre à l'acteur de voir la liste des règles d'analyse.
Acteur principal	Administrateur.
Partie 2 : Description des scénarios.	
Pré-conditions	Présence des règles dans la base de données.
Post-conditions	La liste des règles sera affichée à l'utilisateur sous forme d'une table.
Scénario	**Scénarios alternatifs :** la base de données des règles est vide.

Table 2.5: Description textuelle de cas d'utilisation «Gérer les règles d'analyse : Consulter les règles d'analyse»

Cas d'utilisation «Gérer les règles d'analyse : Créer des règles d'analyse»

Partie 1 : Identification.	
Titre	Gérer les règles d'analyse : Consulter les règles d'analyse.
Résumé	Permettre à l'acteur de créer des règles d'analyse.
Acteur principal	Administrateur.
Partie 2 : Description des scénarios.	
Pré-conditions	L'acteur doit donner des informations valides pour la création du règle.
Post-conditions	La règle sera enregistrée dans la base de données.
Scénario	**Scénario nominal :** L'acteur remplit les informations du règle et clique sur le bouton de Création. **Scénario alternatifs :** Les informations saisies du règle sont invalides.

Table 2.6: Description textuelle de cas d'utilisation «Gérer les règles d'analyse : Créer des règles d'analyse»

2.4.3 Cas d'utilisation «Consulter le tableau de bord»

Partie 1 : Identification.	
Titre	Consulter le tableau de bord.
Résumé	Permettre à l'acteur de suivre les alertes obtenues après la phase d'analyse.
Acteur principal	Administrateur
Partie 2 : Description des scénarios.	
Pré-conditions	La présence des alertes dans la base de données pour éviter un tableau de bord vide.
Scénario	**Scénarios alternatifs :** Un tableau de bord vide.

Table 2.7: Description textuelle de cas d'utilisation «Consulter les tableaux de bord»

2.4.4 Cas d'utilisation «Effectuer un recherche»

Partie 1 : Identification.	
Titre	Effectuer un recherche.
Résumé	Permettre à l'acteur de faire une recherche sur l'ensemble des événements.
Acteur principal	Administrateur.
Partie 2 : Description des scénarios.	
Pré-conditions	**1-** La présence des données saisies par l'utilisateur est nécessaire au système pour effectuer la recherche. **2-** Les événements doivent être indexés afin d'augmenter la vitesse de recherche.
Post-conditions	Le résultat de recherche sera exposé à l'utilisateur comme un tableau sur la même interface.
Scénario	**Scénarios d'exception :** les critères de recherche donnés par l'acteur sont invalides. **Scénario nominal :** L'acteur donne des critères de recherche sur la barre de recherche. **Scénarios alternatifs :** La barre de recherche est vide.

Table 2.8: Description textuelle de cas d'utilisation «Effectuer un recherche»

2.4.5 Cas d'utilisation détaillé «Gérer les utilisateurs».

La figure (fig 2.4) représente le diagramme de cas d'utilisation «Gérer les utilisateurs» qui illustre la possibilité de création et consultation des utilisateurs par l'administrateur principal. Après une consultation de liste des utilisateurs, l'administrateur principal peut soit trouver, soit modifier ou bien supprimer un utilisateur.

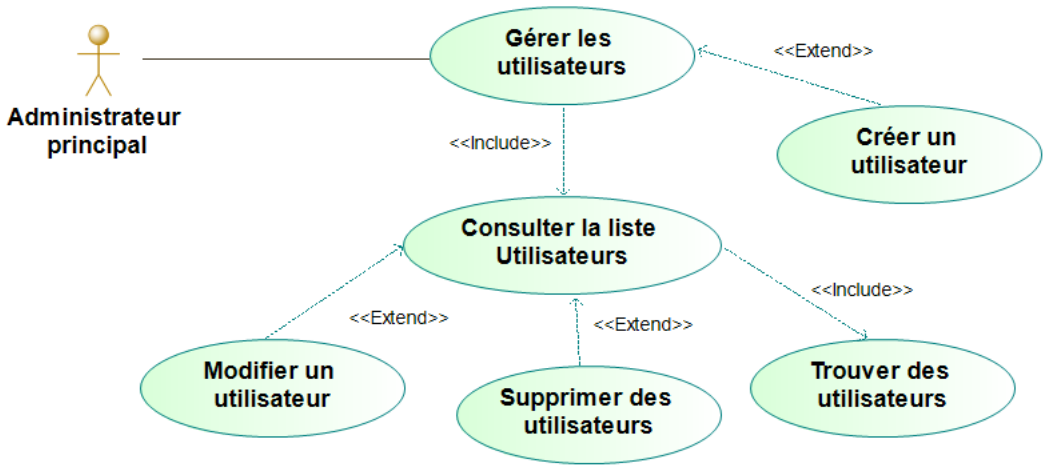


Figure 2.4: Cas d'utilisation détaillé «Gérer les utilisateurs».

Cas d'utilisation « Gérer les utilisateurs : Trouver des utilisateurs »

Partie 1 : Identification.	
Titre	Trouver des utilisateurs.
Résumé	Permettre à l'acteur de filtrer les utilisateurs de notre applications.
Acteur principal	Administrateur principal.
Partie 2 : Description des scénarios.	
Pré-conditions	Existence d'utilisateurs dans la base de données.
Post-conditions	Le résultat sera affiché à l'acteur lors du filtrage de la liste des utilisateurs en utilisant les mots saisis dans la barre de recherche.
Scénario	**Scénario nominal :** L'acteur donne les mots de filtre sur la barre de recherche.

Table 2.9: Description textuelle de cas d'utilisation « Gérer les utilisateurs : Trouver des utilisateurs ».

Cas d'utilisation « Gérer les utilisateurs : Supprimer des utilisateurs »

Partie 1 : Identification.	
Titre	Supprimer des utilisateurs.
Résumé	Permettre de supprimer un ou plusieurs utilisateurs.
Acteur principal	Administrateur principal.
Partie 2 : Description des scénarios.	
Pré-conditions	Existence d'utilisateurs dans la base de données.
Post-conditions	Le résultat sera affiché à l'acteur lors du filtrage de la liste des utilisateurs en utilisant les mots saisis dans la barre de recherche.
Scénario	**Scénarios d'exception :** L'utilisateur veut se supprimer. **Scénario nominal :** L'acteur sélectionne les utilisateurs à supprimer, puis il clique sur le bouton de suppression.

Table 2.10: Description textuelle de cas d'utilisation « Gérer les utilisateurs : Supprimer des utilisateurs ».

2.4.6 Cas d'utilisation détaillé «Gérer les groupes des permissions».

La figure (fig 2.5) représente le diagramme de cas d'utilisation «Gérer les groupes des permissions». Cette gestion est uniquement faite par l'administrateur principal :

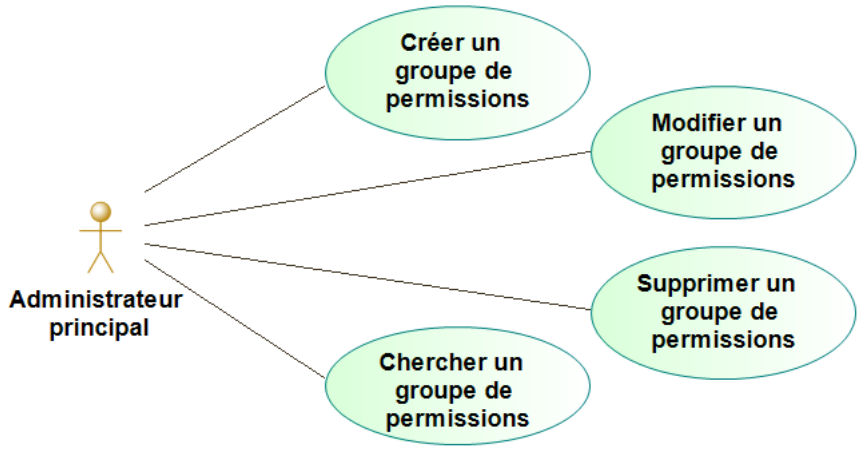


Figure 2.5: Cas d'utilisation détaillé «Gérer les groupes des permissions».

2.5 Architecture de système

Dans cette section, on va traiter l'architecture générale qu'on a conçue pour la réalisation de ce travail.

La figure (fig2.6) montre l'architecture générale de notre système qui se base sur:

1) La collection de fichier log.

2) L'envoie d'événements qui composent le contenu de fichier log au «Normalisateur d'évenement» pour qu'ils soient normalisés.

3) Le stockage d'informations extraites de ce fichier dans la base de données.

4) La récupération d'événements normalisés pour l'analyser, ainsi que d'autres informations utiles dans l'analyse.

5) Le stockage de résultats d'analyse dans la table d'alertes de la base de données.

6) L'envoie de de ces résultats sur le tableau de bord pour qu'ils soient affichés.

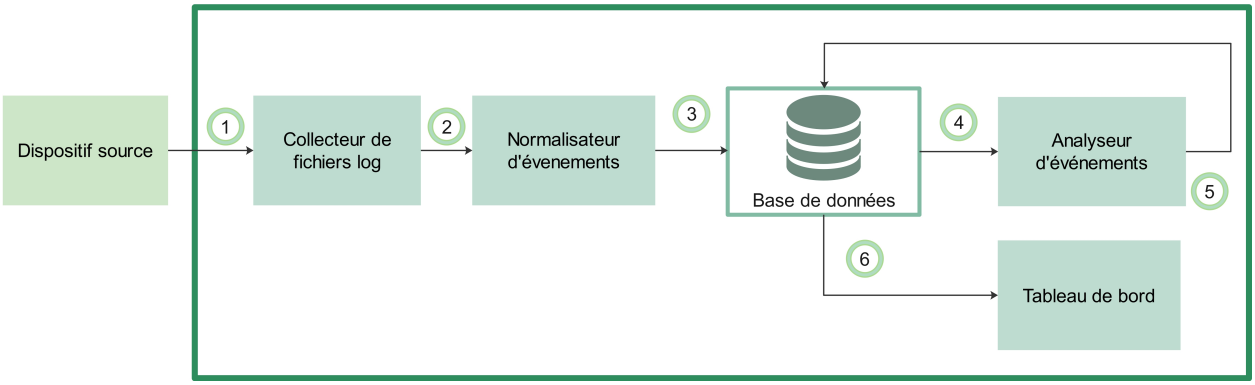


Figure 2.6: Architecture générale de système.

Les sous sections suivantes décriront chaque composant de cette architecture.

2.5.1 Dispositifs des sources

C'est l'emplacement qui stocke les fichiers log, il peut être un serveur de stockage, un emplacement local où l'utilisateur se connecte avec le système ou bien l'emplacement local de serveur.

2.5.2 Collection de fichiers log

La collection peut être effectuée par l'utilisateur de l'application ou par le système de manière dynamique.

► Utilisateur:

Quand un utilisateur veut télécharger un fichier log à l'application, la possibilité de téléchargement apparaît en entrant les données pour établir une connexion avec le serveur de stockage et demander le fichier, en entrant le chemin du fichier local ,ou en utilisant les données stockées (dynamiquement) soit

pour établir une connexion au serveur de stockage et récupérer le fichier, soit pour le récupérer en utilisant le chemin absolu qui part de la racine de fichier.

► **Système:**

La collection par le système se fait dans un moment spécifié par l'utilisateur et enregistré comme un paramètre de collection.

Une fois l'étape de collection est déclenchée par le système, un autre paramètre de collection est vérifie pour reconnue le type de collection.

la collection fait par système peut être distingué par deux type: une collection en utilisant un partage réseau et une collection en utilisant un chemin local.

Après la vérification de type de collection, une récupération des données se fait par le système pour effectuer la récupération de fichier log.

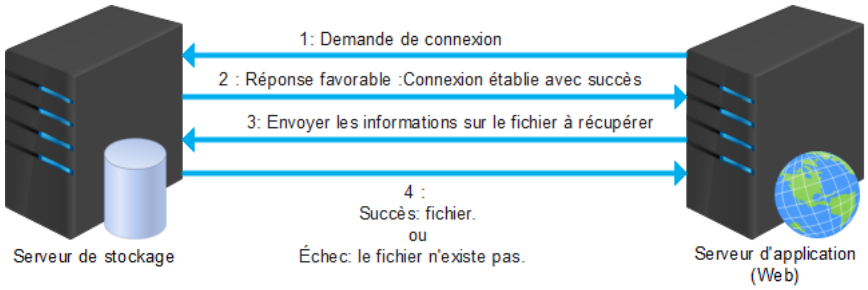


Figure 2.7: Récupération de fichier log par le système après l'établissement de connexion avec le serveur de stockage.

2.5.3 Normalisation d'évènements

Cette étape a pour objectif de normaliser les événements de fichier log collecté afin de faciliter et améliorer le processus d'analyse. Le fonctionnement de cette étape se fait par deux parties comme la figure suivante illustre:

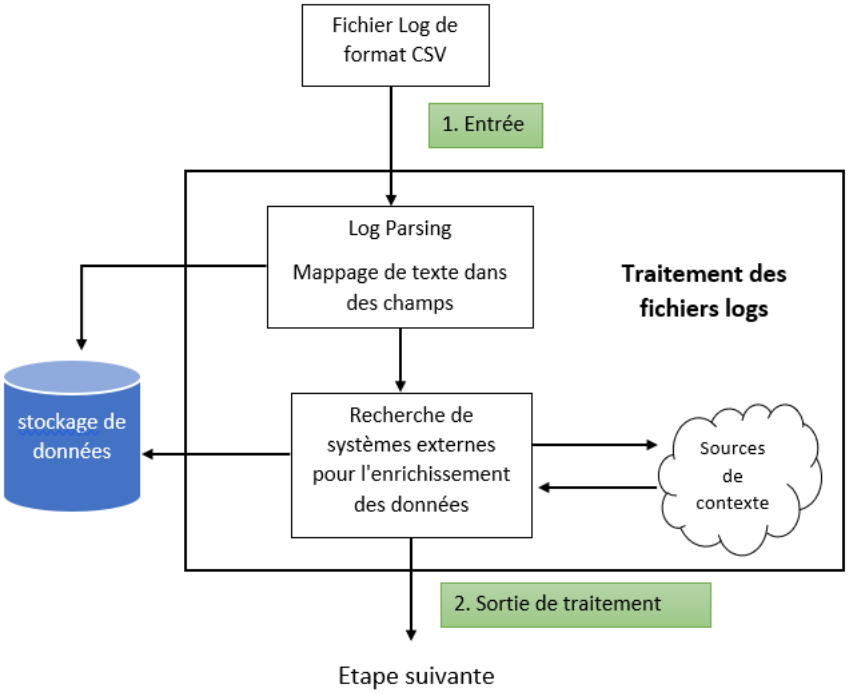


Figure 2.8: Architecture de normalisateur.

A. Parsing (Mappage de texte dans des champs)

Il consiste de prendre un format d'un événement et de le convertir en données structurées, tout on prenons en compte la façon dont les données sont conservées, la position de chaque élément de données dans l'enregistrement et le mot clé associé (fig 2.10).

```
csv_file = request.FILES["csv_file"]
file_data = csv_file.read().decode('ISO-8859-1')
lines = file_data.split("\n")
a=0
for line in lines:
    if a>0 :
        fields = line.split(";")
        """Récupérer le contenu entre chaque (;) comme un élément
        fields[i] tel que 0<= i <= (nombre_elements_de_fields-1)
        """
    a+=a
```

Figure 2.9: Exemple d'une lecture d'un fichier log de format csv avec un délimiteur point-virgule(;).

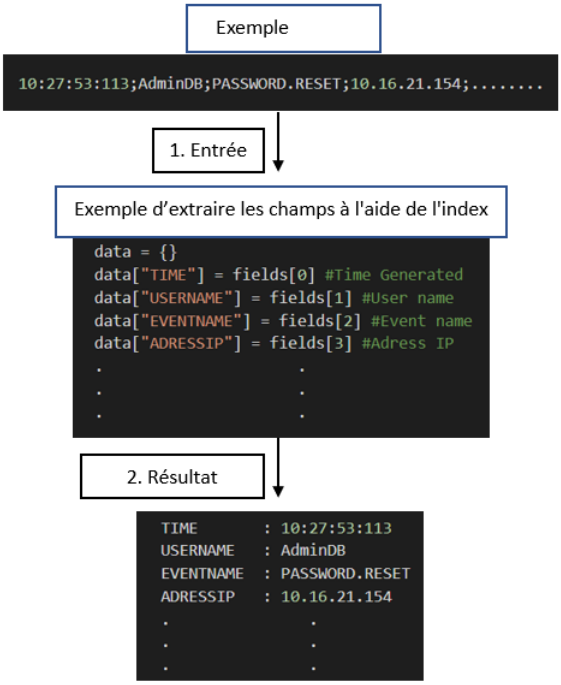


Figure 2.10: Exemple de Parsing d'un log.

B.Enrichissement d'événements

Il consiste d'ajouter d'informations importantes qui peuvent rendre les données plus utiles. Par exemple, si le log d'origine contient des adresses IP, mais pas les emplacements physiques (réels) des utilisateurs accédant au système , on peut définir ou utiliser des services de données pour trouver les emplacements et les ajouter aux données (fig 2.11).

La figure(2.12) représente la récupération de la représentation textuelle correspondant à l'adresse IP et au port / service.

Figure 2.11: Exemple-1- d'enrichissement d'un log.

```
import socket
#Le 'x' représente l'@ip
nameinfo=socket.getnameinfo((x, 0), 0)
if nameinfo[0]==x :
    data["System_Computer_Name"] = "NOCOMPUTERNAME"
else:
    data["System_Computer_Name"] =nameinfo[0]
```

Figure 2.12: Exemple-2- d'enrichissement d'un log.

2.5.4 Analyseur d'événements

L'analyseur d'événements va permettre de connecter les événements de sécurité qui ont été collectés et détermine les relations entre eux afin de déterminer les attaques au sein de l'infrastructure, il va permettre aussi de détecter les actions illégales de l'utilisateur de SI .

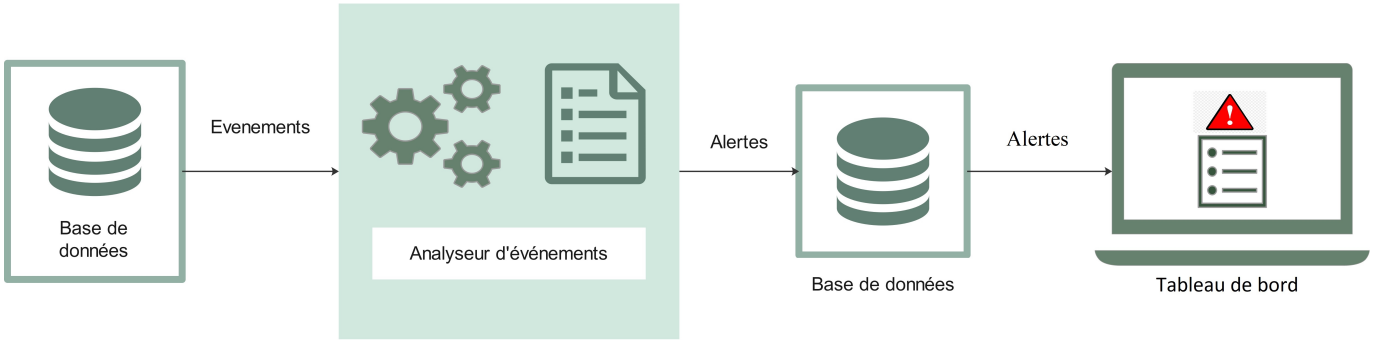


Figure 2.13: Analyseur d'événements.

2.5.5 Règles d'analyse

Une règle d'analyse permet de déclencher des alertes spécifiées aux actions des utilisateurs, autrement dit, des alertes spécifiées aux évènements.

Après avoir créé et stocké une nouvelle règle d'analyse, une analyse est effectuée sur tous les événements normalisés qui sont stockés dans la base de données.

Une fois les événements normalisés sont stockés, une analyse est effectuée d'une manière automatique en utilisant et en appliquant ces règles.

Chaque règle d'analyse possède un statut. Lors de l'analyse, l'analyseur vérifie ce statut, dans le cas où le statut est actif alors cette règle sera utilisée dans l'analyse, sinon elle sera négligée.

La figure suivante illustre le mécanisme d'utilisation d'une règle d'analyse:

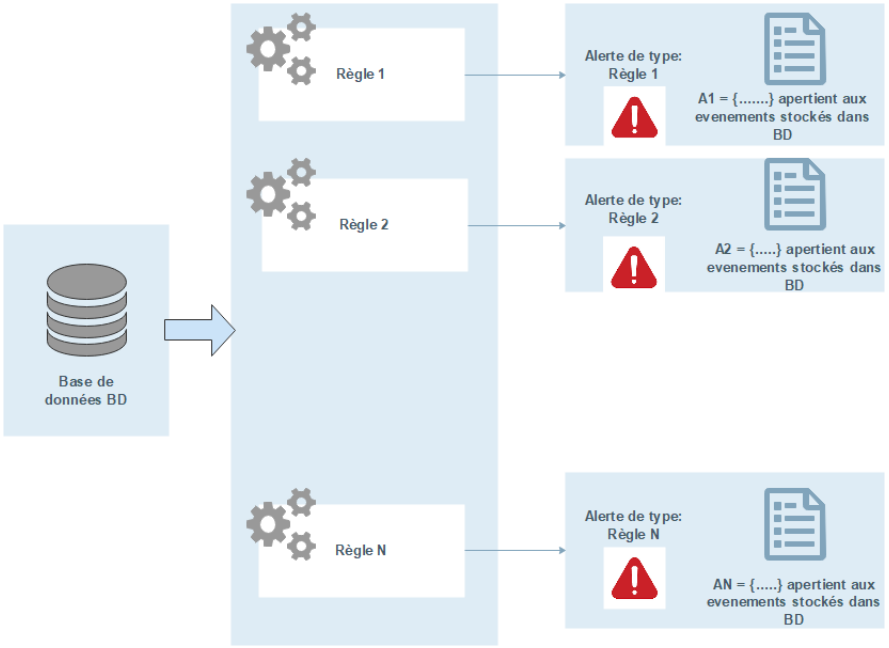


Figure 2.14: Le mécanisme d'utilisation d'une règle d'analyse.

2.5.6 Règles de corrélation

Une règle de corrélation va permettre de connecter les événements de sécurité qui ont été collectés, et détermine les relations entre eux (fig2.15).

Les mêmes propriétés appliquées aux règles d'analyse sont appliquées aux règles de corrélation.

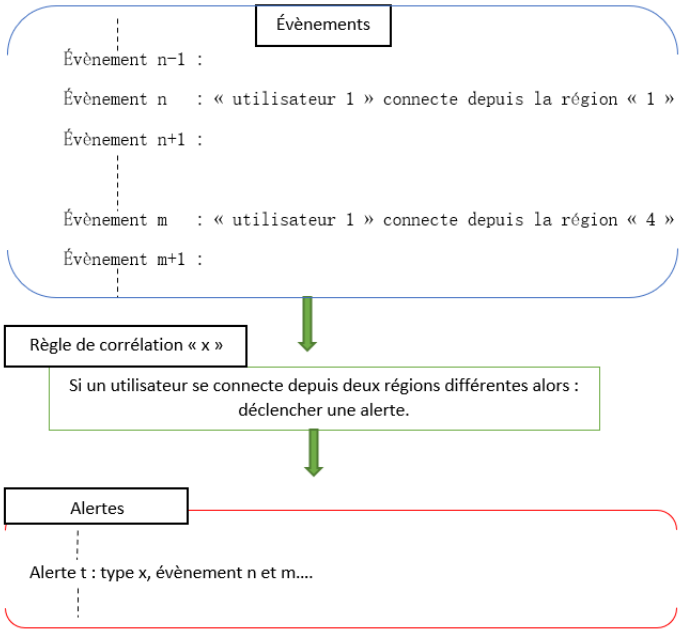


Figure 2.15: Exemple d'une règle de corrélation.

2.5.7 Tableau de bord (Dashboard)

Cette partie va permettre à l'utilisateur de l'application de recevoir un rapport sur les incidents de sécurité qui ont eu lieu à l'intérieur de l'infrastructure. Ce rapport va contient toutes les alertes obtenues après chaque phase d'analyse et qui affichées dans une interface appelée «Tableau de bord».

2.6 Base de Données

Une base de données contient un ensemble d'informations qui sont stockées, accessibles et gérées à l'aide d'un système de gestion de base de données (SGBD). Parmi les différents types de bases de données, il existe deux principaux types de bases de données: SQL(bases de données relationnelles) et NoSQL (bases de données non relationnelles).

Étude comparative et sélection des besoins

SQL et NoSQL font la même chose : stocker des données, mais ils le font chacun à leur manière avec des approches bien différentes. Le tableau suivant présente une comparaison entre une base de données SQL et une base de données NoSQL, afin de mieux justifier le choix de base de données de ce travail.

caractéristique	SQL	NoSQL
Stockage	Utilise des tables pour stocker et récupérer des données (un schéma prédéfini).	Peut utiliser les paires clé-valeur, les documents, les graphes ou autres (un schéma dynamique pour les données non structurées).
Évolutivité	BD SQL sont évolutives verticalement, ce qui signifie qu'ils peuvent augmenter la charge sur un seul serveur en augmentant la RAM, le CPU ou le SSD.	BD NoSQL sont évolutives horizontalement, ce qui signifie qu'ils peuvent gérer plus de charge en ajoutant plus de serveurs ou en partageant.
Type de données à stocker	BD SQL ne conviennent pas au stockage de données hiérarchique (la manipulation des données structurées limitées).	BD NoSQL conviennent mieux au stockage de données hiérarchique. Ils sont hautement préférés pour les ensembles de données volumineux.
Vitesse	BD SQL sont des bases de données normalisées dans lesquelles diverses données sont décomposées en tables logiques pour éviter la redondance et la duplication des données. Dans ce scénario, les bases de données SQL sont plus rapides que leurs homologues NoSQL pour les jointures, les requêtes, les mises à jour, etc.	L'entité de données particulière est stockée ensemble et non partitionnée. Ainsi, effectuer des opérations de lecture ou d'écriture sur une seule entité de données est plus rapide pour les bases de données NoSQL que pour les bases de données SQL. De plus, le NoSQL répond à la problématique actuelle du Big Data(la rapidité d'accès à l'information).

Table 2.11: Comparaison entre BD SQL et BD NoSQL.

2.6.1 Choix de base de données

Le choix de la base de données se repose sur la rapidité d'analyse et de recherche pour les événements normalisées d'une part, et sur la souplesse lors de la création des règles d'analyse d'autre part.
Pour cela, on peut dire que l'utilisation de base de données non relationnelle est la meilleure solution

pour l'analyse et le recherche , surtout que l'ensemble des évènements normalisés est très volumineux . Pour la création des règles, l'utilisation d'une base de données relationnelle est la meilleure solution.

2.7 Diagramme de classe

Pour décrire la structure des entités de système, cette partie va illustrer le diagramme de classe qui montre les classes intervenantes et la dépendance entre ces différentes classes sous la notion des relations.
Notre diagramme de classe est composée de quatre parties essentielles :

- Une pour la gestion des utilisateur de notre système.
- Une pour la gestion des fichiers log.
- Une pour la gestion des fichiers.
- Une pour la gestion des règles.

La figure suivante montre un aperçu de diagramme de classes UML (juste les classes et les relations):

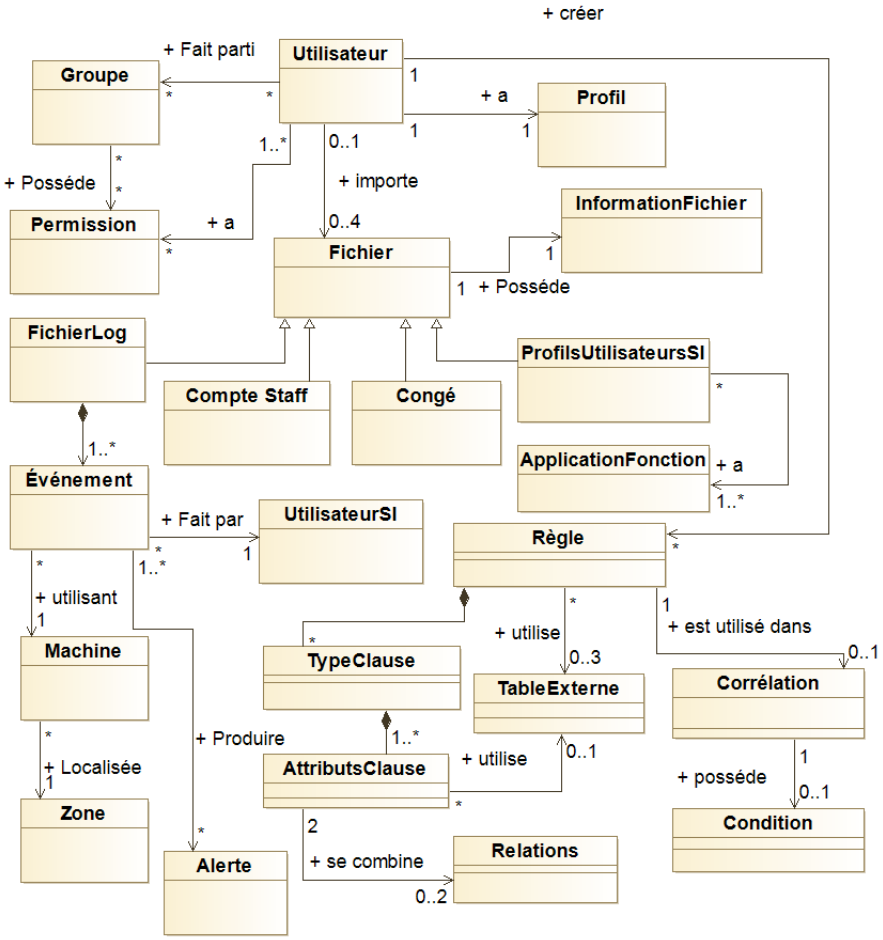


Figure 2.16: Aperçu de diagramme de classe UML (Overview).

Ci-dessous on va décrire les classes apparus dans le diagramme de classe.

Classe	Description
UtilisateurApp	Cette classe représente les utilisateurs de notre système, chaque utilisateur est caractérisé par un identifiant, un nom , un prénom,date de naissance,genre, email,numéro de téléphone, nom d'utilisateur et mot de passe, chaque utilisateur a un ou plusieurs permissions, et il est affecté à un groupe de permission.l'utilisateur qui a le rôle d'administrateur de site, il possède toutes les permissions.
Profil	Cette classe représente le profil d'un utilisateur de système,il est caractérisé par un id, une image. chaque profile est possédé par un utilisateur de système.
Groupe	Cette classe représente les groupes des permissions qu'un administrateur de site peut créer. Chaque groupe est identifié par son nom, et il a ses propres permissions.
Permission	Cette classe représente les permissions dans notre système, chaque permission est caractérisée par id , nom et le contenue au quelle la permission va être appliqué.

Table 2.12: Description textuel – Gestion des utilisateurs –

Nom de la table	Description
FichierLog	Représente les fichiers log entrées dans notre système, chaque fichier log est caractérisé par un identificateur, un nom, une date de téléchargement, une source, la méthode de téléchargement et l'utilisateur qui a téléchargé ce fichier si il existe.
Événement	Cette classe représente l'ensemble des événements que contient le fichier log téléchargé.Cet événement représente une action effectuée par l'utilisateur du système d'information dans l'infrastructure.
Zone	Cette classe représente les zones réseau où les utilisateur du système d'information font leur actions en utilisant des adresses IP qui appartient à ces zones , il est caractérisé par un identificateur, un nom et le réseau effectuer à cette zone.
UtilisateurSI	Elle représente les utilisateurs du système d'information de l'organisme,il est caractérisé par un nom.
Machine	Elle représente les machines par lesquels l'utilisateur du système d'information se connecte, ces machines sont caractérisées par un identifiant,nom d'ordinateur,adresse ip.
Alerte	Cette classe les alertes obtenues après une analyse,Une alerte peut être générée suite à un ou plusieurs événements.

Table 2.13: Description textuel – Gestion des fichiers log–

Nom de la table	Description
Fichier	Représente les fichiers téléchargés au système qui sont CompteStaff, congé, ProfilsUtilisateursSI et log. Elle est caractérisé par un nom.
InformationFichier	Représente les informations des fichiers(fichier log , fichier CompteStaff, fichier congé et fichier profilesUtilisateurSI) qui seront utilisées pour récupérer ces fichiers.
CompteStaff	Représente les comptes des employés de l'organisme qui sont extraites du fichier csv "CompteStaff".
Congé	Représente les informations des congés des utilisateurs du système d'information qui sont extraites du fichier csv "congé"
ProfilsUtilisateursSI	Représente les profiles des utilisateurs de système d'information qui sont extraites du fichier csv "ProfileUtilisateurSI".
ApplicationFonction	Représente les tables des données du système d'information.

Table 2.14: Description textuel – Gestion des fichiers –

Classe	Description
Règle	La classe règles contient les informations d'une règle d'une détection, tel que le statuts informe de son état (active/désactive).
TypeClause	Elle contient les types de clause d'une requête utilisés dans la règle . Exemple de type de clause dans SQL: WHERE,ORDER BY,...
AttributsClause	Elle contient les attributs soumis à la condition, les valeurs auxquelles ils vont être les comparer et les opérateurs de comparaison.
Relations	Elle contient les opérateurs logiques qui relient entre AttributsClause.
TableExterne	Elle représente l'ensemble des tables qui contribuent à la création d'une règle . Dans notre cas, on a trois table externe:Congé, profileUtilisateurSI,ComptesStaff.
Corrélation	Elle Représente l'ensemble des relations qui peuvent êtres entre les événements.Dans cette partie, on vas voir les fonctions d'agrégation comme un type de corrélation (COUNT,MAX..)
Condition	C'est la condition nécessaire pour créer une règle de corrélation, elle utilise la sortie de classe Corrélation comme un attribut de conditions.

Table 2.15: Description textuel des classes de diagramme de classe UML – Gestion des règles –

Pour la partie de gestion des règles, son diagramme de classe est donné par la figure suivante :

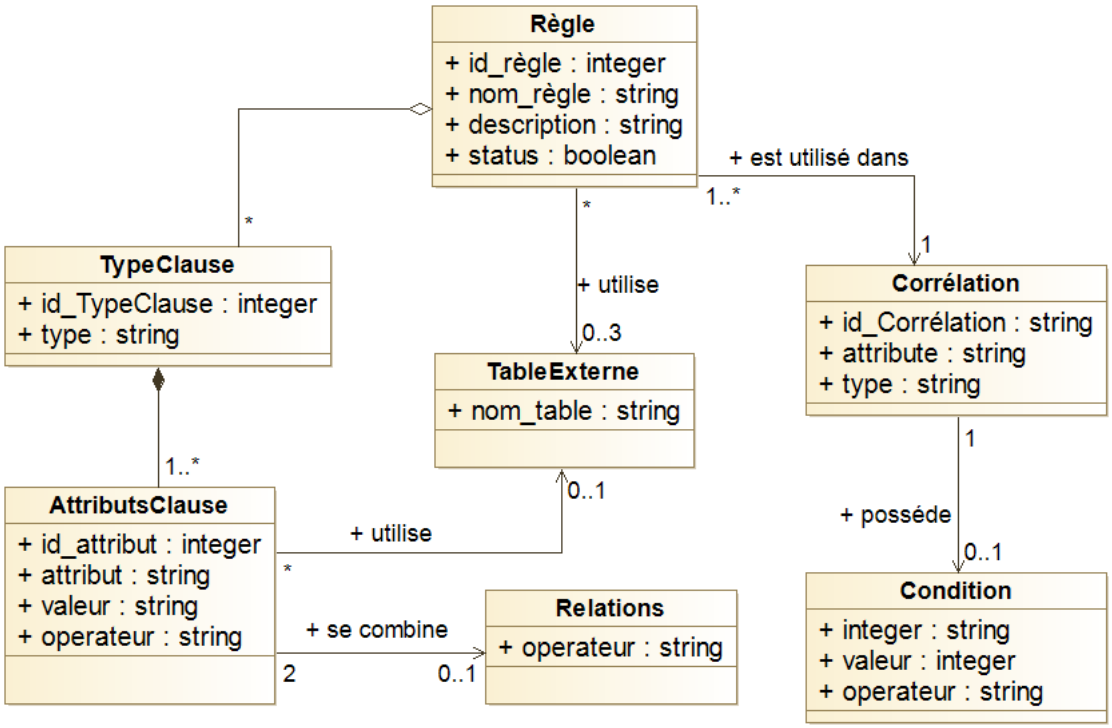


Figure 2.17: Diagramme de classe UML – Gestion des règles –.

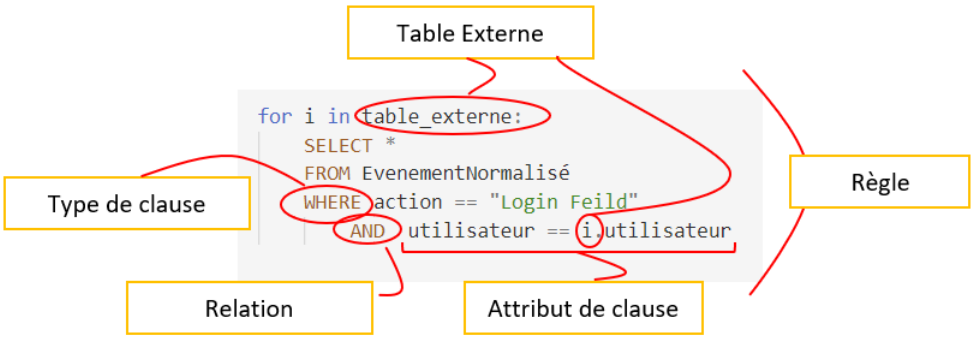


Figure 2.18: Un exemple d'une création d'une règle utilisant le diagramme de classe UML – Gestion des règles – .

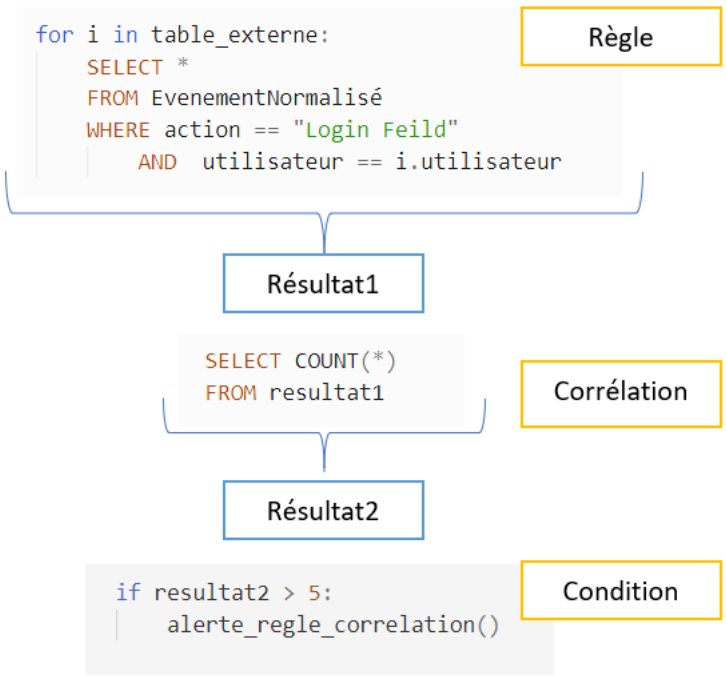


Figure 2.19: Un exemple d'une utilisation de diagramme de classe UML – Gestion des règles – .

2.8 Diagramme de séquence

Le diagramme de séquence représente les échanges de messages entre les acteurs et le système de manière chronologique.

Dans ce qui suit, on va présenter quelques scénarios de notre solution, on choisisse les scénarios relatifs à :

- Authentification d'un utilisateur
- Gérer les tables de système
- Gérer l'analyse des évènements

2.8.1 Diagramme de séquence «authentification d'un utilisateur»

Le processus d'authentification est faite par un couple (nom d'utilisateur, mot de passe), il est considéré comme une étape préliminaire pour vérifier et confirmer l'identité d'une entité qui veut accéder à l'espace des utilisateurs.

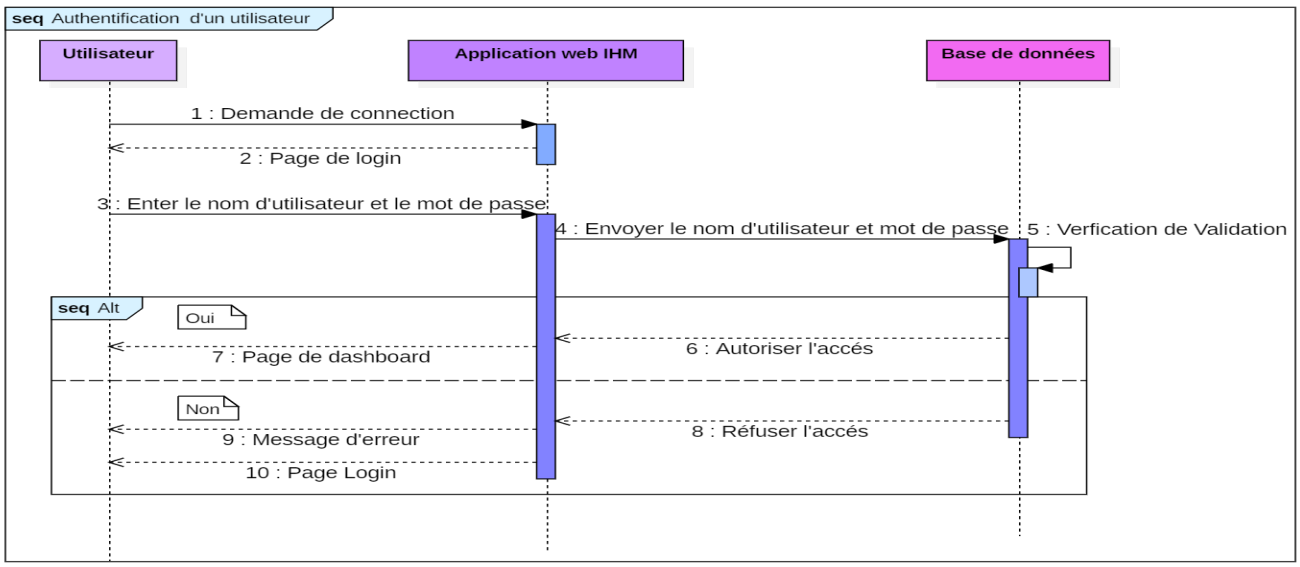


Figure 2.20: Diagramme de séquence «authentification d'un utilisateur».

2.8.2 Diagramme de séquence «Gérer les entrées du système»

Ce diagramme décrit le cas où l'utilisateur télécharge le fichier log à l'aide du protocole sftp (le protocole de transfert de fichiers SSH) qui fournit des capacités de transfert de fichiers sécurisées.
L'utilisateur peut télécharger le fichier en remplissant le formulaire pour établir une connexion; Ou cliquer simplement sur le bouton de téléchargement dynamique, dans ce cas, le système récupérera les informations de fichier de la base de données et établira une connexion avec le serveur de stockage afin qu'il puisse récupérer le fichier journal.

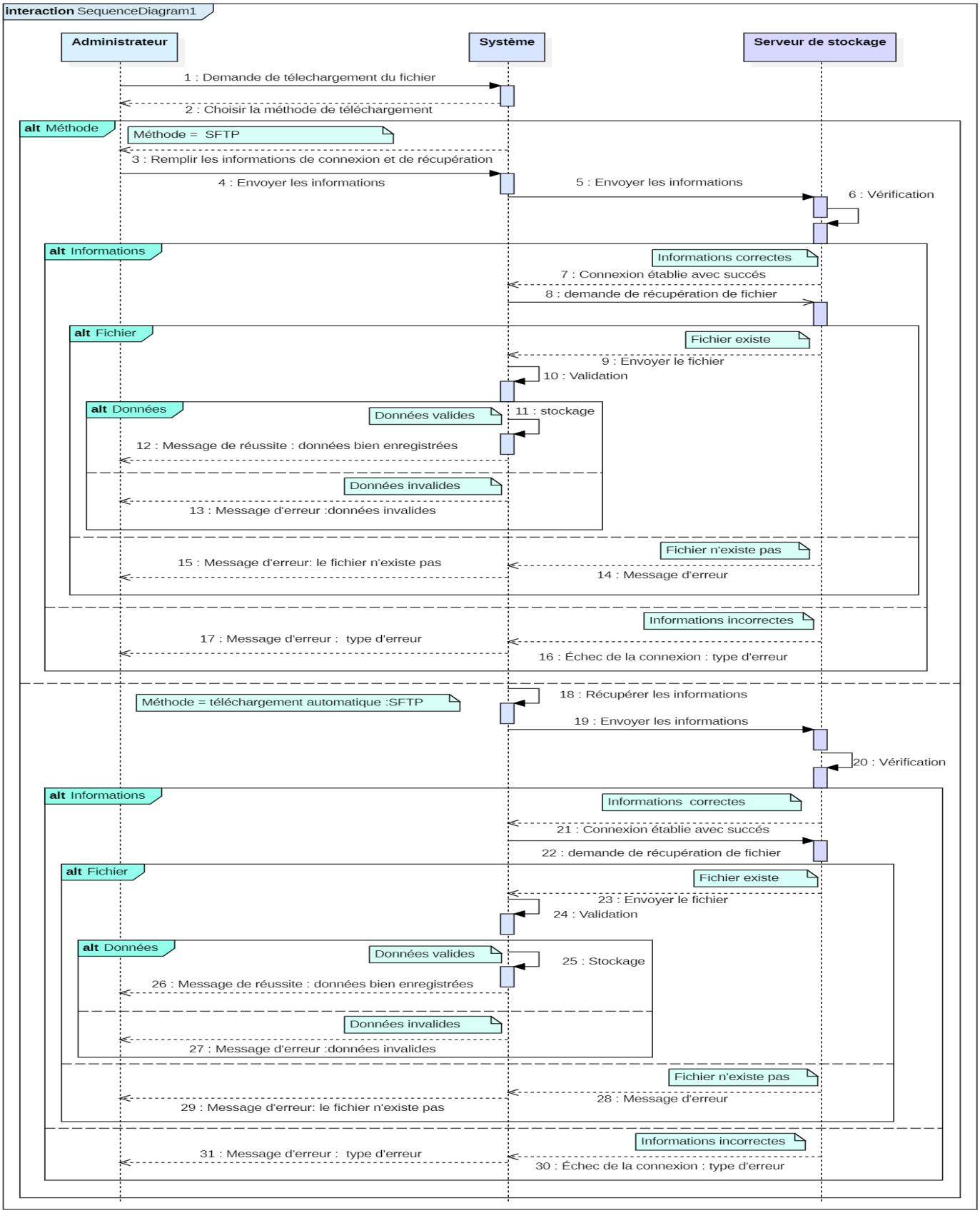


Figure 2.21: Diagramme de séquence «authentification d'un utilisateur».

2.8.3 Diagramme de séquence «Créer des règles»

Cette action permet à l'utilisateur de système de créer des règles, c'est-à-dire ajouter une règle à la table de base de données des règles. La figure suivante représente le diagramme de séquence de «Créer des règles»: Figure

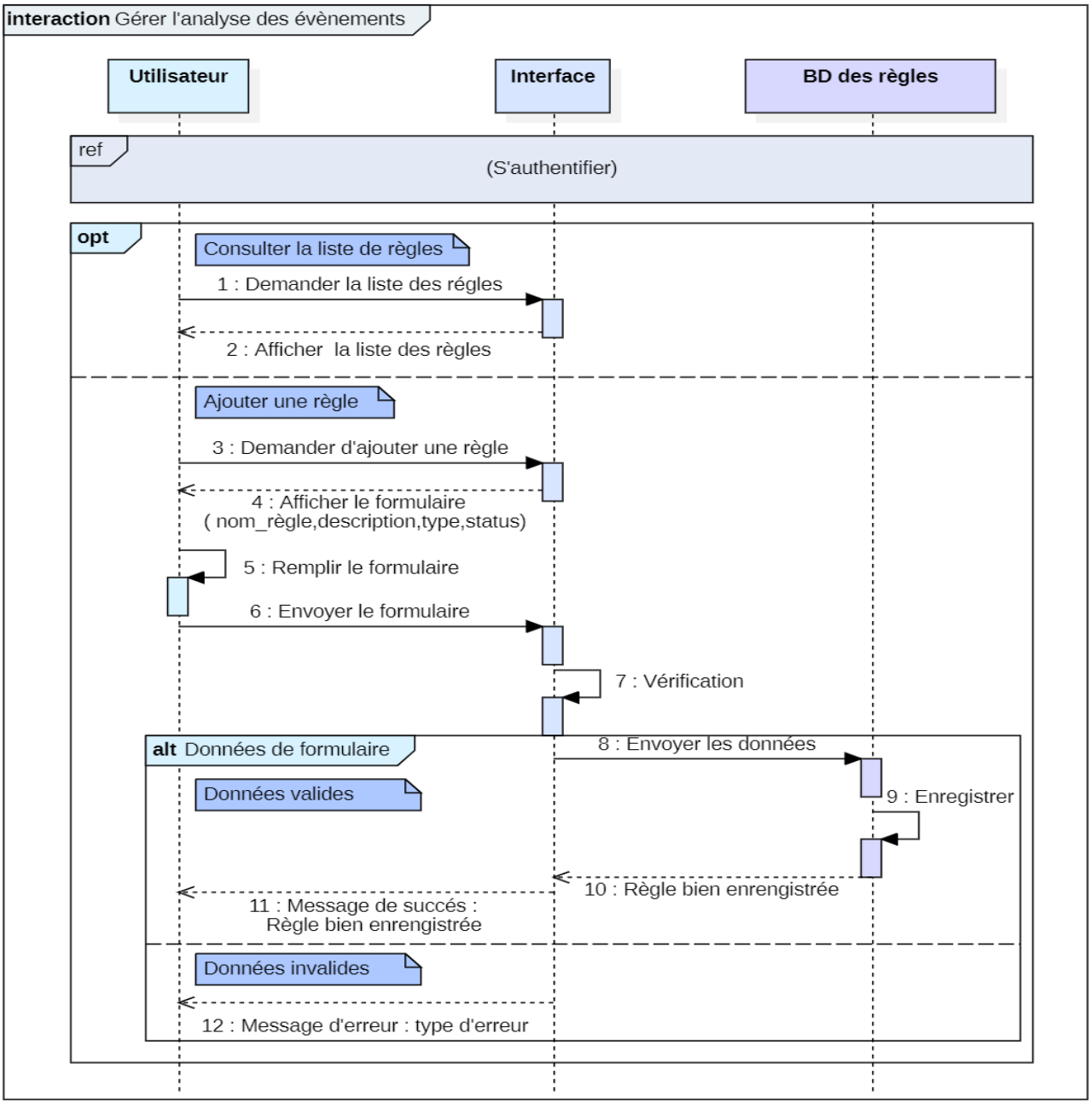


Figure 2.22: Diagramme de séquence «Créer des règles»

2.9 Modèle de navigation

La figure 2.23 représente le modèle de navigation de notre application qui donne une vision générale sur le design d'application et une illustration de la navigation entre ses différentes interfaces :

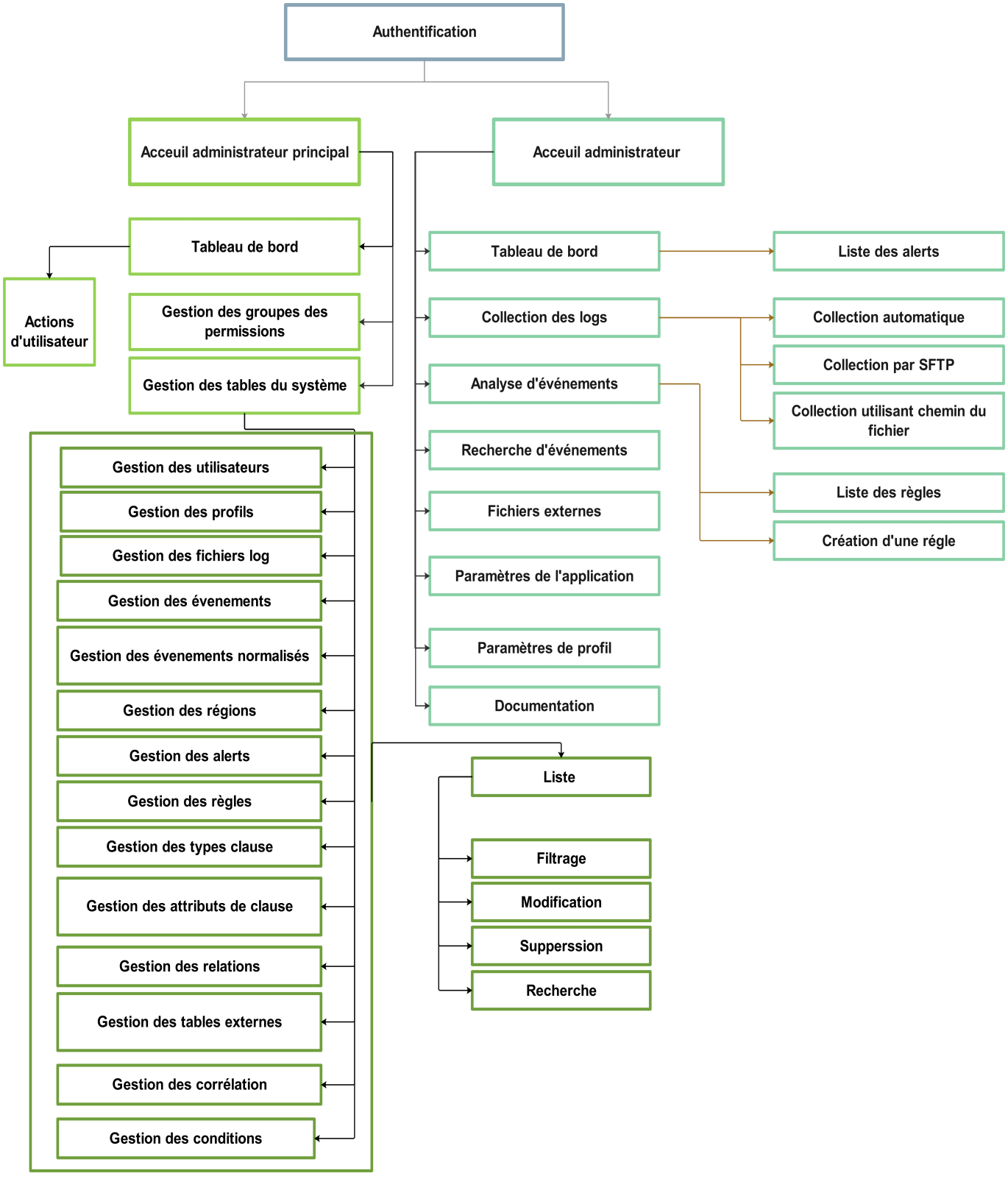


Figure 2.23: Diagramme de navigation

2.10 Conclusion

Dans ce chapitre, on a présenté la conception de la solution qu'on a proposé pour répondre au besoin exprimé par l'organisme d'accueil; en identifiant les diagrammes de cas d'utilisation, de séquence, de classe et de navigation pour obtenir une bonne vision et conclure une meilleure structure afin de faciliter la réalisation de cette solution.

Dans le chapitre suivant, on va montrer toutes les étapes,qu'on a suivies pour implémenter et réaliser notre solution.

Chapter 3
Mise en place la solution

3.1 Introduction

L'implémentation est la phase la plus importante après celle de la conception. l'objectif principale de cette phase est de mettre en œuvre la solution décrite dans le chapitre précèdent. Pour ce faire, on va commencer tout d'abord par préciser l'environnement de développement et définir les besoins utilisées dans le développement de front-end et back-end. Ensuite, on va présenter les différentes étapes d'installations des outils utilisées et la manière dont ils ont été implémentés. Enfin, on va présenter les différentes interfaces avec des tests afin de vérifier l'efficacité et le bon fonctionnement de notre application.

3.2 Environnement de développement

Dans cette section on va présenter l'environnement de développement qui est constitué par deux parties nommées environnement matériel et environnement logiciel.

3.2.1 Environnement matériel

On a utilisé pour les besoins de ce projet un ordinateur portable qui a les caractéristiques suivantes:

Processeur	Inlel(R) Core(TM) i5-6300u CPU @ 2.40GHz 2.50 GHz
Mémoire RAM	8,00GO
Type du Système	Système d'exploitation 64bits
Système d'exploitation	Windows 10

Table 3.1: Les caractéristique techniques d'environnement matériel

3.2.2 Environnement logiciel

L'environnement logiciel consiste les composants suivants :

A.**Logiciels utilisés :**

Les logiciels utilisés sont:

► **Celery(Distributed Task Queue)**

Est un système distribué simple, flexible et fiable pour traiter de grandes quantités de messages, tout en fournissant aux opérations les outils nécessaires pour maintenir un tel système. Il s'agit d'une file d'attente de tâches axée sur le traitement en temps réel, tout en prenant en charge la planification des tâches.[24]

► **Elasticsearch**

Elasticsearch est un moteur de recherche et d'analyse distribué et en open source pour tout type de données, y compris les données textuelles, numériques, géo spatiales, structurées et non structurées. Réputé pour ses API REST simples, sa nature distribuée, sa vitesse et sa scalabilité.[25]
La figure suivante montre les composants d'Elasticsearch, suivis d'une explication de chacun d'eux:

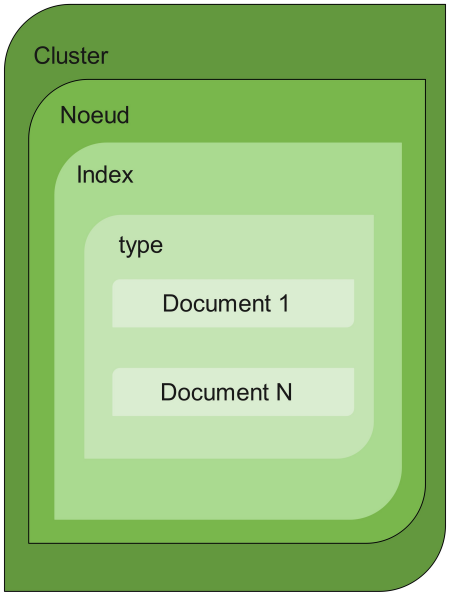


Figure 3.1: Elasticsearch composant.

• Cluster :
Un cluster est un ensemble d'un ou plusieurs nœuds (serveurs) qui stockent toutes les données. Il permet d'indexer et de rechercher des données dans l'ensemble des nœuds. Les clusters Elasticsearch sont dotés de partitions principales et de copies pour fournir un basculement en cas de panne d'un nœud. Lorsqu'une partition principale tombe en panne, la copie prend sa place.[26]
• Noeud :
Un nœud est un serveur unique qui fait partie de cluster, stocke les données et participe aux fonctionnalités d'indexation et de recherche du cluster.[26]

• Index :

Un index est une collecte de documents en lien les uns avec les autres.[26]

• type:

Est un type logique d'un index dont la sémantique est complète. Il est défini pour les documents qui ont un ensemble de champs communs. On peut définir plus d'un type dans dans un seul index. [26]

• document :

Est une unité d'information de base qui peut être indexé. Il est démontré dans le JSON.[26]

► Redis(Remote Dictionary Server)

Est un système de stockage de données clé-valeur en mémoire, open source et rapide, pour une utilisation en tant que base de données, cache, courtier de messages et file d'attente.Redis offre désormais des temps de réponse inférieurs à la milliseconde permettant des millions de demandes par seconde pour des applications en temps réel dans plusieurs domaines(jeu, technologie publicitaire,etc.).[27]

► SolarWinds SFTP / SCP Server

SolarWinds SFTP SCP Server est un outil gratuit pour le transfert sécurisé de fichiers réseau. Bien qu'il soit adapté au transfert de fichiers en toute sécurité sur Internet, ce programme a été conçu pour les administrateurs réseau. SFTP SCP Server est un excellent outil pour gérer les sauvegardes de configuration et les mises à jour système des services et des ressources cloud.[**?**]

B.**Modélisation conceptuelle :**

► Modelio Open Source 4.0

Modelio est un environnement de modélisation, prenant en charge une large gamme de modèles et de diagrammes UML / BPMN, et fournissant des fonctionnalités d'assistance de modèle.Le logiciel prend en charge les tests UML2, BPMN, MDA, XMI, TOGAF, SoaML.[28]

► StarUmL 3.2.2

Est un outil de modélisation UML (Unified Modeling Language) open source, il fournit onze types de diagramme différents et accepte la notation UML 2.0. Il support l'approche MDA (Model Driven Architecture) en prenant en charge le concept de profil UML en permettant de générer du code pour plusieurs langages.[29]

► EDRAW MAX

Est une solution logicielle permettant de réaliser des diagrammes, des graphiques et des dessins professionnels pour de nombreux besoins. Les développeurs, commerciaux et ingénieurs pourront l'utiliser pour concevoir des diagrammes de flux, des cartes mentales détaillées, des dessins électriques ou encore des plans pour les architectes.[30]

C.Langage de Modélisation:

► - UML (Unified Modeling Language)

Est un langage de modélisation normalisé composé d'un ensemble intégré de diagrammes, développé pour aider les développeurs de systèmes et de logiciels à spécifier, visualiser, construire et documenter les artefacts des systèmes logiciels, ainsi que pour la modélisation métier et autres systèmes non logiciels. . L'UML utilise principalement des notations graphiques pour exprimer la conception de projets logiciels. [31]

D.Éditeur de fichiers CSV:

► Sublime Text

Sublime Text est un éditeur de texte codé en Python et C++ qui présente une interface originale ainsi que de nombreuses fonctionnalités. Parmi ces dernières, on a la coloration syntaxique, l'autocomplétions et de plusieurs outils de recherche. Une barre latérale, aussi appelée minimap offre la possibilité de naviguer et prévisualiser rapidement le code source. De plus, il utilise des macros pour automatiser les tâches et simplifier le travail. Dans la même lignée, le logiciel intègre la sauvegarde automatique de projets. Sublime Text se distingue également grâce aux nombreux langages de programmation compatibles avec les standards JavaScript, C, C++, C#, LaTeX, Perl, PHP, Ruby, CSS, SQL, XML ou encore XLS. Cet éditeur de texte affiche une interface épurée et soignée qui s'avère simple et agréable à prendre en main. [32]

E.Framework:

► Django : [33]

Django est un framework Python de haut niveau, Créé par des développeurs expérimentés.Il permet un développement rapide de sites web sécurisés et maintenables. Django prend en charge la plupart des problèmes du développement web ,il aide à éviter les erreurs de sécurité classique en fournissant une infrastructure conçue pour "faire ce qu'il faut" pour protéger les sites web automatiquement. Ces objectifs se traduisent concrètement par :

- Un moteur de template très puissant implémentant un concept d'héritage de templates.
- Une gestion du routage d'URL (contrôleur frontal des applications) élégante, fondée sur les paterns d'expressions régulières écrit en Python .
- Django est l'un des frameworks web les plus matures pour Python. Ses règles de conception se concentrent largement sur la réduction du temps de développement d'applications Web.
- Il utilise des helpers et autres outils pratiques pour ne pas se répéter (Don't Repeat Yourself)
- Parmi les bons principes adoptés par Django, on peut également mentionner la facilité d'accès au testing unitaire conféré par un framework de test très simple à mettre en œuvre mais puissant.
- Django possède un système de sécurité intégré. Autrement dit, créer une application sécurisée, et optimisée.
- Les fonctions de sécurité intégrées fournies par Django aident les développeurs à protéger les applications Web contre une variété d'attaques – cross-site scripting, injection SQL,ect. En même temps, le framework Web améliore la sécurité des applications Web en évitant les erreurs de sécurité courantes liées au codage en Python.

*F.***Environnement de développement intégré**

► **Visual Studio Code**

Visual Studio Code est un éditeur de code open-source, gratuit et multi-plateforme (Windows, Mac et Linux), développé par Microsoft. Il est livré avec un support intégré pour JavaScript, TypeScript et Node.js et dispose d'un riche écosystème d'extensions pour d'autres langages (tels que C ++, C #, Java, Python, PHP, Go) et des environnements d'exécution (tels que .NET et Unity) .[34]

3.3 Développement Front-End

Il s'agit finalement des éléments du site que l'on voit à l'écran et avec lesquels on peut interagir. Ces éléments sont composés de HTML, CSS et de Javascript contrôlés par le navigateur web de l'utilisateur. Dans mon travail, j'ai utilisé html5, css3 , javascript ,jquery , boostrap4 .

► **HTML5 (Hypertext Markup Language 5)**

Est une version du célèbre format HTML utilisé pour concevoir les sites web, développé par le W3C (World Wide Web Consortium). Celui-ci se résume à un langage de balisage qui sert à l'écriture de l'hypertexte indispensable à la mise en forme d'une page Web. Lancée en octobre 2014, cette version

HTML5 apporte de nouveaux éléments et de nouveaux attributs par rapport à la version précédente. Elle offre par exemple la possibilité de définir le contenu principal d'une page Web, d'ajouter une introduction en header, d'insérer un sous-titre à un contenu multimédia de type vidéo.[35]

► CSS3 (Cascading Style Sheets 3)

Est un langage informatique utilisé sur Internet pour la mise en forme de fichiers et de pages HTML , il se présente comme une alternative à la mise en forme via des balises, notamment HTML. Un peu plus complexe à maîtriser, il permet un gain de temps considérable dans la mise en forme d'une page web par rapport à ces balises. Il permet d'appliquer des règles de mise en forme (titrage, alignement, polices, couleurs, bordures, etc.) à plusieurs documents simultanément.[36]

► JavaScript

JavaScript est un langage de script utilisé pour créer et contrôler le contenu dynamique d'un site Web, c'est-à-dire «tout ce qui bouge», actualise ou change d'une autre manière sur l'écran de client sans l'actualisation manuelle de la page Web. Il possède des fonctionnalités telles que: graphiques animés, diaporamas photo, formulaires interactifs etc.[37]

► jQuery

Est une bibliothèque JavaScript gratuite, libre et multiplateforme. Compatible avec l'ensemble des navigateurs Web (Internet Explorer, Safari, Chrome, Firefox . . .), Jquery facilite l'écriture de scripts. Il permet d'agir sur les codes HTML, CSS, JavaScript et AJAX et s'exécute essentiellement côté client.[38]

► Bootstrap 4

Est un framework open source développé par l'équipe du réseau social Twitter. Il utilise les langages HTML, CSS et JavaScript, fournit aux développeurs des outils pour créer facilement des sites web . Ce framework est pensé pour développer des sites avec un design responsive, qui s'adapte à tout type d'écran, et en priorité pour les smartphones. Il fournit des outils avec des styles déjà en place pour des typographies, des boutons, des interfaces de navigation et bien d'autres encore.[39]

3.4 Développement Back-End

Il s'agit du partie de code exécutée par le serveur, ce code doit être écrire dans le langage serveur. On a utilisé comme des technologies Backend les langages suivants:

► Python

Python est un langage de programmation interprété, orienté objet et de haut niveau .Ses structures de données intégrées de haut niveau, associées à un typage dynamique et à une liaison dynamique, le rendent très attractif pour le développement rapide des applications. Python prend en charge les modules et

les packages, ce qui encourage la modularité du programme et la réutilisation du code.[40]

► **SQLite**

SQLite (est une bibliothèque écrite en langage C qui propose un moteur de base de données relationnelle accessible par le langage SQL.Contrairement aux serveurs de bases de données traditionnels, comme MySQL ou PostgreSQL, sa particularité est de ne pas reproduire le schéma habituel client-serveur mais d'être directement intégrée aux programmes. L'intégralité de la base de données (déclarations, tables, index et données) est stockée dans un fichier indépendant de la plateforme. [41]

3.5 Mise en œuvre de la solution

Dans cette partie, on vas décrit les étapes de mise en œuvre de la solution et comment utiliser les outils définis précédemment.

3.5.1 Installation et configuration des outils

La première étape est centrée sur l'installation et la configuration des outils utilisés dans le développement.

► **Django**

Pour installer Django, il suffit de suivi les étapes saisies dans le site officiel de Django: https://docs.djangoproject.com/en/3.1/howto/windows/.

Ensuite, installer les packages essentiels à l'aide de « **pip**» (Un gestionnaire de packages python). On cite l'essentiel:

- $pip\ install\ pysftp$
- $pip\ install\ dnspython$
- $pip\ install\ elasticsearch == 7.8.0$
- $pip\ install\ elasticsearch-dsl == 7.2.1$
- $pip\ install\ celery == 4.4.6$
- $pip\ install\ redis == 3.5.3$

► **Elasticsearch**

On doit tout d'abord installer Java et le définir comme une variable d'environnement.(La version de Java doit être entre 7 et 8).Ensuite, télécharger le fichier zip élastique à partir de site officiel: « https://www.elastic.co/downloads/elasticsearch ».

► **Celery**

Celery va être utilisé comme un package python pour gérer les tâches planifiées dans le back-end de notre application.

► **Redis**

Il va être utiliser comme un courtier de messages. Cela signifie qu'il gère la file d'attente des "messages" entre Django et Celery.

Tout d'abord, On doit installer Redis à partir de la page de téléchargement officielle: https://redis.io/download .

Une configuration se fait au niveau de fichier de réglages et configuration de notre projet Django **Settings.py**.

Comme une première étape, on doit ajouter les modules **django_ elasticsearch_ dsl** et **celery** à la liste des **INSTALLED_APPS** dans le fichier **settings.py** du projet (fig3.2).

Ensuite,on ajoute la configuration de Elasticsearch qui consiste à saisir le numéro de port d'écoute du serveur Elasticsearch et son adresse IP(fig3.3).

```
31  # Application definition
32
33  INSTALLED_APPS = [
34      'django.contrib.admin',
35      'django.contrib.auth',
36      'django.contrib.contenttypes',
37      'django.contrib.sessions',
38      'django.contrib.messages',
39      'django.contrib.staticfiles',
40      'user',             # Define (user app) as module of INSTALLED_APPS
41      'crispy_forms',
42      'django_elasticsearch_dsl',        # Define elasticsearch_dsl as module of INSTALLED_APPS
43      'celery' # Celery
44  ]
45
```

Figure 3.2: Capture de configuration d'applications définies dans Django.

```
141
142 ELASTICSEARCH_DSL={
143     'default': {
144         'hosts': 'localhost:9200'
145     },
146 }
147
```

Figure 3.3: Capture de configuration de Elasticsearch dans python.

La figure suivante montre la configuration des paramètres de Celery dans le fichier **Settings.py**. Ces paramètres indiquent à Celery d'utiliser Redis comme courtier de messages ainsi que de l'endroit où se connecter. Ils indiquent également à Celery de s'attendre à ce que les messages soient transmis dans les deux sens entre les files d'attente de tâches Celery et le courtier de messages Redis.

```
# Tell Celery that it should send and read messages from a redis Broker
# celery

CELERY_BROKER_URL = 'redis://localhost:6379'
CELERY_RESULT_BACKEND = 'redis://localhost:6379'
CELERY_ACCEPT_CONTENT = ['application/json']
CELERY_RESULT_SERIALIZER = 'json'
CELERY_TASK_SERIALIZER = 'json'
CELERY_TIMEZONE = 'Europe/Paris'
#CELERY_TIMEZONE = 'Africa/Algiers'
```

Figure 3.4: Configuration des paramètres de Celery dans Django.

3.5.2 Base de données

► Configuration de la base de données relationnelle

La configuration de base de données se fait au niveau de fichier **Settings.py** (fig3.5).

```
83  DATABASES = {
84      'default': {
85          'ENGINE': 'django.db.backends.sqlite3',
86          'NAME': os.path.join(BASE_DIR, 'db.sqlite3'),
87      }
88  }
```

Figure 3.5: Configuration de la base de données la base de données relationnelle.

► Construction des table de la base de données la base de données relationnelle

Une définition d'un modèle dans **Django** implique une définition d'une table de base de données où chaque champ de modèle est défini comme un attribut correspond à une colonne de cette table. la figure (fig3.6). montre quelques tables définies dans notre projet.

```
118  # Define the correspondence  regions/IP_subnets model
119  class Region(models.Model): ···
123  # Define Leave model (External table)
124  class Leave(models.Model): ···
135  # Define StaffAccount model (External table)
136  class StaffAccount(models.Model): ···
139  # Define "profile of SI user's" model (External table)
140  class ProfileSiUsers(models.Model): ···
153  # Define DirectoryFiles model
154  # This table contains file information that the user want to upload it
155  class DirectoryFiles(models.Model): ···
169  # Define Alert model
170  class Alert(models.Model): ···
177
178  class Rules(models.Model): ···
195
196  class ExternalTables(models.Model): ···
207
208  class ClauseType(models.Model): ···
221
222  class ClauseAttribute(models.Model): ···
275
276  class Relation(models.Model): ···
286
287  class Correlation(models.Model): ···
328
329  class Condition(models.Model): ···
```

Figure 3.6: Capture d'une partie de construction des table de la base de données.

la figure suivante montre la définition de table **Règles** de base de données.

```python
# Define Rules model
class Rules(models.Model):
    id_rule = models.AutoField(primary_key=True)
    rule_name = models.CharField(max_length=200, null=False, blank=False)
    rule_description  = models.CharField(max_length=99999, null=False, blank=False)
    status = models.IntegerField(default = 1,
                                 blank = True,
                                  null = True,
                                  choices =(
                                  (1, 'Active'), (0, 'Inactive')
                                  ))

    def __str__(self):
        return self.rule_name

    class Meta:
        # Add verbose name
        verbose_name = 'Rule'
```

Figure 3.7: Code source de la table Règles.

► Configuration de la base de données non relationnelle

On donne l'adresse IP d'Elasticsearch et Django va automatiquement envoyer les événements normalisés vers l'instance d'Elasticsearch(fig:3.3).

► Construction de l'index de la base de données non relationnelle

La figure suivante montre une partie de la création de l'index d'évènements normalisés.

```python
# write an index (like a model in models.py)
@registry.register_document
class NormalizedEventDocument(Document):
    # Declare the fields name (username and sorce_region) as KeywordField
    # KeywordField is used fo filtering, aggrigation
    k_user = fields.KeywordField(
        fields={'raw': fields.KeywordField()}
    )
    source_region = fields.KeywordField(
        fields={'raw': fields.KeywordField()}
    )
    client_ip_address = fields.KeywordField(
        fields={'raw': fields.KeywordField()}
    )
```

Figure 3.8: Capture d'une partie de la construction de l'index d'évènements normalisés.

Le Document mentionné dans la figure sert de wrapper (couverture) pour permettre à l'index d'être rédigé comme modèle de Django.

3.5.3 Récupération des fichier

Dans cette partie, on va parler sur la manière de téléchargement des fichiers sur le système.

La collection à partir d'un partage réseau

Cette partie repose sur le module **pysftp** et les informations nécessaires pour établir une connexion au serveur sur lequel les fichiers journaux sont stockés.

Au début ,une récupération des informations se fait au niveau de base de données, une partie de ces informations vas êtres utiliser pour que l'application puisse se connecte au **serveur SFTP** cible, et l'autre partie est utilisée pour récupérer de fichier et de l'enregistrer sur le serveur d'application.
La figure suivante montre le code source de la récupération de fichier log en utilisant le **sftp**.

```
obj = get_object_or_404(DirectoryFiles,filename = 'Log file')
cnopts = pysftp.CnOpts()
cnopts.hostkeys = None
try:
    # Establish a connection to the SFTP server
    with pysftp.Connection(host=obj.host, username=obj.username,password=obj.password,cnopts=cnopts) as srv:
        # Set the remote location
        remoteFilePath = obj.remote_file_path
        localFilePath = obj.local_file_path
        # Define the local path where the file will be saved
        # As a definition, add the string "_v_sftp" to the file name to be different from the local file
        filename, file_extension = os.path.splitext(localFilePath)
        localFilePath = filename+"_v_sftp"+file_extension
        # Upload the file from the remoteFilePath  to localFilePath
        srv.get(remoteFilePath, localFilePath)
    # Closes the connection
    srv.close()
```

Figure 3.9: Le code source de la récupération de fichier log en utilisant le sftp.

La collection à partir d'un path local de fichier

La figure suivante montre le code source de la récupération de fichier log en utilisant le **sftp**.

```
# if the form has been submitted by user is file_logfile_localpath
if request.method == 'POST' and 'file_logfile_localpath' in request.POST:
    obj = get_object_or_404(DirectoryFiles,filename = 'logfile')
    try:
        # open file in read mode
        with open(obj.local_file_path,'r', encoding='ISO-8859-1') as file:
```

Figure 3.10: Le code source de la récupération de fichier log en utilisant le path local de fichier.

Après la réception de fichier par l'application, une prochaine étape est se fait pour ce fichier qui est la normalisation.

3.5.4 Normalisation des événements

Après une collection d'un fichier Log, le système va automatiquement lire tout le fichier, il sépare les lignes pour que chaque événement va être traiter

Enrichissement d'événement

Une étape d'enrichissement se fait pour enrichir les événements avec des informations utiles provenant de sources externes qui ne sont pas disponibles dans l'événement d'origine. On a basé pour enrichir nos événements sur les éléments suivant:

- **Enrichir les données du fichier log**

  ► **Le nom d'analyste :** dans le cas où l'analyste importe le fichier journal, une information est ajoutée qui contient le nom d'utilisateur de notre application qui a téléchargé le fichier au notre système.

  ► **Le nombre d'événements :** Chaque fichier importer via le système ou l'analyste est contient un ensemble des événements, le nombre de ces événements est considéré comme une information d'enrichissement.

  ► **L'adresse ip et le nom de serveur :** Ces informations sont récupérées lors de l'importation du fichier journal.

  ► **La date et le temps de collecte :** Cette information représente une capture de moment où la phase de collection de fichier journal débute.

- **Enrichir les données du réseau**

  ► **Le réseau :** en utilisant le champ ≪ *Network* ≫ de la table ≪ *Region* ≫ de base de données et l'adresse ip d'utilisateur du système bancaire, un test se fait en créant un objet ≪ `ip_network` ≫ pour vérifier si une adresse ip appartient aux réseaux reconnues de la banque.

  ► **La zone :** à partir d'un segment de réseau où l'utilisateur du système bancaire se connecte, on peut déduire la zone affecté à ce segment de réseau.

- **Enrichir les données du machine**

  ► **Nom de la machine:** La méthode GetnameInfo () effectue une recherche DNS pour récupérer la représentation textuelle correspondant à l'adresse IP. Lorsque la recherche ne parvient pas à obtenir la représentation textuelle, la représentation numérique est renvoyée(fig:3.11).

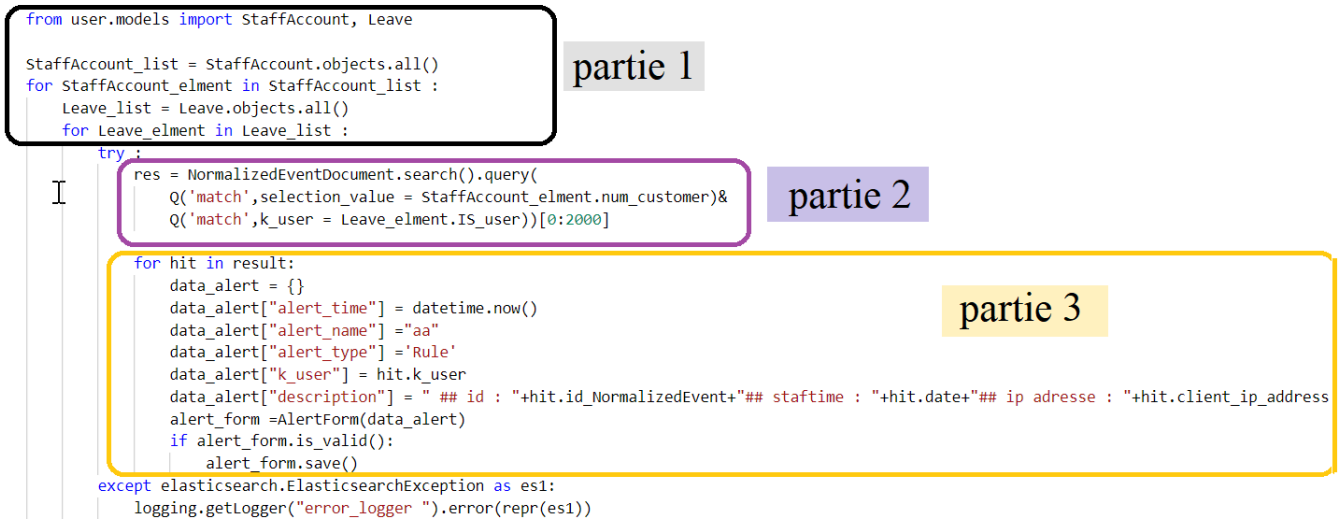
```
if adress_ip=="": #to deal with the problem of empty adress ip field in csv file
    adress_ip="0.0.0.0"
data_normalizedevent["system_computer_name"] = "NOCOMPUTERNAME"
nameinfo=socket.getnameinfo((adress_ip, 0), 0)
if nameinfo[0]==adress_ip :
    data_normalizedevent["system_computer_name"] = "NOCOMPUTERNAME"
else:
    data_normalizedevent["system_computer_name"] =nameinfo[0]
```

Figure 3.11: Obtenir le nom de la machine.

3.5.5 Analyse d'événements

Pour l'analyse d'événements, j'ai implémenté une fonction pour générer puis exécuter les règles d'analyse.

La partie de création des règles d'analyse sera divisée en trois parties(fig:3.12).




```
from user.models import StaffAccount, Leave

StaffAccount_list = StaffAccount.objects.all()
for StaffAccount_elment in StaffAccount_list :
    Leave_list = Leave.objects.all()
    for Leave_elment in Leave_list :
        try :
            res = NormalizedEventDocument.search().query(
                Q('match',selection_value = StaffAccount_elment.num_customer)&
                Q('match',k_user = Leave_elment.IS_user))[0:2000]

            for hit in result:
                data_alert = {}
                data_alert["alert_time"] = datetime.now()
                data_alert["alert_name"] ="aa"
                data_alert["alert_type"] ='Rule'
                data_alert["k_user"] = hit.k_user
                data_alert["description"] = " ## id : "+hit.id_NormalizedEvent+"## staftime : "+hit.date+"## ip adresse : "+hit.client_ip_address
                alert_form =AlertForm(data_alert)
                if alert_form.is_valid():
                    alert_form.save()
        except elasticsearch.ElasticsearchException as es1:
            logging.getLogger("error_logger ").error(repr(es1))
```

Figure 3.12: Exemple d'une règle d'analyse.

Tel que:

- **Partie 1:**pour importer les classes de modèle, récupérer les données des tables externes et accéder aux éléments de ces tables (les boucles).
- **Partie 3:** pour récupérer le résultat de la requête Elasticsearch.
- **Partie 3:** pour sauvegarder les alertes dans la base de données.

La figure suivante montre une partie de cette fonction concernée par la création de la première partie de la règle d'analyse:

```
    # For each rule in rules_list
    for r in rules_list :
        # Define a string for the "reused part" of code
        reused_code_string = ""
        # Define a string for the "loop part" of code
        loop_string = ""
        # Get the names of the external tables used in this rule
        external_table_list = ExternalTables.objects.filter(rule=r)
        # Define a counter. it will be used as a tabs that be left on the  next line of  "loop part"
        cmp_external_table_list = 0
        # If external_table_list is not none
        if external_table_list :
            for etp in external_table_list :
                if cmp_external_table_list == 0 :
                    reused_code_string += "\nfrom user.models import "+ str(etp.table_name)
                    cmp_external_table_list +=1
                else :
                    reused_code_string += ", "+ str(etp.table_name)
                    cmp_external_table_list +=1
                for i in range(cmp_external_table_list-1):
                    loop_string += "\t"
                loop_string += str(etp.table_name)+"_list = "+str(etp.table_name)+".objects.all()"+"\n"
                for i in range(cmp_external_table_list-1):
                    loop_string += "\t"
                loop_string += "for "+str(etp.table_name)+"_elment in "+str(etp.table_name)+"_list :"+"\n"
            reused_code_string +="\n\n"

```

Figure 3.13: Création de la première partie de la règle d'analyse.

La figure ci-dessous montre la partie qui permet d'écrire la requête DSL Elasticsearch (la deuxième partie de la règle d'analyse).

```
49          string_rule = "NormalizedEventDocument.search()"
50          clause_type_list =ClauseType.objects.filter(rule=r)
51          print(clause_type_list)
52          for  ct in clause_type_list :
53              if ct.type == "filter" or ct.type =="query":
54                  string_rule += "."+str(ct.type)+"("
55                  clause_attribute =ClauseAttribute.objects.filter(clause_type=ct)
56                  for ca in clause_attribute :
57                      if ca.attribute == "time" :
58                          value = "datetime.strptime('"+ca.value+"','%Y-%m-%d %H:%M:%S.%f')"
59                      elif ca.attribute == "date":
60                          value = "datetime.strptime('"+ca.value+"','%Y-%m-%d').date()"
61                      else:
62                          value = ca.value
63                      # for external table
64                      if ca.values_table is not None :
65                          value =str(ca.values_table.table_name)+"_elment."+ str(ca.value)
66                          if ca.attribute == "time" :
67                              value = "datetime.strptime(str("+value+"),'%Y-%m-%d %H:%M:%S.%f')"
68                          elif ca.attribute == "date":
69                              value = "datetime.strptime(str("+value+"),'%Y-%m-%d').date()"
70                      if ca.operator =='=' :
71                          if  ca.attribute == "time" or ca.attribute == "date" :
72                              string_rule += "Q('match',"+str(ca.attribute)+" = "+value+")"
73                          elif ca.values_table is not None :
74                              string_rule += "Q('match',"+str(ca.attribute)+" = "+value+")"
75                          else:
76                              string_rule += "Q('match',"+str(ca.attribute)+" = '"+value+"')"
```

Figure 3.14: la partie de création d'une raquette Elasticsearch DSL .

3.6 Tests et résultats:

On va illustrer dans cette section les diverses interfaces de notre application avec quelques tests effectués.

3.6.1 Espace administrateur

Cet espace est destiné à un administrateur qui est un super utilisateur qui dispose des privilèges spéciaux nécessaires pour administrer et maintenir l'application.

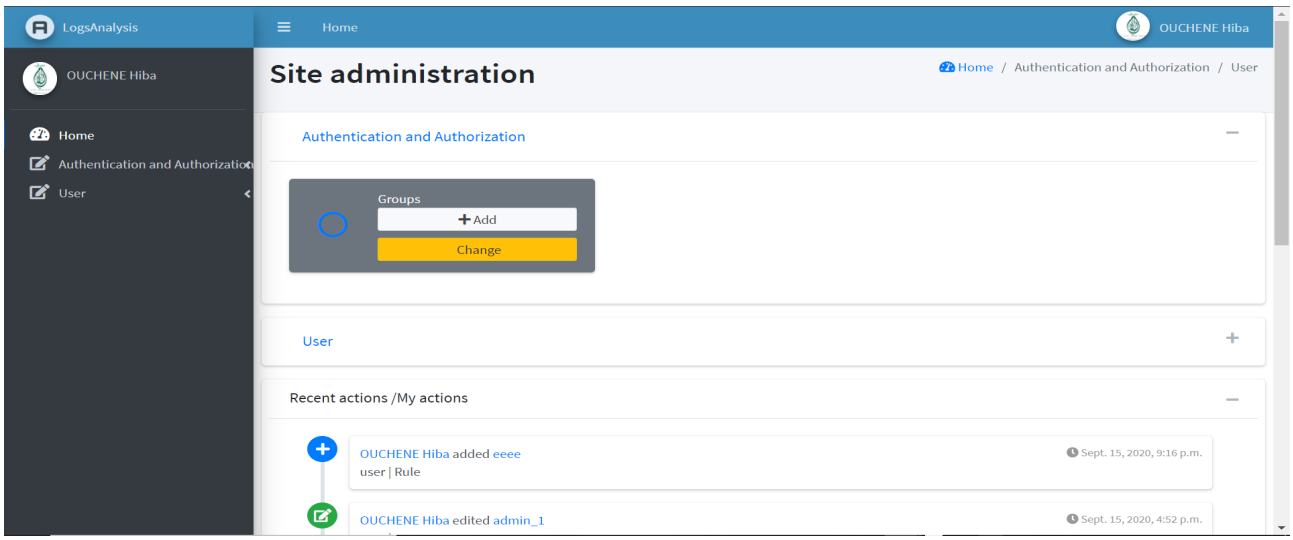


Figure 3.15: Espace administrateur(Superuser) .

Un administrateur peut faire plusieurs actions qui sont catégorisés comme suite:

a. Authentification et autorisation

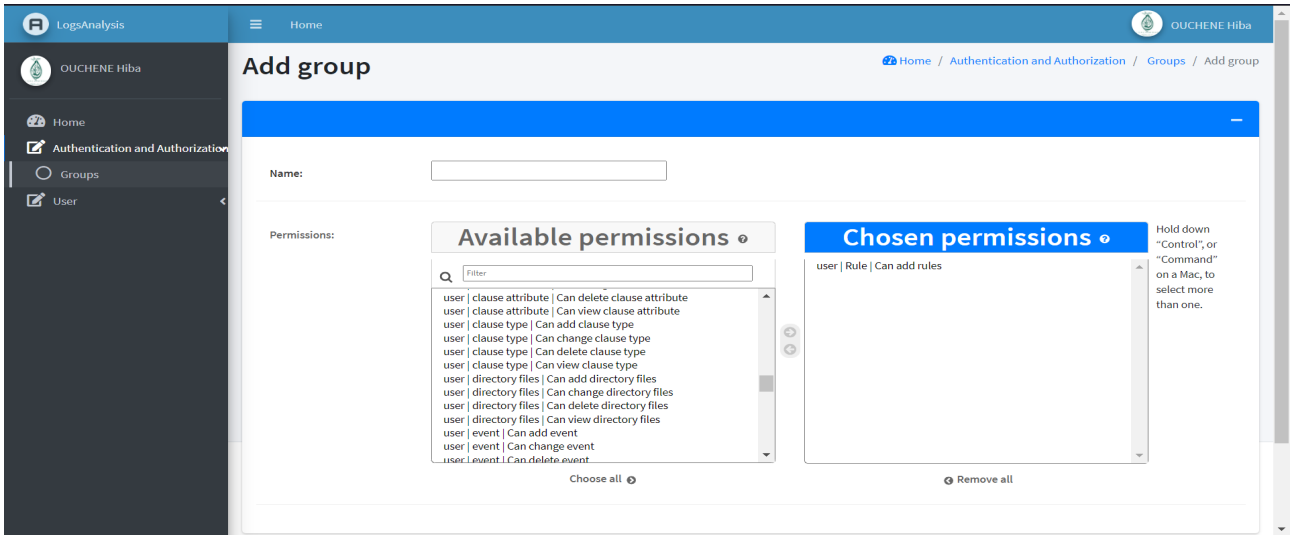


Figure 3.16: Créer un groupe de permission .

b. Tables de base de données d'application

Dans cette partie, l'administrateur peut faire plusieurs opérations sur des tables de base de données qui sont spécifiées lors de développement d'application.

► Table Utilisateur

L'administrateur peut gérer les utilisateurs, leurs groupes et permissions (ajouter, modifier, supprimer, bloquer...)comme représenté dans les figures suivantes :

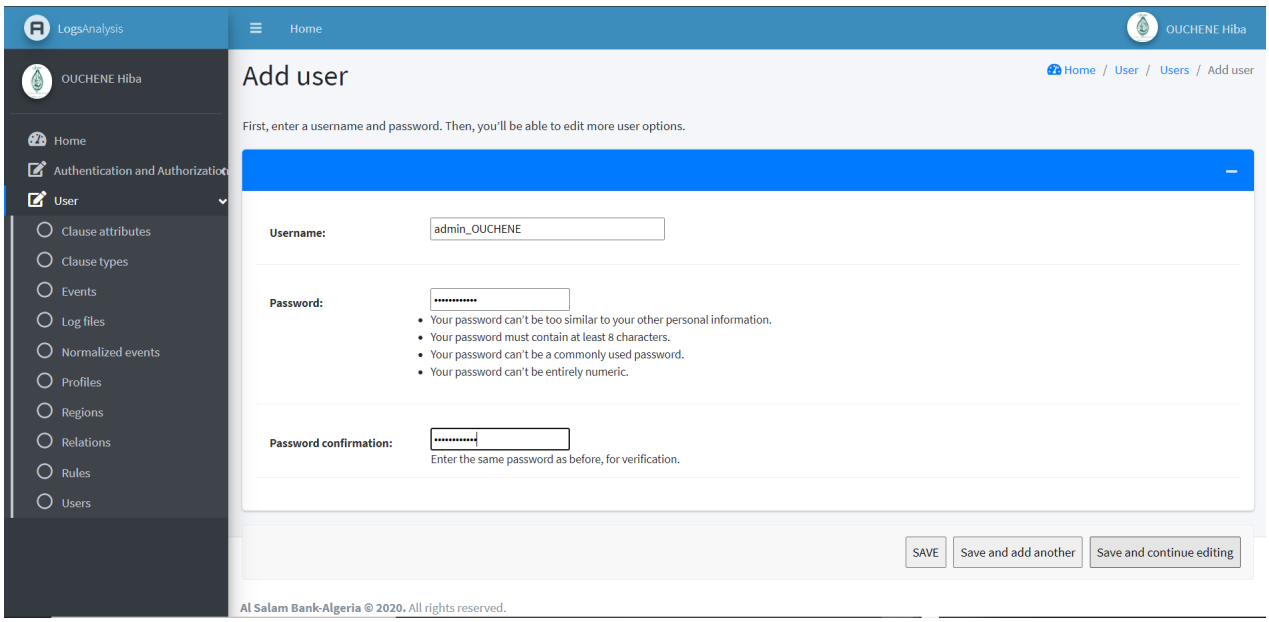


Figure 3.17: Ajouter un utilisateur.

Après l'enregistrement de ce utilisateur, on contenue la modification de ces informations qui se fait sur trois niveaux:

- informations d'authentification
- informations personnel
- permissions
- des dates importantes

Un bouton "HISTORY"offre la possibilité de consulter l'historique des modifications apportées à cet utilisateur.

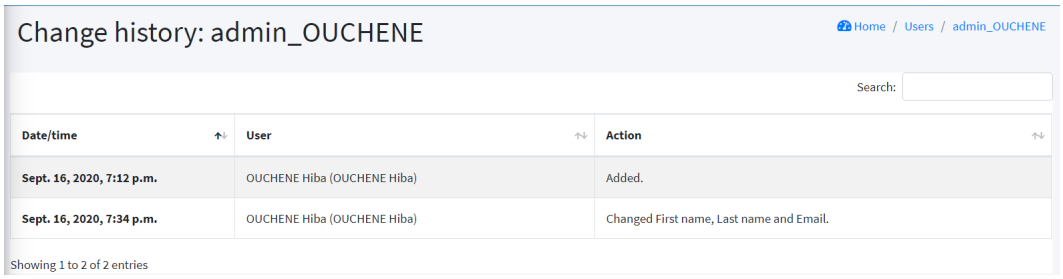


Figure 3.18: Consulter l'historique des changements.

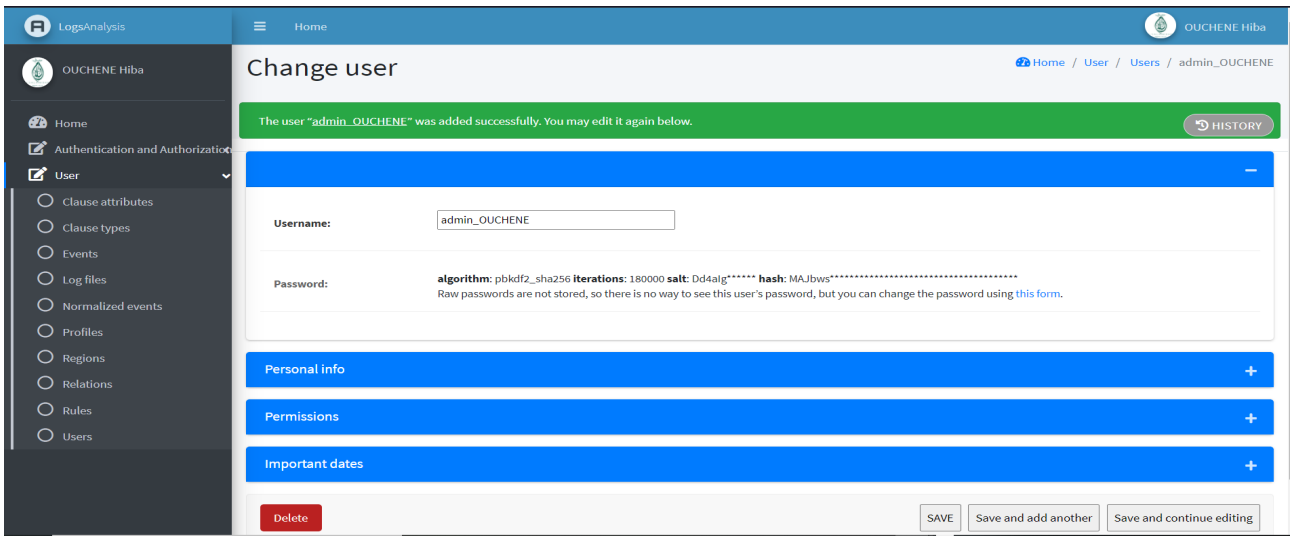


Figure 3.19: Ajouter un utilisateur.

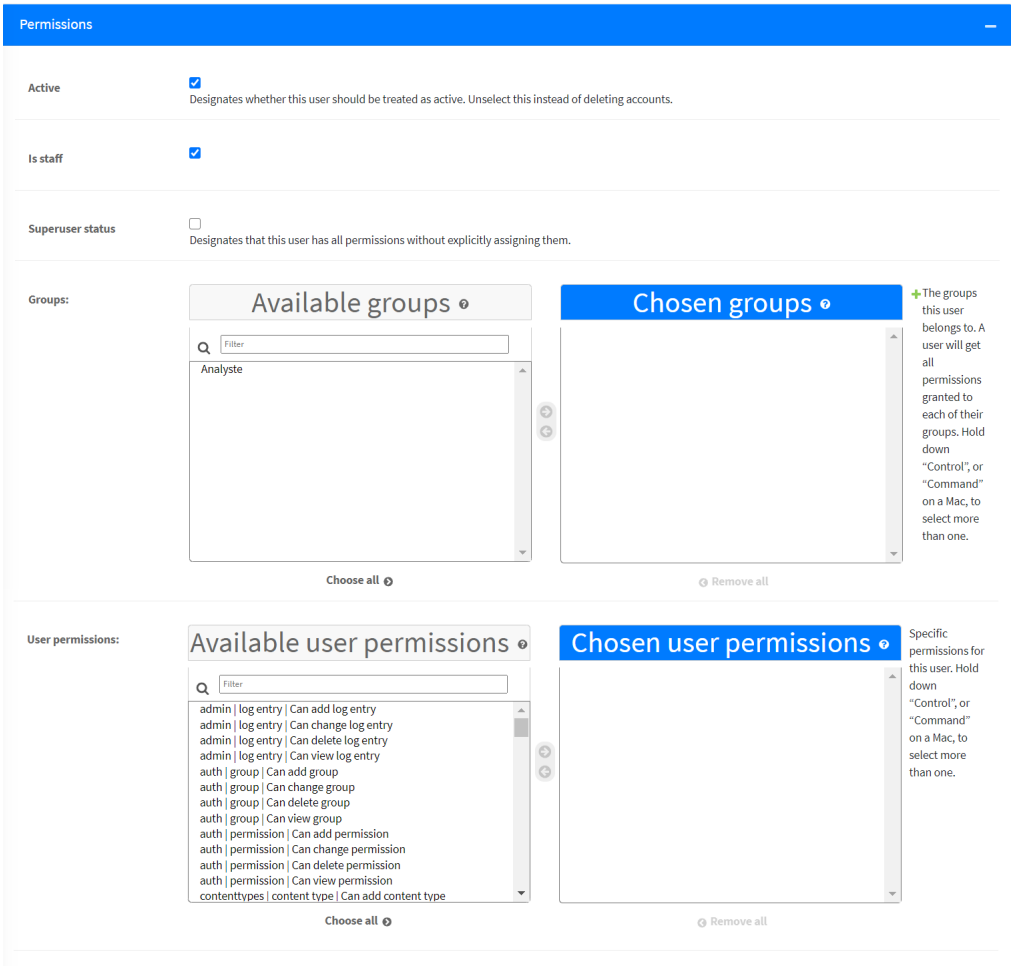


Figure 3.20: Gérer les permissions d'un utilisateur.

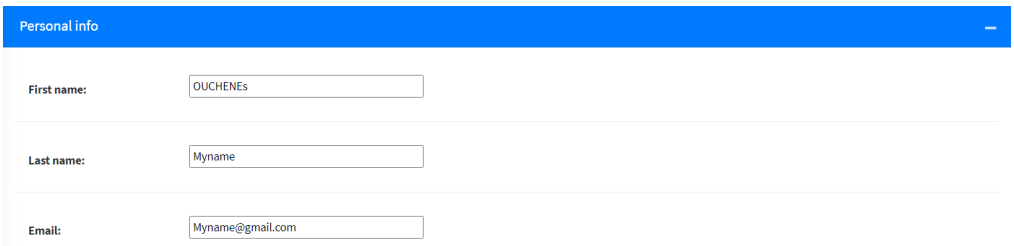


Figure 3.21: changer les informations personnels.

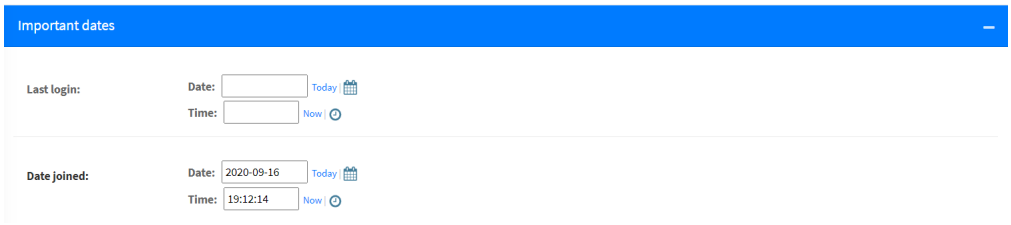


Figure 3.22: Les dates importants d'un utilisateur.

► Des autres tables

Plusieurs actions sont possible sur quelques tables spécifies lors de développement pour êtres gérer par l'administrateur d'application sont situées dans la figure suivante:

Figure 3.23: Quelques tables qui peuvent être modifiées par l'administrateur.

Des recherches,des filtres, des suppressions, des modifications et des créations sont autorisés sur ces tables. Quelques exemples de ces opérations illustrés dans les figures ci-dessus.

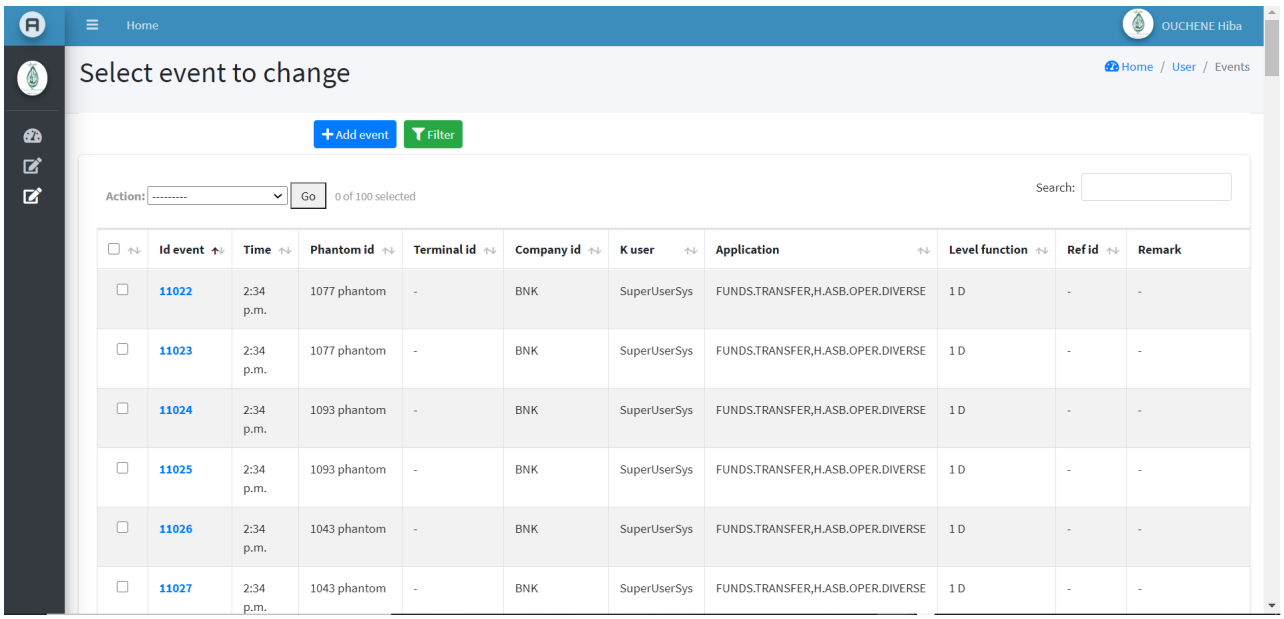


Figure 3.24: Prise d'écran de la table Event.

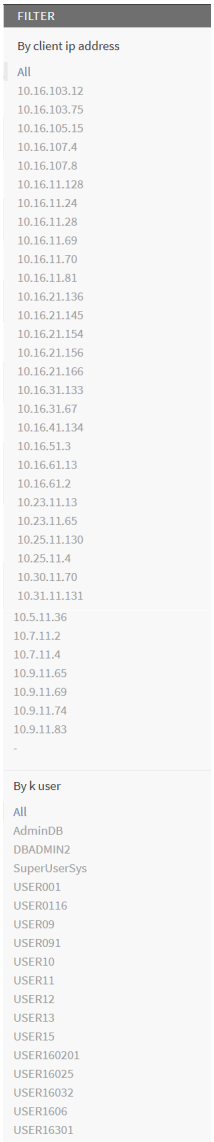


Figure 3.25: les filtres de la table Event.

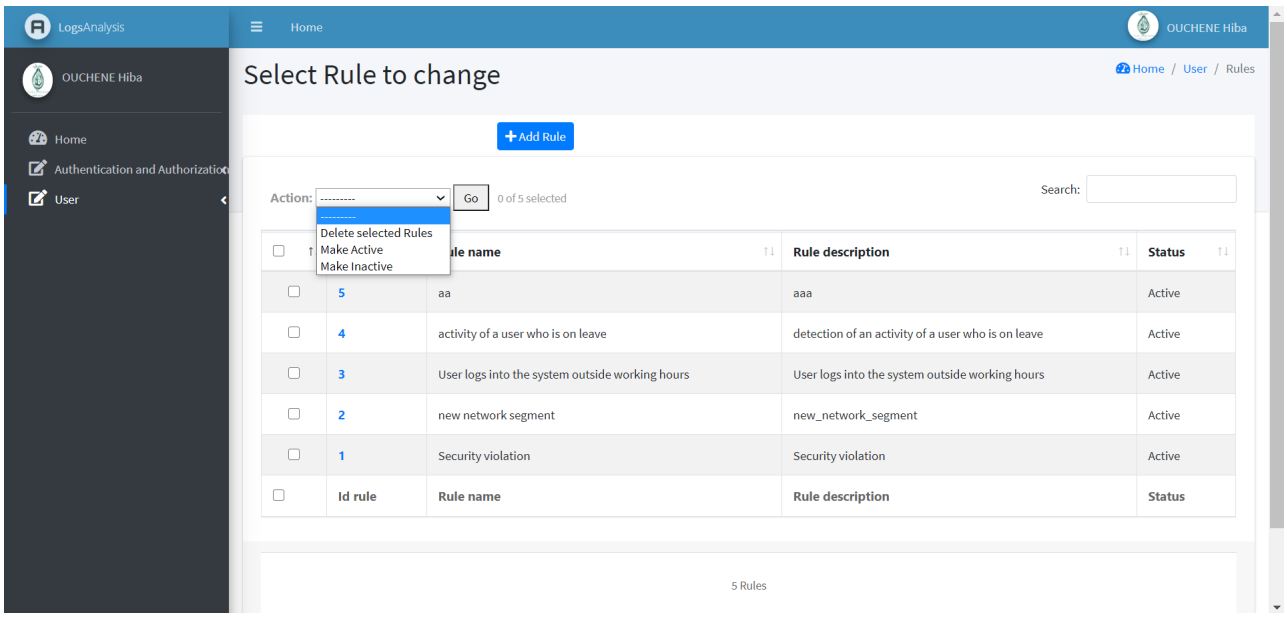


Figure 3.26: Les actions autorisés sur la table Rule.

3.6.2 Espace Utilisateur:

Cet espace rassemble tous les utilisateurs sur les mêmes permissions et les mêmes interfaces, afin qu'ils puissent exécuter différentes fonctions applicatives.

Dans cette partie on va illustrer les diverses fonctionnalités de notre application on fournissant des prises de vue de chaque fonctionnalité implémentée à partir de l'interface de notre application.

► **Consulter le tableau de bord**

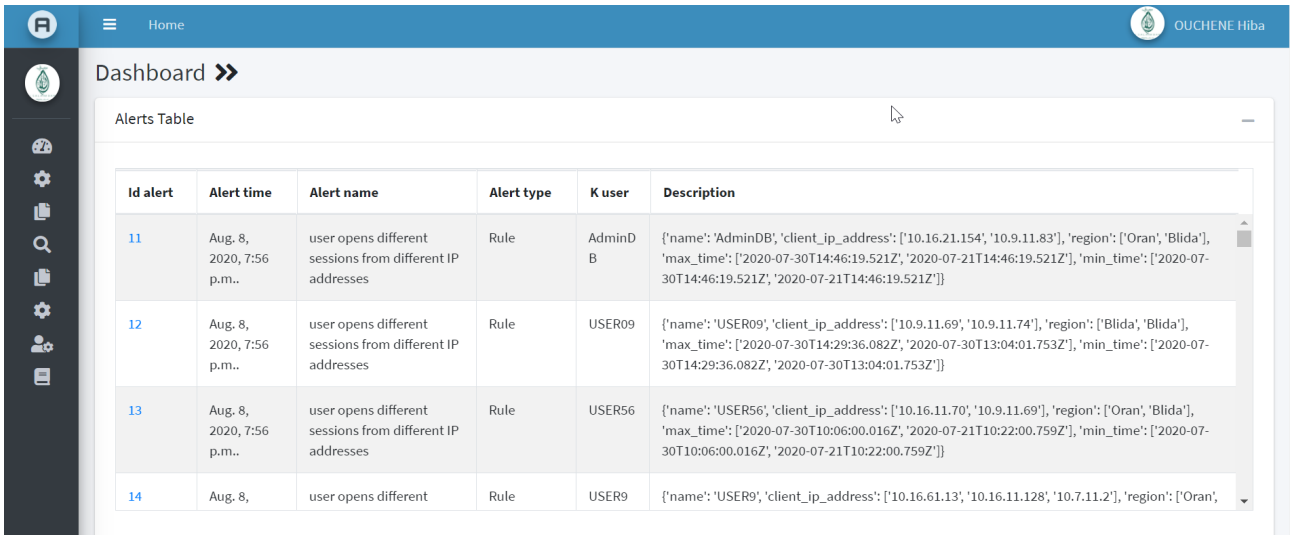


Figure 3.27: Les paramètres d'application.

La figure ci-dessus montre l'interface principale de l'application qui visualise les alertes qui ont été déclenché par l'analyseur d'événements.

► **Récupération de fichier log**

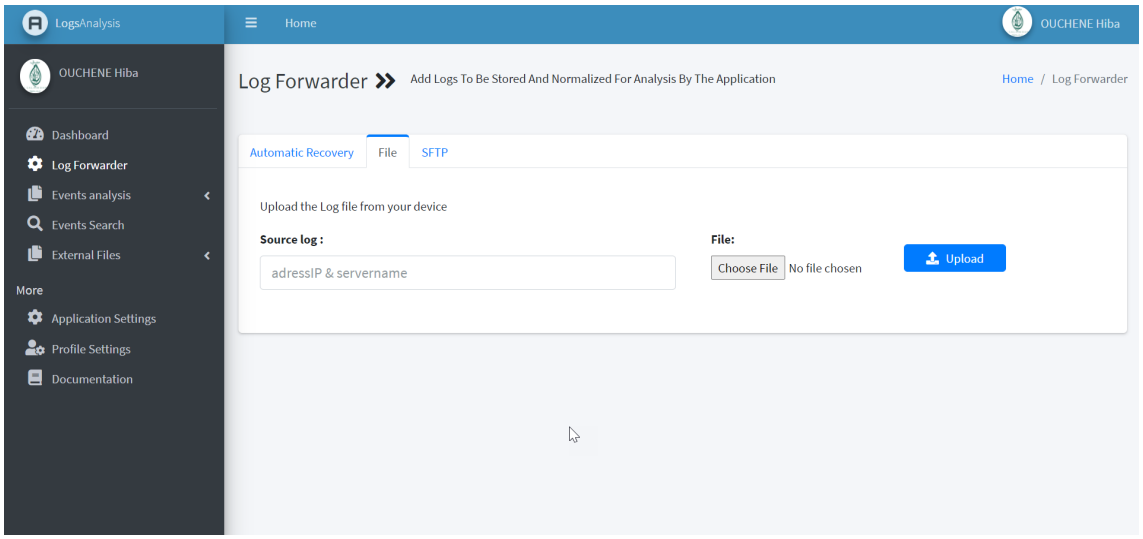


Figure 3.28: Une capture d'écran de l'interface de récupération de log.

La figure ci-dessus montre l'interface qui va permettre à l'utilisateur de l'application de récupérer le fichier log et l'envoyer vers le normalisateur d'événements. L'utilisateur doit choisir l'une des méthodes de récupération parmi les quatre disponibles (Fichier, SFTP, Fichier automatique,SFTP automatique).

► **Fichiers Externes**

Cette interface permet à l'utilisateur de récupérer les fichiers Congé.csv, CompteStaff.csv et ProfilUserSI.csv, comme apparu dans la figure suivante:

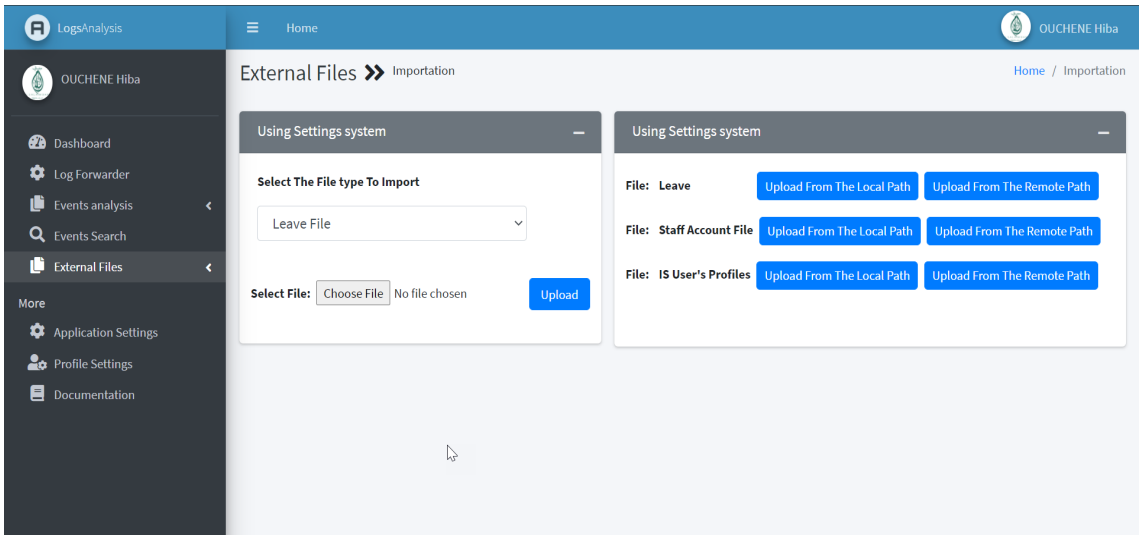


Figure 3.29: Une capture d'écran de l'interface des fichiers externes.

► **Les paramètres d'application**

La figure suivante montre l'interface des paramètres de l'application qui permet de modifier les informations de tables externes:

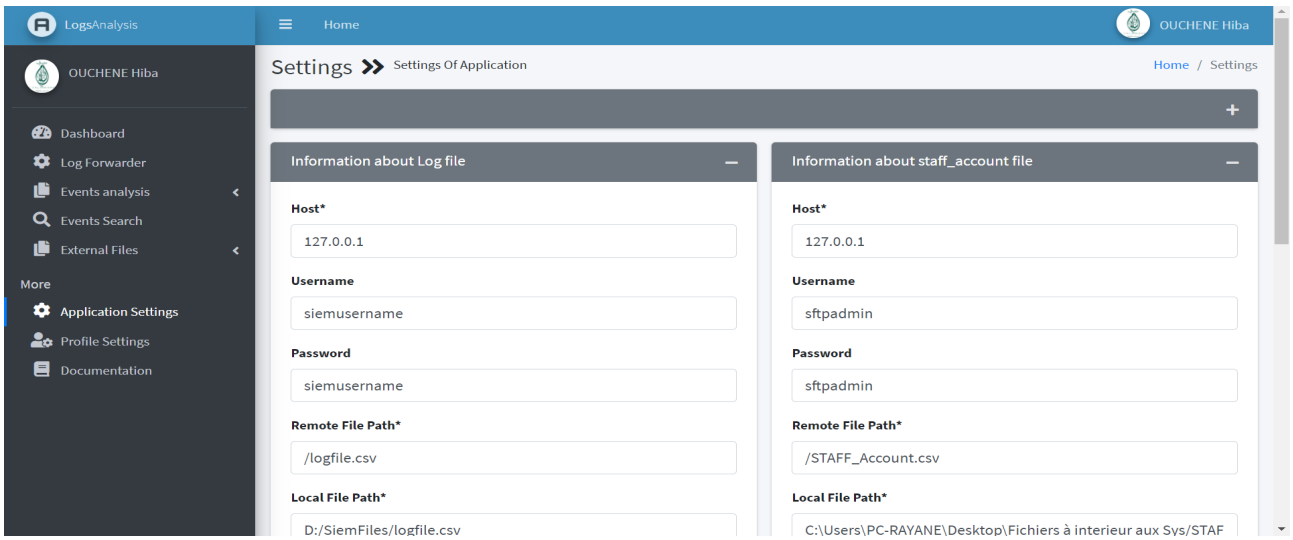


Figure 3.30: Une capture d'écran de l'interface des paramètres d'application.

► **Profile utilisateur**

Possibilité de consulter le profil de l'utilisateur pour modifier les informations présentes dans la figure (fig: 3.31).

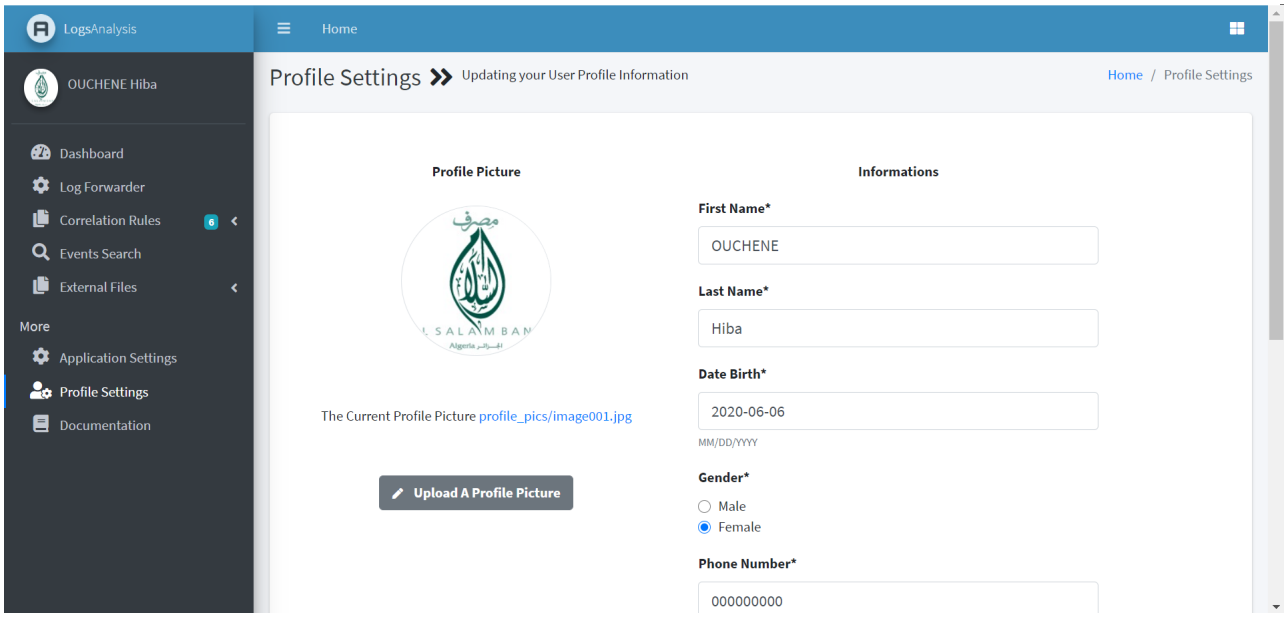


Figure 3.31: Une capture d'écran de l'interface de profile d'un utilisateur.

3.6.3 Test

Après avoir implémenté notre solution, on procéde à des tests de quelques scénarios afin de vérifier l'efficacité et le bon fonctionnement de l'analyseur d'événements.

► **Les violations de la politique de sécurité (Security violation)**

Une violation de sécurité est définit comme suite:

If classification is equal to ''SECURITY'' and the remark is not empty then trigger an alert.

tel que:

- le champ **«remark»** est utilisé lorsque des violations de sécurité sont enregistrées, pour expliquer pourquoi le système n'autorise pas l'activité tentée.
- le champ **«classification»** est égal à 'SECURITY' lorsqu'une action liée à la sécurité est enregistrée, cette action peut être une violation de sécurité.

La figure suivante illustre une prise d'écran de la saisie des informations de règle d'analyse 'Violation de sécurité' dans les tables: Règle, TypeClause, AttributsClause, Relation. Cette figure est suivie par la figure (3.33), qui montre la fonction écrite par le générateur de règles d'analyse.

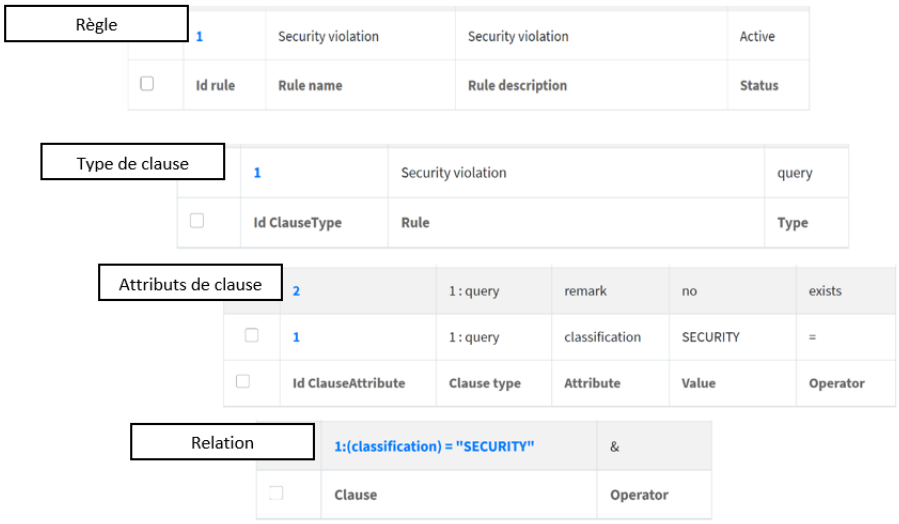


Figure 3.32: Prise d'écran des tables pour la règle d'analyse 'Violation de sécurité'.

```
# 1 : Security violation
try:
    result = NormalizedEventDocument.search().query(
        Q('match', classification='SECURITY') & Q('exists', field='remark'))[0:2000]

    for hit in result:
        data_alert = {}
        data_alert["alert_time"] = datetime.now()
        data_alert["alert_name"] ="Security violation"
        data_alert["alert_type"] ='Rule'
        data_alert["k_user"] = hit.k_user
        data_alert["description"] = " ## id : "+hit.id_NormalizedEvent+"## staftime : "+hit.date+"## ip adresse : "+hit.client_ip_address
        alert_form =AlertForm(data_alert)
        if alert_form.is_valid():
            alert_form.save()

except elasticsearch.ElasticsearchException as es1:
    logging.getLogger("error_logger ").error(repr(es1))
```

Figure 3.33: La fonction de la régle d'analyse 'Violation de sécurité' écrite par le générateur de règles d'analyse.

► Ouverture d'une session à partir d'un nouveau segment réseau

"Ouvrir une session à partir d'un nouveau segment de réseau" est définit par l'action de connexion d'un utilisateur au système avec une adresse IP qui n'appartient pas aux réseaux que le système connaît, il est donc définit par l'expression suivante:

If source_region is equal to ''NORegion'' and application is equal to ''SIGN.ON'' then trigger an alert

La valeur "NORegion" signifie que ce réseau n'a pas été reconnu par le système, qui implique que ce réseau n'existe pas. Cette valeur a été traitée lors de la phase de normalisation.
La figure suivante montre une prise d'écran de la saisie des informations de règle d'analyse 'Ouverture d'une session à partir d'un nouveau segment réseau' dans les tables: Règle, TypeClause, AttributsClause, Relation de base de données.

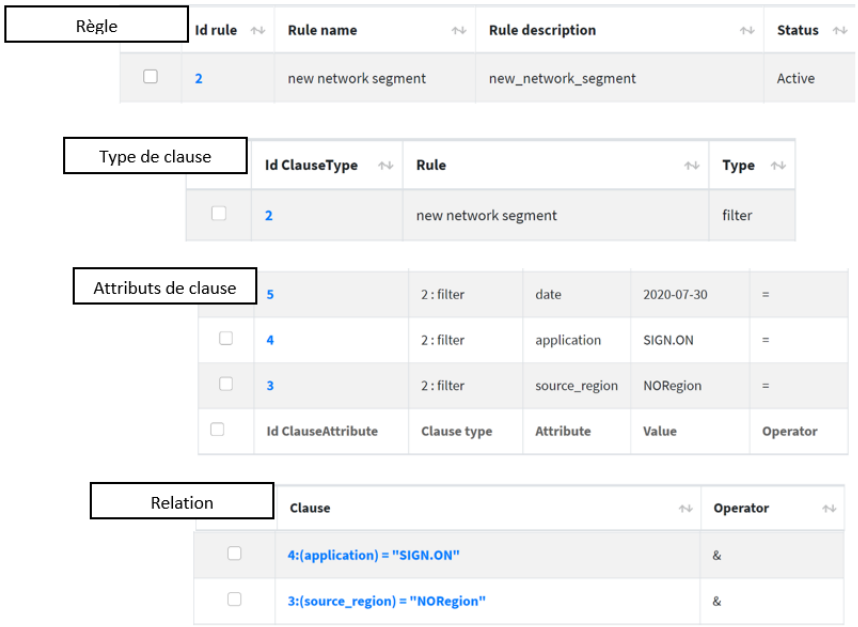


Figure 3.34: Prise d'écran des tables pour la règle d'analyse 'Ouverture d'une session à partir d'un nouveau segment réseau.

Le code de la détection de "l'ouverture d'une session à partir d'un nouveau segment réseau" écrit par le générateur de règles d'analyse est comme suite:

```
# 2 : new network segment
try :
    result = NormalizedEventDocument.search().filter(
        Q('match',source_region = 'NORegion')&
        Q('match',application = 'SIGN.ON')&
        Q('match',date = datetime.strptime('2020-07-30','%Y-%m-%d').date()))[0:2000]

    for hit in result:
        data_alert = {}
        data_alert["alert_time"] = datetime.now()
        data_alert["alert_name"] ="New network segment"
        data_alert["alert_type"] ='Rule'
        data_alert["k_user"] = hit.k_user
        data_alert["description"] = " ## id : "+hit.id_NormalizedEvent+"## staftime : "+hit.date+"## ip adresse : "+hit.client_ip_address
        alert_form =AlertForm(data_alert)
        if alert_form.is_valid():
            alert_form.save()
```

Figure 3.35: Ouverture d'une session à partir d'un nouveau segment réseau.

► Utilisation d'une session d'un utilisateur en congé

Il définit le cas où on trouve dans le fichier log un utilisateur en congé:

For i in congé_table:

If event.username == i.username and i.start_date<= event.date <= i.resumption_date then trigger an alert.

Les informations de règle seront saisies comme suit:

Règle

Id rule	Rule name	Rule description	Status
4	activity of a user who is on leave	detection of an activity of a user who is on leave	Active

Type de clause

Id ClauseType	Rule	Type
5	activity of a user who is on leave	filter

Attributs de clause

Id ClauseAttribute	Clause type	Attribute	Value	Operator	Values table
11	5 : filter	date	start_date	gt	Leave
10	5 : filter	date	resumption_date	lt	Leave
9	5 : filter	k_user	IS_user	=	Leave

Relation

Clause	Operator
10:(date) lt (Leave.resumption_date)	&
9:(k_user) = (Leave.IS_user)	&

Table externe

Id external table	Rule	Table_name
1	activity of a user who is on leave	Leave

Figure 3.36: Prise d'écran des tables pour la règle d'analyse 'Utilisation d'une session d'un utilisateur en congé.

Le résultat de la création de la règle 'Utilisation d'une session d'un utilisateur en congé' est illustré dans la figure suivante:

```python
# 4 : activity of a user who is on leave
from user.models import Leave

Leave_list = Leave.objects.all()
for Leave_elment in Leave_list :
    try :
        res = NormalizedEventDocument.search().filter(
            Q('match',k_user = Leave_elment.IS_user)&
            Q('range',date={'lt':datetime.strptime(str(Leave_elment.resumption_date),'%Y-%m-%d').date()})&
            Q('range',date={'gt':datetime.strptime(str(Leave_elment.start_date),'%Y-%m-%d').date()}))[0:2000]

        for hit in result:
            data_alert = {}
            data_alert["alert_time"] = datetime.now()
            data_alert["alert_name"] ="Activity of a user who is on leave"
            data_alert["alert_type"] ='Rule'
            data_alert["k_user"] = hit.k_user
            data_alert["description"] = " ## id : "+hit.id_NormalizedEvent+"## staftime : "+hit.date+"## ip adresse : "+hit.client_ip_address
            alert_form =AlertForm(data_alert)
            if alert_form.is_valid():
                alert_form.save()
        except elasticsearch.ElasticsearchException as es1:
            logging.getLogger("error_logger ").error(repr(es1))
```

Figure 3.37: Utilisation d'une session d'un utilisateur en congé.

3.7 Conclusion

Dans ce dernier chapitre, on a présenté les différents outils utilisés pendant le développement de ce projet,ainsi que comment ceux-ci ont été utilisés pour implémenter les différents composants de l'application.On a ensuite présenté les différentes fonctionnalités de cette application, illustré le tout avec des prises d'écran.

Conclusion

Ce projet de fin d'études porte sur la conception et le développement d'une application web pour la gestion des évènements et des informations de sécurité pour Al Salam Bank afin de surveiller les actions des utilisateurs de son système d'information. Le présent manuscrit détaille toutes les étapes par lesquelles je suis passée pour arriver au résultat attendu.

En prenant compte de ce qui a précédé, on peut dire que les objectifs fixés au début de ce projet ont été atteints. J'ai conçu un système qui permet de Récupérer un fichier logs, normaliser, stocker, indexer, analyser et superviser les événements de ce fichier en générant des alertes dès qu'une condition sur un ou plusieurs événements est réalisée. Ces alertes sont basées sur des règles d'analyse définies par l'utilisateur qui permettent de différencier et de détecter les actions anormales effectuées. Une fois ces actions sont détectées, un stockage se fait au niveau de la table de la base de données d'alertes. Les alertes stockées seront affichées afin que l'utilisateur de l'application puisse les prendre directement en charge.

Ce travail a été développé en utilisant le framework Django qui offre une facilité et une simplicité en termes de développement d'applications sécurisées.

J'ai également utilisé le moteur de recherche le plus populaire Elasticsearch connu pour son indexation rapide et sa facilité de déploiement et d'utilisation.

Enfin, en termes de perspective on peut noter qu'il est possible d'améliorer:

- Ajouter ou supprimer certains modules ou interfaces pour une meilleure adoption.
- Fusionner les deux interfaces (administrateur et utilisateur) en une seule contenant des modules qui diffèrent d'un utilisateur à l'autre.
- Ajouter plus de fonctionnalités, telles que le suivi des actions d'un utilisateur SI qui a effectué une action suspecte.
- Enrichir le tableau de bord avec des courbes et des graphes sur les données à visualiser , sur la consommation de ressources CPU et l'utilisation de la RAM, des indicateurs de vitesse de collecte , d'indexation...
- Ajouter une documentation afin d'expliquer comment utiliser cette application par l'utilisateur.

Ce travail a été l'occasion pour moi en tant qu'étudiante en Sécurité des Systèmes d'information de voir concrètement une partie du monde professionnel et de mettre en œuvre les connaissances théoriques acquises durant mon cursus.

Bibliography

[1] C. Bohic, "Siem : ce que gartner reproche aux 7 leaders de son magic quadrant 2020," *https://www.silicon.fr/siem-gartner-magic-quadrant-2020-334500.html*, 24 février 2020.

[2] A. Chuvakin, "The complete guide to log and event management," *White Paper*, 2010.

[3] K. Kent and M. Souppaya, "Guide to computer security log management," *NIST special publication*, vol. 92, pp. 1–72, 2006.

[4] S. Al-Fedaghi and F. Mahdi, "Events classification in log audit," *Int J Netw Secur Appl (IJNSA)*, vol. 2, no. 2, pp. 58–73, 2010.

[5] B. SIX, "Log formats – a (mostly) complete guide," 2020.

[6] M. Rouse, "Security information and event management (siem). techtarget search security," February 2020.

[7] M. K. Pratt, "What is siem software? how it works and how to choose the right tool," *csoonline.com*, 2017.

[8] S. Ganapathy, "The absolute guide to siem," 2018.

[9] J.-P. Lang. (2006-2017) Aperçu, prelude oss project. [Online]. Available: https://www.prelude-siem.org/#Prelude-OSS-project

[10] SIEMONSTER, *https://siemonster.com/*, 2019.

[11] B.V and Elasticsearch, "What is the elk stack? why, it's the elastic stack." *https://www.elastic.co/what-is/elk-stack*, 2020.

[12] T. Graylog, "Graylog," *https://www.capterra.fr/software/183539/graylog*.

[13] I. a. i. a. Gartner, "Gartner magic quadrant," *https://www.gartner.com/en/research/methodologies/magic-quadrants-research*, 2020.

[14] S. Vijay, Ruslan Desyatnikov, "Top 11 best siem tools in 2020 for real-time incident response and security." *https://www.softwaretestinghelp.com/siem-tools/*, 2020.

[15] VanRoey.be, "Exabeam," *https://www.vanroey.be/fr/caution/ex-faiseur/*.

[16] www.capterra.fr, "Outils siem (security information event management)," *https://www.capterra.fr/directory/31239/siem/software*, 2020.

[17] A. Kingatua, "9 meilleurs outils de réponse aux incidents de sécurité pour les petites et les entreprises," *https://geekflare.com/fr/security-incident-response-tools/*, mai 27, 2020.

[18] R. S. L. or its affiliates, "Threat detection and response," *https://www.rsa.com/en-us/products/threat-detection-response*, 2020.

[19] L. SolarWinds Worldwide, "Security event manager," *https://www.solarwinds.com/security-event-manager*, 2020.

[20] FutureMarketInsights, "Security information and event management software market: Global industry analysis 2012 – 2016 and opportunity assessment; 2017 – 2027," *https://www.futuremarketinsights.com/reports/security-information-and-event-management-software-market*, 2020.

[21] MarketsandMarkets, "Security information and event management market by component, application, deployment mode, organization size, vertical (information, finance and insurance, healthcare and social assistance, utilities), and region - global forecast to 2025," *https://www.marketsandmarkets.com/Market-Reports/security-information-event-management-market-183343191.html*, 2020.

[22] wilfried Erisco, "DÉfinition des besoins fonctionnels et des besoins non fonctionnels," *https://www.memoireonline.com/02/09/1973/m-conception-et-developpement-dune-application-de-la-gestion-dune-bibliotheque5.html*.

[23] O. Capuozzo, "Cas d'utilisation, une introduction," *France, Editions CERTA*, 2004.

[24] "Celery 5.0.1 documentation," *https://docs.celeryproject.org/en/stable/*.

[25] "Elasticsearch," *https://www.elastic.co/fr/what-is/elasticsearch*, 2020.

[26] "Elasticsearch : Tout ce que vous devez savoir sur elasticsearch," *http://www.mupmag.fr/elasticsearch/*, 2020.

[27] "What is redis?" *https://aws.amazon.com/redis/*, 2020.

[28] LinuxLinks, "Modelio – open source uml and bpmn modeling environment," *https://www.linuxlinks.com/modelio-open-source-uml-and-bpmn-modeling-environment/*.

[29] F. Martinig, "Staruml - open source uml tool," *http://www.methodsandtools.com/tools/staruml.php*.

[30] "Edraw max," *https://www.clubic.com/telecharger-fiche54328-edraw-max.html*.

[31] V. Paradigm., "What is unified modeling language (uml)?" *https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what-is-uml/*, 2020.

[32] C. SAS., "Télécharger sublime text pour windows," *https://www.clubic.com/telecharger-fiche430809-sublime-text.html*, 2020.

[33] "Django introduction," *https://developer.mozilla.org/en-US/docs/Learn/Server-side/Django/Introduction*.

[34] Microsoft., "Visual studio code," *https://code.visualstudio.com/docs*, 2020.

[35] "Html5 (hypertext markup language 5) : définition de ce langage informatique," *https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203257-html5-hypertext-markup-langage5-definition-traduction/*, 14-02-2019.

[36] "Css (cascading style sheets) : définition, traductione," *https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203277-css-cascading-style-sheets-definition-traduction/*, 10-01-2019.

[37] S. Morris, "Tech 101: What is javascript?" *http://www.mupmag.fr/elasticsearch/*.

[38] "jquery : définition simple," *https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203587-jquery-definition/*, 08-01-2019.

[39] "Bootstrap : définition, tutoriels, astuces, pratiques," *https://www.journaldunet.com/web-tech/developpeur/1159810-bootstrap-definition-tutoriels-astuces-pratiques/*, 28-08-2019.

[40] "What is python? executive summary," *https://www.python.org/doc/essays/blurb/*, 2001-2020.

[41] "Sqlite," *https://fr.wikipedia.org/wiki/SQLite*.