RÉPUBLIQUE DU BÉNIN
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ D'ABOMEY-CALAVI

INSTITUT DE FORMATION ET DE RECHERCHE EN INFORMATIQUE

BP 526 Cotonou Tel : +229 01 55 02 80 70/01 53 97 30 60
https://www.ifri-uac.bj Courriel : contact@ifri.uac.bj

MÉMOIRE

pour l'obtention du

Diplôme de Licence en Informatique

Option : Sécurité Informatique

Présenté par :
Andromède Céleste Mahoussé **KASSA**

Mise en place d'un système de surveillance des accès de production en télétravail: cas du Port Autonome de Cotonou

Sous la supervision :
Ing Mawuénan Hervé Guy **AKAKPO**

Membres du jury :

SANYA Max Fréjus	Docteur	Président
GOHOUE Rose	Ingénieure	Examinatrice
AKAKPO Mawuénan Hervé Guy	Ingénieur	Rapporteur

Année Académique : **2023-2024**

Sommaire

Dédicace ii

Remerciements iii

Résumé iv

Abstract v

Table des figures vi

Liste des tableaux viii

Liste des acronymes ix

Glossaire x

Introduction 2

1 Revue de littérature 4

2 Choix et mise en place de la solution 10

3 Tests et Résultats 29

Conclusion 39

Bibliographie 40

Webographie 41

Table des matières 42

Dédicace

A
Ma mère Jocelyne ADJOVI
Mon père Hervé KASSA
Mon frère Alvin KASSA
Tous mes proches

Remerciements

La réalisation de ce mémoire a été possible grâce au soutien et à l'accompagnement de plusieurs personnes à qui nous voulons exprimer toute notre gratitude. Nous tenons ainsi à remercier sincèrement :

- Le professeur Eugène C. EZIN, Directeur de l'Institut de Formation et de Recherche en Informatique (IFRI), pour son engagement envers la formation des étudiants ;
- M. Hervé AKAKPO, Chef du Département des Systèmes d'Informations de la Caisse Autonome de Gestion de la Dette du Bénin (CAGD), mon maître de mémoire, pour son accompagnement, ses conseils avisés et sa disponibilité tout au long de ce travail ;
- M. Jospy GOUDALO, agent du Département des Systèmes d'Informations de la CAGD, assistant de mon maître mémoire pour son aide précieuse et son suivi tout au long de ce travail ;
- M. Symphorien DOFFON, le Responsable de la Sécurité des Systèmes d'Informations (RSSI) du Port Autonome de Cotonou (PAC), mon maître de stage, qui m'a accueillie et guidée tout le long de ce stage avec bienveillance dans l'exécution de ce projet ;
- L'ensemble du corps professoral de l'IFRI, dont l'enseignement et les conseils pratiques ont servis de base pour l'acquisition des compétences nécessaires à la réalisation de ce travail ;
- Ma famille, mes amis pour leur encouragement continu, leur patience et leur soutien moral tout au long de cette aventure académique.

Que toutes les personnes ayant contribué de près ou de loin à l'aboutissement de ce mémoire trouvent ici l'expression de notre sincère reconnaissance.

Résumé

Avec l'expansion du télétravail, les entreprises autorisent les employés à accéder aux ressources internes en passant par Internet. Afin de garantir la sécurité des accès distants, plusieurs mesures de sécurité sont mises en place. Parmi celles-ci, il y a la surveillance et l'analyse des activités du réseau. Une méthode adéquate est l'utilisation d'un système de centralisation et d'analyse des journaux d'événements. Dans la suite de ce travail, le choix s'est porté sur Elastic Stack, pour la collecte, l'analyse, la visualisation de ces journaux d'événements ou logs provenant des systèmes critiques. Par ailleurs, une architecture a été mise en place dans le but d'évaluer les performances de l'outil. Les résultats obtenus de ces tests ont donné des informations pertinentes et ont démontré l'efficacité de la suite pour la surveillance des accès distants.

Mots clés : télétravail, mesures de sécurité, surveillance, logs, Elastic Stack.

Abstract

With the expansion of remote work, companies allow employees to access internal resources via the Internet. To ensure the security of remote access, several security measures are implemented. Among these, network activity monitoring and analysis are essential. An appropriate method is the use of a system for centralizing and analyzing event logs. In this work, we opted for the Elastic stack, for collecting, analyzing, and visualizing event logs from critical systems. We set up an architecture to evaluate the performance of the tool. The results obtained from these tests provided relevant information and demonstrated the effectiveness of the stack for remote access monitoring.

Key words: remote work, security measures, monitoring, logs, Elastic Stack.

Table des figures

2.1 Flux de données dans la suite ELK . . . 13
2.2 Architecture réseau . . . 15
2.3 Vérification du fonctionnement de Elasticsearch . . . 16
2.4 Demande du token dans Kibana . . . 17
2.5 Demande du code de vérification . . . 17
2.6 Affichage du statut de Kibana . . . 18
2.7 Finalisation de la configuration de Kibana . . . 18
2.8 Page d'accueil de Kibana . . . 19
2.9 Affichage du statut de Logstash . . . 19
2.10 Configuration de l'input filestream . . . 20
2.11 Configuration des inputs syslog et netflow . . . 21
2.12 Configuration de l'output de Filebeat . . . 21
2.13 Configuration du module auditd . . . 22
2.14 Configuration des modules system et file integrity . . . 23
2.15 Configuration de l'input de Logstash . . . 24
2.16 Configuration de l'output de Logstash . . . 24
2.17 Configuration pour l'envoi des mails . . . 25
2.18 Exemple de configuration du plugin throttle . . . 25
2.19 Configuration de l'adresse IP de Kibana . . . 26
2.20 Configuration de l'adresse IP de Elasticsearch . . . 26
2.21 Stack Management . . . 27
2.22 Emplacement de "Snapshot et Restore" . . . 27

3.1 Emplacement de l'index management . . . 29
3.2 Liste des indexs . . . 30
3.3 Interface des vues de données . . . 30
3.4 Création d'une vue de données . . . 31
3.5 Vue de données de Filebeat . . . 31
3.6 Vue de données d'Auditbeat . . . 32
3.7 Flux réseau d'un utilisateur distant . . . 32
3.8 Heures d'activité de l'utilisateur distant "*user*" . . . 33
3.9 Heures d'activité de l'utilisateur distant "*administrateur*" . . . 33
3.10 Dashboard . . . 34
3.11 Processus exécutés . . . 34
3.12 Terminal de l'utilisateur . . . 35
3.13 Alertes de l'IDS Snort pour l'attaque par brute force . . . 36
3.14 Alertes de l'IDS Snort pour le deni de service . . . 36

3.15 Mail d'alerte sur le système 37

Liste des tableaux

1.1 Avantages et Inconvénients de quelques SIEM 8

2.1 Critères de choix et méthodes d'évaluation 11
2.2 Grille comparative de quelques SIEM 12

Liste des acronymes

DNS :

Domain Name System 15,

ELK :

Elasticsearch Logstash Kibana 12, 13, 15, 26,

IDS :

Intrusion Dectection System vi, 15, 35–37,

IP :

Internet Protocol vi, 16, 26, 32, 34, 36,

IPFIX :

Internet Protocol Flow Information Export 20,

NAT :

Network Address Translation 39,

SFTP :

Secure Shell File Transfert Protocol 15,

SIEM :

Security Information and Event Management viii, 3, 6–12, 28, 39, 42,

SSH :

Secure Shell 15, 36,

TCP :

Transmission Control Protocol 20,

UDP :

User Datagram Protocol 20,

VPN :

Virtual Private Network 5, 15, 32, 35, 37,

Glossaire

DNS : Système qui traduit les noms de domaine, par exemple "www.example.com" en adresses IP, par exemple "192.168.1.1". Il permet d'accéder facilement aux sites web et services en ligne sans avoir à mémoriser des adresses IP complexes.

Framework d'Audit : Un ensemble de bonnes pratiques, de méthodes et d'outils pour l'évaluation de la conformité, la sécurité et la performance des systèmes d'information d'une organisation.

IDS : Système de détection d'intrusion qui surveille le trafic réseau ou les journaux d'événements pour identifier une possible anomalie.

IP : Protocole de communication utilisé pour l'identification et l'acheminement des données à travers un réseau, notamment sur Internet. Il définit les règles permettant aux appareils (ordinateurs, routeurs, serveurs) de s'identifier et de communiquer entre eux via des adresses IP uniques.

IPFIX : Une évolution du protocole NetFlow standardisé, permettant une meilleure flexibilité dans la collecte et l'analyse des flux réseau pour la gestion et la surveillance de la sécurité.

Logs : Enregistrements horodatés générés par un système informatique. Il relate les événements qui se sont produits sur le système au cours du temps.

NAT : Technique permettant de modifier les adresses IP source ou destination des paquets traversant un routeur ou un pare-feu, afin de permettre l'accès à Internet pour plusieurs appareils utilisant une seule adresse IP publique.

Netflow : Protocole développé par Cisco pour la collecte des données sur les flux du trafic réseau.

Phishing : Technique de cyberattaque ayant pour but de soutirer des informations sensibles auprès d'une personne, en se faisant passer pour une personne de confiance. Cette technique se fait par l'envoi de mails, des messages instantanés, etc.

Plugin : Un module logiciel additionnel permettant d'étendre les fonctionnalités d'une application principale. Les plugins sont couramment utilisés dans les navigateurs web, les SIEM et les logiciels de surveillance pour ajouter des capacités spécifiques.

RFC3164 : Protocole Syslog utilisé pour l'envoi de logs système sur un réseau avec UDP sur le port 514 et suit un format simple basé sur du texte.

RFC5424 : Version améliorée de la RFC 3164 qui standardise le format Syslog avec un encodage structuré, l'ajout d'horodatages précis et la possibilité d'utiliser un autre protocole pour sécuriser les transmissions.

SFTP : Protocole de communication utilisé pour le transfert et la gestion des fichiers. Il utilise le protocole SSH pour ajouter une couche de sécurité dans le transfert.

Spams : Messages électroniques non sollicités qui sont envoyés en masse, souvent à des fins publicitaires ou frauduleuses. Ils peuvent contenir des liens malveillants menant à du phishing ou à des logiciels malveillants.

SSH : Protocole réseau utilisé pour administrer des machines à distance en ligne de commande et transférer des fichiers de manière sécurisée. Il permet d'établir une connexion chiffrée entre un client et un serveur.

TCP : Protocole de communication orienté connexion qui assure une transmission fiable des données en garantissant leur arrivée dans l'ordre d'envoi et leur intégrité.

UDP : Protocole de communication non orienté connexion, qui sert à transmettre des données rapidement, mais sans la garantie de la transmission.

Virtualisation : Technologie qui permet de partager des ressources d'une machine physique pour la création des machines virtuelles.

VPN : Réseau privé virtuel permettant de sécuriser les connexions entre un utilisateur et un réseau distant ou entre deux réseaux distants, par la création d'un tunnel chiffré sur Internet. Il est utilisé pour assurer un accès sécurisé aux ressources d'une entreprise à distance.

Introduction Générale

Dans un monde de plus en plus connecté à la recherche d'une productivité accrue, le télétravail s'est avéré être une option alliant flexibilité, efficacité tout en travaillant à distance. Depuis l'avènement du Covid19, le nombre d'entreprises ayant adopté cette approche a augmenté. Les employés peuvent remplir leurs fonctions depuis leur domicile et combiner vie privée et professionnelle. Bien que cette méthode présente de nombreux avantages, elle a néanmoins des inconvénients. Ces inconvénients sont variés et l'un d'eux est une augmentation des risques de sécurité pour la structure l'utilisant. Plusieurs mesures de sécurité doivent être ainsi prises en compte afin de minimiser au maximum les risques. Parmi celles-ci, on note la journalisation des activités des utilisateurs distants, dans l'optique d'avoir une vue sur les événements ayant eu lieu dans l'environnement de travail et par conséquent de détecter des comportements suspects.

1. Contexte et justification

Le travail à distance prend de plus en plus d'ampleur au fil des années. Avec lui, les mesures de sécurité augmentent pour garantir un environnement de travail fiable et protégé. Il est à noter que des entreprises comme le Port autonome de Cotonou, qui assure une permanence pour répondre aux besoins des utilisateurs et assurer la continuité des activités, ont recours à cette méthode. Une fois chez eux, certains employés doivent accéder au réseau interne de l'entreprise. Cependant, ces entreprises se trouvent confrontées à une situation où elles peuvent identifier les utilisateurs distants connectés, mais sans assez d'informations détaillées sur leurs actions. Ce manque de visibilité peut poser des risques en matière de sécurité et de conformité. Afin d'y remédier, cette étude propose la mise en place d'un outil de centralisation des logs qui permettra de collecter, analyser et tracer les événements réseau. Cela offrira une meilleure visibilité sur les activités des utilisateurs, qu'ils soient présents dans l'entreprise ou à distance, et aidera à détecter d'éventuelles anomalies.

2. Problématique

Le télétravail est une solution optimale aidant les entreprises à assurer une continuité des activités avec des employés en dehors de leurs locaux. Il nécessite un accès aux ressources privées de l'entreprise depuis Internet. Internet est reconnu comme un environnement exposé aux menaces. Une mauvaise configuration pourrait être une faille propice aux attaques. Bien que les systèmes mis en

place permettent de combler cette faille, ils n'offrent pas une visibilité sur les actions effectuées par les utilisateurs externes une fois dans le réseau internet. La problématique actuelle est la suivante : comment suivre les actions des utilisateurs externes une fois dans l'environnement de l'entreprise ?

3. Objectifs

L'objectif général de ces travaux est de fournir une visibilité plus ou moins complète des actions exécutées par les utilisateurs en particulier ceux distants afin de renforcer la sécurité et la traçabilité des accès. Plus précisément, il s'agit d'être en mesure de :

- mettre en place une solution efficace de centralisation des logs ;
- surveiller le flux réseau émis par les utilisateurs distants ;
- corréler et analyser les données obtenues afin de les relier aux utilisateurs concernés et identifier les comportements suspects.

4. Présentation de la structure

Le Port Autonome de Cotonou est l'entité responsable de la direction et de la coordination des opérations portuaires. Reconnu comme "poumon de l'économie du Bénin", il joue un rôle important dans le commerce sous-régional en favorisant l'import-export. Il s'occupe également du bon déroulement des infrastructures et des projets d'agrandissement pour optimiser son efficacité et sa compétitivité.

5. Organisation du mémoire

Ce mémoire est structuré en trois chapitres. Le premier est consacré à une revue de littérature, abordant des notions sur le télétravail, les mesures de sécurité à prendre en compte ainsi qu'une présentation de quelques SIEM (System Information and Event Management). Le second chapitre présente la méthode utilisée pour choisir notre outil et son déploiement. Le dernier chapitre quant à lui présente les différents tests effectués suivi des résultats obtenus pour chacun d'eux.

Chapitre 1

Revue de littérature

Introduction

Ce chapitre, passe en revue les notions en lien avec le télétravail ainsi que les aspects à prendre à compte, relatifs au maintien de la sécurité des données. Une étude comparative de quelques outils qui permettraient de conserver cette sécurité est également exposée.

1.1 Télétravail

1.1.1 Définition

Selon le Dictionnaire Le Robert, le télétravail est défini comme une "activité professionnelle exercée à distance (d'une entreprise), grâce aux outils de télécommunication" [1]. Il définit un environnement au sein duquel une activité exercée au sein des locaux d'une société peut être à présent faite en dehors du lieu de travail, et cela grâce aux nouvelles technologies. Une fois les conditions réunies, ils peuvent rester en communication avec ceux qui sont sur site et assurer ensemble la continuité des activités pour l'atteinte des objectifs.

1.1.2 Avantages et inconvénients

Le télétravail de plus en plus utilisé présente de nombreux avantages. En voici quelques-uns [2] :

- Amélioration de la qualité de vie : le télétravail donne aux employés la possibilité de définir eux-mêmes les moments adéquats pour le travail, ce qui leur donne l'occasion de mieux concilier vie privée et professionnelle. Cette flexibilité et le lieu où ils sont les épargnent du stress occasionné par plusieurs facteurs, dont la présence des supérieurs ou des collègues, tout ceci contribuant à un travail productif et à un cœur apaisé ;
- Réduction des dépenses : avec le travail à distance, les dépenses liées au déplacement, la nourriture, et pleins d'autres sont réduites et permettent aux employés de faire des économies. Il en est de même pour les factures d'électricité ou d'eau (en ce qui concerne les employeurs) ;

- Accès à un bassin de talents plus large : le télétravail offre aux employeurs la possibilité de recruter des talents présents dans diverses localités capables de fournir un travail de qualité profitable pour l'entreprise, sans prendre en compte le facteur emplacement [3].

Bien qu'ayant des avantages, le télétravail présente également des inconvénients pour toutes les parties prenantes. En voici quelques-uns [3] :

- Isolement et perte de connexion sociale : les employés à distance se retrouvent dans un environnement routinier où ils n'ont pas réellement de contact avec l'extérieur ce qui les handicape lorsqu'ils sont amenés plus tard à travailler en groupe ou en communauté.
- Problèmes de compréhension et de communication entre collaborateurs : travailler ensemble à distance avec les facteurs qui entrent en jeu comme la qualité de la connexion Internet peut s'avérer difficile et occasionner des malentendus et une mauvaise compréhension sur des aspects essentiels du travail.
- Indépendance des employés : c'est à la fois un avantage et un inconvénient. L'absence des employés dans les locaux empêche le contrôle de leurs tâches par un supérieur hiérarchique, ce qui pourrait mettre un doute sur la qualité du travail fourni par ces derniers.
- Risque de sécurité : l'environnement créé par le télétravail expose les données de l'entreprise à l'extérieur et peut être une faille exploitable par des personnes malfaisantes.

1.1.3 Le facteur sécurité

Grâce au télétravail, les employeurs et les employés peuvent continuer leurs activités ayant trait à l'entreprise depuis n'importe quel emplacement hors des bureaux de cette dernière. Le réseau de l'entreprise ou les actifs qui ne devraient pas être disponibles au réseau externe se retrouvent "accessibles". Cette accessibilité peut être perçue comme une exposition, des biens et des informations confidentielles, ce qui est un problème en termes de sécurité informatique. Dans ce contexte, la sécurité informatique dispose d'une place capitale dans le contexte du télétravail, comme il dit dans "*Impact of Teleworking and Remote Work on Business : Productivity, Retention, Advancement, and Bottom Line*" [1], et cela en prenant en compte certains aspects dont :

- Un accès sécurisé à distance : il est important de définir un accès distant sécurisé aux réseaux de l'entreprise aisni qu'à ses ressources. Cela implique l'utilisation d'un VPN (Virtual Private Network) et d'une authentification à plusieurs facteurs. Le VPN est une technologie qui permet une connexion à distance de manière sécurisée à travers Internet, à un autre réseau. Il crée un tunnel entre les machines où il est configuré afin de dissimuler toutes les informations qui passeront. Cela garantit une confidentialité des données ainsi que leur intégrité ;
- Des mises à jour régulières sur les ordinateurs : les mises à jour du système et des logiciels corrigent les failles de sécurité qui pourraient être exploitées par une personne malveillante ;

En dehors de ces aspects, il est important de prendre en compte d'autres éléments comme :

- L'installation des anti-virus : l'installation et l'utilisation d'un antivirus renforcent la sécurité des équipements. Ces outils détectent les logiciels malveillants et les sites web à risque, assurant une navigation et un stockage protégés ;

- La surveillance des activités des collaborateurs : il s'agit de suivre et d'enregistrer les actions des collaborateurs sur les systèmes auxquels ils ont accès. En cas d'attaque, cela permet d'identifier rapidement la source du problème et de réagir efficacement pour limiter les dommages ;

Une autre mesure de sécurité consiste à établir une politique de sécurité couvrant plusieurs aspects importants, en rappelant que les règles de l'organisation restent applicables même en télétravail ; ce que nous rappelle **Jagdish Krishanlal Arora** dans "*Cyber Security*" [2]. Cela permet de garantir la protection des données et le respect des règles existantes.

1. Inventaire des appareils utilisés : il est important de connaître les appareils utilisés par les collaborateurs et de savoir qui les utilise. Cela permet de suivre les incidents éventuels et de s'assurer que les employés respectent les droits qui leur sont attribués ;
2. Séparation des appareils professionnels du personnel : certains employés utilisent leurs équipements de travail pour des activités personnelles, ce qui peut entraîner des risques de fuite ou de corruption des données sensibles de l'entreprise ;
3. Utilisation des mots de passe robustes : une politique de mots de passe pour exiger l'utilisation de majuscules, minuscules, chiffres et caractères spéciaux. De plus, ces mots de passe doivent expirer et être renouvelés périodiquement ;
4. Authentification à double facteur : il est recommandé d'utiliser une authentification à deux facteurs pour sécuriser l'accès aux ressources. Cela ajoute une couche de protection en vérifiant l'identité de l'utilisateur par deux moyens distincts.

Ces mesures sont essentielles pour réduire les risques d'attaques virtuelles liées au télétravail, bien qu'elles n'offrent pas une protection totale. Il est également crucial de penser à la sécurité physique des équipements de l'entreprise, qui peuvent être exposés au vol, aux dommages ou à la destruction lorsqu'ils sont hors des bureaux.

Pour que ces pratiques soient efficaces, il est indispensable de former les employés concernés. Ils doivent comprendre l'importance de ces mesures et adopter de bonnes pratiques, comme l'utilisation de réseaux sécurisés (éviter les connexions publiques dans les cafés ou bibliothèques), la modification des informations de connexion par défaut sur les routeurs personnels, et la notification immédiate des incidents tels que le phishing ou les spams.

1.2 Journalisation des activités

La journalisation des activités consiste à enregistrer et analyser les événements du système, tels que les authentifications, les connexions réseau, les changements d'accès, les alertes de sécurité, etc. En ce qui concerne le télétravail, elle est essentielle pour suivre les connexions à distance des employés, détecter des anomalies et identifier rapidement les intrusions ou attaques potentielles. A l'aide d'outils comme les SIEM (Security Information and Event Management), cette collecte d'informations sur les flux de données et les comportements suspects permet de garantir la sécurité du réseau et de réagir à temps en cas de menace.

1.2.1 Définition et présentation d'un SIEM

Un SIEM est un système centralisé utilisé pour la collecte, l'analyse et la gestion des événements et des informations de sécurité dans un réseau. Les informations recueillies ont diverses sources, telles que les serveurs, les pare-feu, les antivirus et autres équipements réseau. Il offre un aperçu en temps réel du système d'informations de l'entreprise, permettant de détecter rapidement les anomalies et vulnérabilités. Le SIEM facilite une réponse rapide aux incidents de sécurité, aidant ainsi à diminuer les dommages potentiels grâce à une surveillance [4].
Le SIEM combine deux concepts de sécurité : le SIM (Security Information Management), qui récupère et gère les logs et données de sécurité, et le SEM (Security Event Management), qui analyse et génère des alertes en temps réel. Ces deux composantes fonctionnent ensemble au sein du SIEM afin d'offrir une solution de sécurité intégrée, assurant ainsi la gestion complète des informations de sécurité.
Il existe deux catégories de logiciels qui sont aussi valables pour les SIEM : les logiciels open source et les logiciels propriétaires [5] :

- Un logiciel dit propriétaire est un système dont le code source appartient à une organisation particulière, laquelle en détient tous les droits et bloque toute modification de la part d'autres utilisateurs. L'accès à ces logiciels se fait généralement par abonnement ou achat, bien que certaines versions limitées en fonctionnalités puissent être proposées gratuitement.

- Une solution open source est un logiciel dont le code source est rendu public et accessible à tous, sans frais. Des versions plus avancées ou améliorées peuvent être proposées moyennant un paiement, souvent accompagnées de services de maintenance et de support.

Les solutions open source offrent plus de flexibilité, de personnalisation et permettent d'avoir des outils performants à moindre coût. C'est la raison pour laquelle dans la suite, l'étude sera focalisé exclusivement sur ceux qui sont open source.

1.2.2 Présentation de quelques SIEM

1.2.2.1 Elastic Stack

Elastic Stack, est une suite d'outils composée de **Elasticsearch**, **Logstash**, **Kibana**, et des **Beats**, qui ensemble offrent une solution complète de gestion et d'analyse de données. **Elasticsearch** est un moteur de recherche et d'analyse distribué permettant de centraliser les informations et d'en faciliter la recherche et l'indexation. **Logstash** collecte, transforme et transfère les données vers Elasticsearch, tandis que **Kibana** fournit une interface utilisateur agissant comme un miroir interactif entre les données stockées et les utilisateurs, pour simplifier les analyses et la visualisation. Les **Beats**, agents légers de collecte, transmettent des données issues de diverses sources directement vers Logstash ou Elasticsearch. Par sa flexibilité et ses fonctionnalités, cette suite permet de surveiller les activités réseau et de détecter efficacement les anomalies. L'Elastic Stack offre une visibilité complète sur l'infrastructure, faisant de lui un outil approprié pour la surveillance efficace d'un environnement de production.

1.2.2.2 Graylog

Graylog est une solution de gestion centralisée des journaux, offrant aux entreprises une vue plus ou moins globale de la sécurité de l'infrastructure. Idéal pour les petites et moyennes entreprises,

il est abordable et capable de détecter les anomalies pour garantir le bon fonctionnement du système. Graylog s'adapte aux évolutions du réseau et à l'augmentation du volume de données sans perte d'efficacité. Il s'appuie sur une architecture solide, composée d'un serveur pour le traitement et l'analyse des logs, d'**Elasticsearch** ou **OpenSearch** pour le stockage et l'indexation rapide des logs, et de **MongoDB** pour la gestion des métadonnées et des configurations. Avec des pipelines de traitement, des dashboards personnalisables et des alertes automatisées, il permet de centraliser les logs et d'améliorer la visibilité sur l'infrastructure IT ce qui permet de répondre aux besoins d'audit, et de performance.

1.2.2.3 Wazuh

Wazuh est un outil centralisé pour l'agrégation et l'analyse en temps réel des données et des logs de sécurité, ce qui aide à identifier les menaces et à sauvegarder les systèmes. Elle repose sur trois composantes principales : **Wazuh Agent**, installé sur les machines cibles pour lire et récupérer les événements de sécurité, les informations système et d'autres données pertinentes ; **Wazuh Manager**, qui centralise le traitement et la corrélation des informations obtenues afin d'identifier des activités suspectes ou malveillantes ; et **Wazuh Dashboard**, une interface web intuitive, qui facilite la gestion des alertes, la configuration du système et le suivi des analyses de conformité. En complément, Wazuh propose une fonctionnalité de vérification de conformité aux normes. Grâce à sa capacité à surveiller en temps réel et à offrir une analyse détaillée des événements, Wazuh se révèle être un outil incontournable pour une surveillance proactive, surtout dans des contextes de télétravail où les menaces peuvent émerger de divers points d'accès.

Quelques avantages et inconvénients de ces trois SIEM sont répertoriés et présentés dans le tableau 1.1.

TABLEAU 1.1 : Avantages et Inconvénients de quelques SIEM

Outils	Avantages	Inconvénients
Elastic Stack	- Scalabilité - Grande flexibilité pour différents cas d'utilisation - Analyse des données en temps réel	- Configuration avancée requise - Nécessité de grandes ressources matérielles pour des grands volumes de données - Besoin de compétences techniques pour la maintenance
Graylog	- Alertes personnalisables - Recherche rapide et détaillées des logs - Stabilité et évolutivité	- Dépendance à Opensearch et MongoDB
Wazuh	- Analyse des données en temps réel - Intégration avec des outils tiers - Sécurité et conformité aux normes - Alertes et notifications automatisées	- Complexité de déploiement - Personnalisation limitée

Conclusion

Dans ce chapitre, un rappel des concepts clés en rapport avec le télétravail a été présenté ainsi que quelques solutions SIEM qui pourraient répondre à la problématique. Le chapitre suivant est dédié à la présentation de la méthode de sélection de l'outil ainsi qu'à son déploiement.

Chapitre 2

Choix et mise en place de la solution

Introduction

Ce chapitre porte sur le choix de l'outil à utiliser dans ce travail. Dans un premier temps, les critères pris en compte ont été exposés, ainsi que la méthode d'évaluation de ces derniers au niveau de chaque outil. Dans un second temps, l'outil révélé comme le plus adéquat à partir des résultats de l'évaluation, a été mis en place.

2.1 Choix de la solution

2.1.1 Présentation des critères et de la méthode d'évaluation des solutions

Dans le but de choisir le SIEM approprié pour effectuer le travail, l'analyse s'est basée sur des critères définis qui aideront à apprécier les trois outils à comparer. Ces critères sont en lien avec des aspects attendus au niveau du SIEM. La solution doit être :

- simple dans le déploiement et l'utilisation sans la nécessité d'une expertise technique avancée des employés ou d'une formation dédiée, afin d'être adaptée à tout type d'entreprise et d'avoir un gain de temps ;
- capable de s'intégrer aisément dans une infrastructure existante et de fonctionner en complémentarité avec les systèmes et protocoles déjà mis en place ;
- malléable afin de s'adapter aux besoins spécifiques de l'entreprise, par exemple au niveau de l'affichage des données et au niveau de la création d'alertes personnalisables obtenues avec des conditions définies par l'administrateur ;
- en mesure de présenter les données collectées dans une interface graphique pour faciliter la recherche et l'analyse des données ;
- de fonctionner sans consommer trop de ressources matérielles pour ne pas impacter la disponibilité ou le fonctionnement des autres services présents.

Les critères issus de ces besoins sont récapitulés dans le tableau 2.1 accompagnés de la méthode d'évaluation utilisée.

TABLEAU 2.1 : Critères de choix et méthodes d'évaluation

Critères	Explication	Méthode d'évaluation
Facilité d'utilisation et de déploiement	Mesure le niveau de difficulté lié à l'installation et à l'utilisation de l'outil	Analyse basée sur les tutoriels disponibles, la qualité de la documentation fournie et sur les avis des utilisateurs
Intégration avec d'autres outils	Evalue la compatibilité de l'outil avec d'autres solutions et technologies utilisées dans l'environnement	Analyse réalisée à partir de la documentation officielle, et des avis des utilisateurs
Visualisation graphique	Evalue la capacité de l'outil à présenter les données collectées de manière claire, intuitive et exploitable grâce à des représentations visuelles	Evaluation faite à partir de l'analyse des tutoriels, de la documentation officielle, et des exemples de tableaux de bord ou visualisations disponibles
Flexibilité et personnalisation	Mesure à quel point la solution peut s'adapter aux besoins spécifiques d'une organisation	Evaluation basée sur la documentation et les tutoriels ainsi que sur des avis des utilisateurs
Consommation des ressources	Mesure l'impact de l'outil sur les ressources système, comme la mémoire, le processeur et l'espace disque	Analyse basée par la lecture de la documentation et des avis des utilisateurs

2.1.2 Comparaison de quelques SIEM

Parmi les SIEM open source les plus utilisés, trois ont retenu notre attention, Elastic Stack, Wazuh et Graylog. Il s'agit de ceux qui apparaissent sur plusieurs avis pour le top 10 des meilleurs SIEM. Une comparaison de ces derniers a été faite dans le but de déterminer si l'un d'eux est la solution adéquate aux besoins. A chacun des critères, une note est attribuée.
Le tableau 2.2 présente les notes attribuées dans une grille comparative des différentes solutions.

TABLEAU 2.2 : Grille comparative de quelques SIEM

	Objectif	Elastic Stack	Graylog	Wazuh
Facilité d'utilisation et de déploiement	5	4.2	4.2	3.8
Intégration avec d'autres outils	5	4.4	4.4	4.3
Visualisation graphique	5	5	4	5
Flexibilité et personnalisation	5	4.5	4.3	4.6
Consommation des ressources	5	3	3	3
Total	30	21.1	19.9	20.7

En ce qui concerne le critère *"Visualisation graphique"*, Kibana est reconnu pour ses puissantes fonctionnalités d'exploration et de visualisation des données [6]. La documentation de Wazuh [7] indique que son interface graphique, Wazuh Dashboard, est basée sur Kibana et bénéficie donc de fonctionnalités similaires. En revanche, Graylog présente certaines limites en matière de visualisation. Des utilisateurs ont notamment signalé l'absence d'une barre de recherche et de filtres temporels dans certaines configurations, ce qui peut restreindre l'affichage et l'exploration des tableaux de bord [8]. Par ailleurs, chacun de ces outils requiert l'installation de plusieurs éléments, tels qu'un moteur d'analyse, une base de données et une interface de visualisation. Avec l'augmentation du volume des données collectées et analysées, la consommation des ressources système peut devenir un enjeu majeur en termes de performance et de scalabilité. Ces différentes informations justifient les notes attribuées à chaque SIEM pour la visualisation graphique et la consommation des ressources.
Les notes attribuées pour tous les autres critères proviennent de Gartner [9] et résultent de l'analyse des avis des utilisateurs..
Comme le montre le tableau 2.2, Elastic Stack est celui qui a obtenu le plus grand score, par conséquent, l'outil utilisé dans la suite pour effectuer le travail, est Elastic Stack.

2.2 Mode de fonctionnement de la solution

2.2.1 La suite Elastic

L'Elastic Stack ou ELK (Elasticsearch Logstash Kibana) est une suite d'outils qui offre des fonctionnalités d'analyse et de traitement des logs. Son architecture est extensible et peut s'adapter aux besoins de l'entreprise. Chaque composant a une fonction précise dans le processus, de la collecte des données jusqu'à leur visualisation.

- **Beats** : ces agents légers collectent les données directement sur les machines sources. Chaque type est conçu et idéal pour une catégorie de données spécifique (métriques, logs, paquets

réseau, etc), ce qui réduit la surcharge inutile et améliore l'efficacité de la collecte. Les données collectées sont ensuite transférées à Logstash ou Elasticsearch.

- **Logstash** : il sert de pipeline de traitement et d'enrichissement des données, qui sont par la suite transférées à Elasticsearch. Il est utilisé pour effectuer les tâches suivantes :
  - centralisation et normalisation des logs de différentes sources ;
  - filtrage et structuration des logs ;
  - corrélation des événements pour identifier des anomalies
  - déclenchement d'alertes en cas de comportement suspect
- **Elasticsearch** : il est le cœur de la suite Elastic. IL s'occupe du :
  - stockage et indexation des logs des accès distants
  - rétention et archivage des logs
- **Kibana** : Quant à Kibana, c'est l'interface utilisateur, qui utilisé pour :
  - la visulisation graphique dynamique des logs ;
  - l'utilisation de filtres et requêtes pour identifier des éléments importants ;

Ensemble, ces outils assurent une chaîne complète : collecte (Beats), traitement (Logstash), stockage et recherche (Elasticsearch), et visualisation (Kibana). Cette architecture rend l'Elastic Stack adéquat pour la surveillance et l'investigation.
La figure 2.1, montre les différents beats et le flux de données dans le suite ELK.

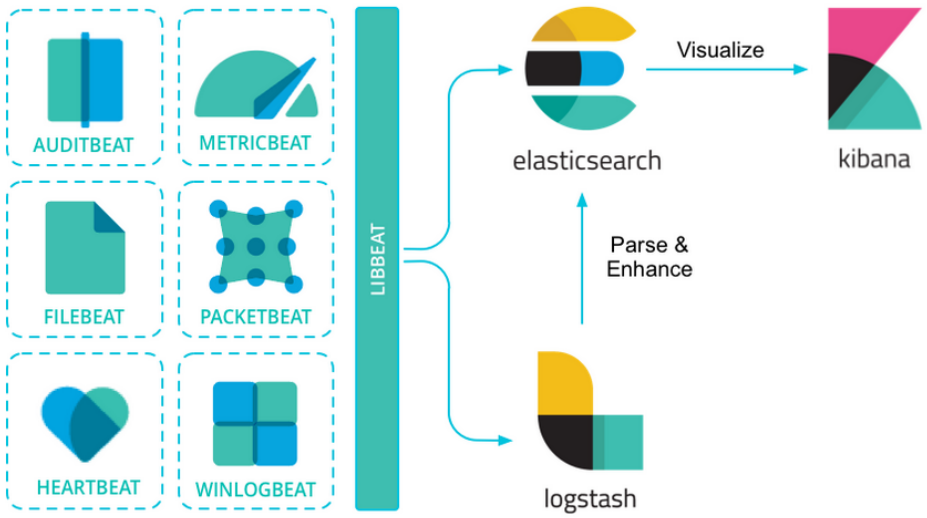


FIGURE 2.1 : Flux de données dans la suite ELK [10]

2.2.2 Présentation des différents beats

Six beats sont disponibles pour la collecte de différentes données.

- **Auditbeat :** est utilisé pour l'audit des événements de sécurité sur un système notamment sous Linux, où il exploite auditd, pour surveiller et enregistrer les activités des utilisateurs ainsi que

les processus exécutés sur le système. Il surveille également les modifications sur les fichiers et identifie en fonction des règles pré-établies des infractions possibles à la politique de sécurité. Sur d'autres systèmes, comme Windows, il s'appuie sur les journaux d'événements pour collecter des informations similaires. ;

- **Filebeat** : a pour rôle de collecter les informations depuis les fichiers mais aussi à partir d'autres sources comme syslog pour tout système d'exploitation ;
- **Heartbeat** : est utilisé pour vérifier de manière périodique l'état des différents services afin d'identifier s'ils sont disponibles ou accessibles ;
- **Metricbeat** : est utilisé pour analyser les performances et l'état des ressources du système par la collecte des métriques adéquates sur le processeur, la mémoire vive, le disque, etc ;
- **Packetbeat** : est un analyseur de paquets réseau qui capture le trafic réseau et récupère les données des protocoles de la couche applicative dans l'optique d'évaluer les performances applicatives ;
- **Winlogbeat** : est utilisé pour les événements qui ont eu lieu sur un système Windows, en rapport avec le système, la sécurité et les applications.

Ce travail vise à suivre les activités des utilisateurs sur les systèmes critiques. Auditbeat répond à ce besoin en renseignant sur les processus et sur les utilisateurs. Pour compléter ces données, il est important d'avoir également les logs générés pour chaque action, ce que fait Filebeat. Par ailleurs, le pare-feu est un équipement qui ne permet pas l'installation d'un beat. Cependant, il est possible de récupérer les événements syslog de ce dernier par le transfert vers un serveur où Filebeat, à l'aide de son module syslog dédié, pourra les récupérer. En tenant compte de ces paramètres, les beats adaptés sont Auditbeat et Filebeat.

2.3 Déploiement de la solution

Pour effectuer le travail, une architecture réseau simple a servi de base, la voici présentée sur la figure 2.2.

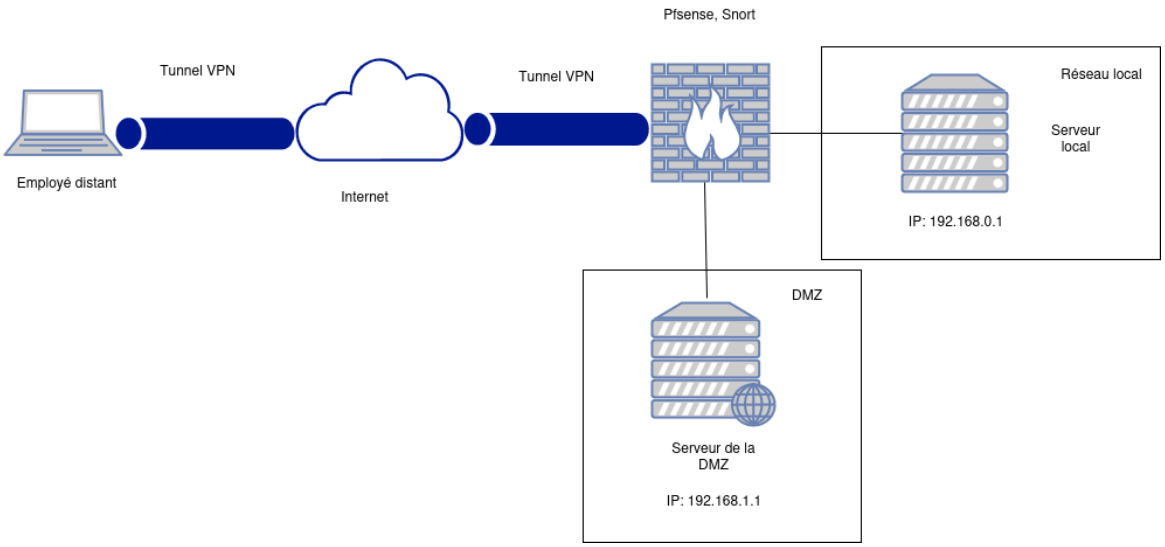


FIGURE 2.2 : Architecture réseau

L'architecture est composée de deux serveurs sous Linux. Le premier, situé dans une zone démilitarisée (DMZ), héberge des services accessibles depuis Internet, il s'agit des services DNS et web. Le service DNS est utilisé à la fois par le réseau internet et l'extérieur. Le second serveur est situé dans le réseau local. Il héberge ELK ainsi qu'un service de fichiers avec le protocole SFTP. Les deux serveurs sont accessibles à distance par SSH. Ces deux serveurs, étant dans des réseaux différents, sont reliés par un pare-feu Pfsense qui est combiné à un Snort, un IDS (Intrusion Detection System) afin de présenter un environnement sécurisé. Le pare-feu est l'équipement sur lequel le VPN est configuré et qui représente la porte d'entrée au réseau de l'entreprise. Des services essentiels ont été privilégiés pour tester ELK, et déployer l'ensemble de l'architecture sur une seule machine pour simplifier l'environnement. Ce déploiement sera fait avec l'utilisation de l'outil de virtualisation Virtualbox de Oracle.

2.3.1 Installation de la solution

Voici les différentes étapes de l'installation de l'outil.

- Importation de la clé PGP de Elasticsearch

```
wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo gpg
--dearmor -o /usr/share/keyrings/elasticsearch-keyring.gpg
```

- Installation du paquet apt-transport-https

```
sudo apt-get install apt-transport-https
```

- Ajout du dépôt officiel Elasticsearch à la liste des sources APT

```
echo "deb [signed-by=/usr/share/keyrings/elasticsearch-keyring.gpg]
https://artifacts.elastic.co/packages/8.x/apt stable main" | sudo tee
/etc/apt/sources.list.d/elastic-8.x.list
```

- Mise à jour des paquets et installation du service Elasticsearch

```
sudo apt-get update
sudo apt-get install elasticsearch
```

- Activation et démarrage du service elasticsearch

```
sudo systemctl daemon-reload
sudo systemctl enable elasticsearch.service
sudo systemctl start elasticsearch.service
```

Afin de s'assurer que l'installation s'est bien effectuée, il faut se rendre sur l'url "https://ip-address:9200", comme le montre la figure 2.3, où *"ip-address"* représente l'adresse IP de la machine où Elasticsearch est installé.

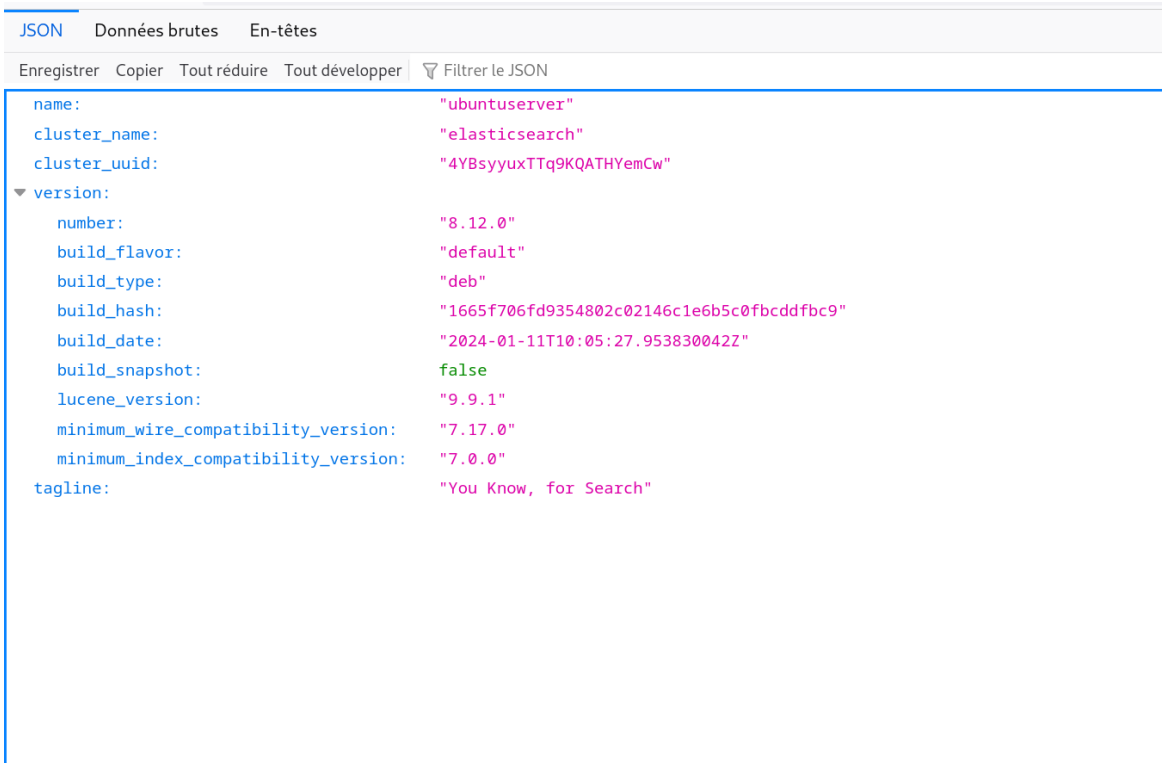


FIGURE 2.3 : Vérification du fonctionnement de Elasticsearch

- Installation de Kibana

```
sudo apt-get install kibana
```

- Activation et démarrage de Kibana

```
sudo systemctl enable kibana.service
sudo systemctl start kibana.service
```

Pour entamer la configuration de Kibana, accédons à une page web via l'url "http://ip-address:5601", où *"ip-address"* représente l'adresse IP de la machine où Kibana est installé. Voilà l'interface en question sur la figure 2.4.

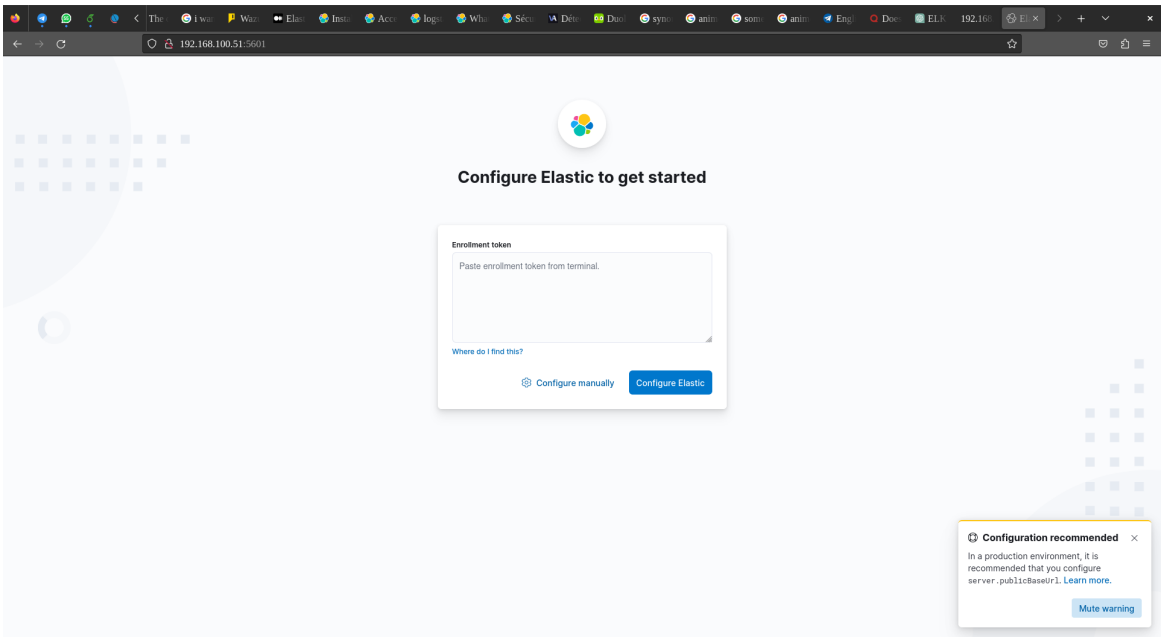


FIGURE 2.4 : Demande du token dans Kibana

Le token demandé, qui figure parmi les informations affichées à la fin de l'installation d'Elasticsearch, peut être obtenu en exécutant la commande suivante.

```
/usr/share/elasticsearch/bin/elasticsearch -create-enrollment-token -s kibana
```

Une fois le token inséré, un code de vérification est demandé, comme présenté sur la figure 2.5. Ce dernier peut être obtenu en vérifiant l'état ou le statut du service kibana en cours d'exécution.

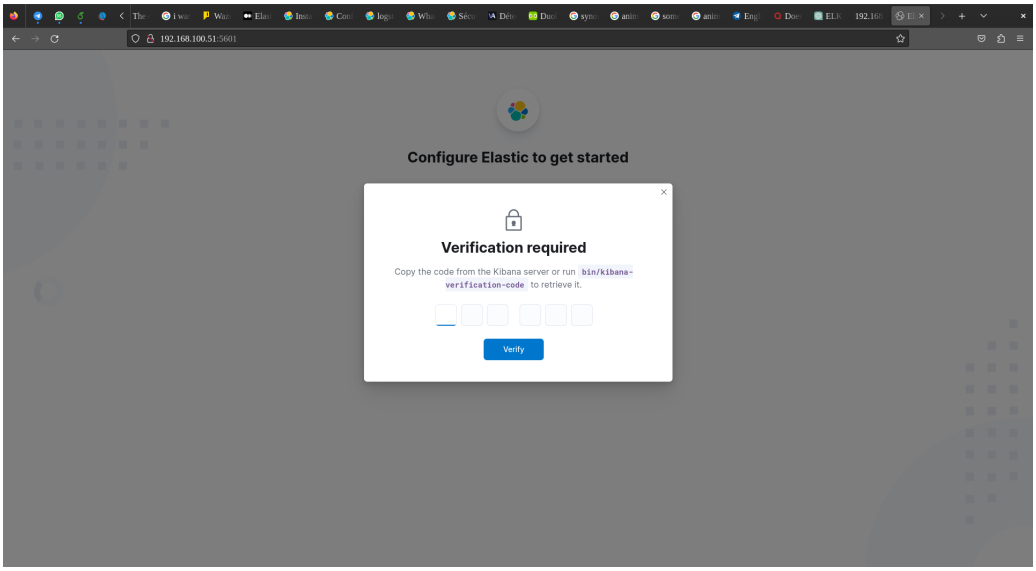


FIGURE 2.5 : Demande du code de vérification

Comme le montre la figure 2.6, le code se situe sur la dernière ligne affichée.

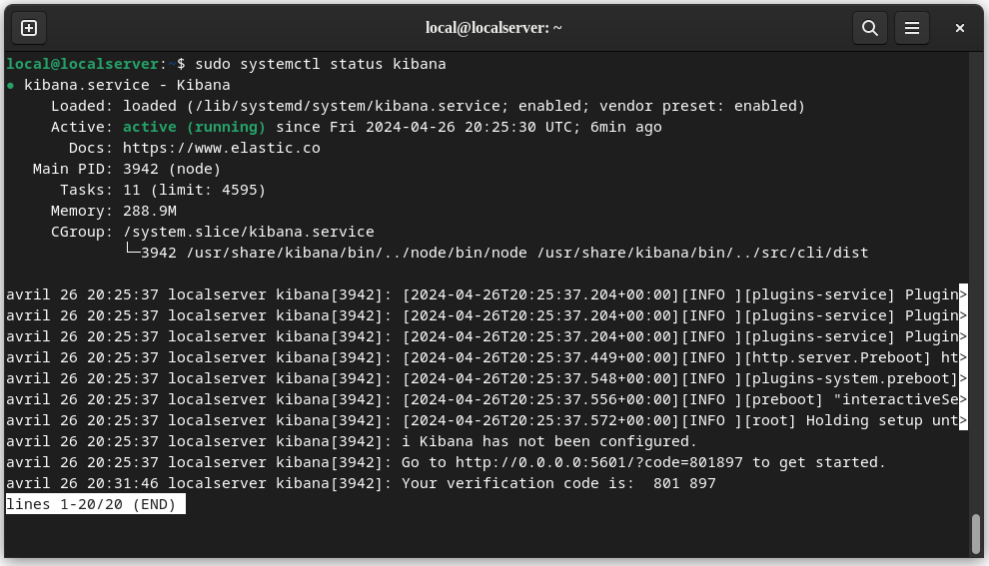


FIGURE 2.6 : Affichage du statut de Kibana

Une fois le code inséré, l'installation de Kibana continue et s'achève, comme le montre la figure 2.7.

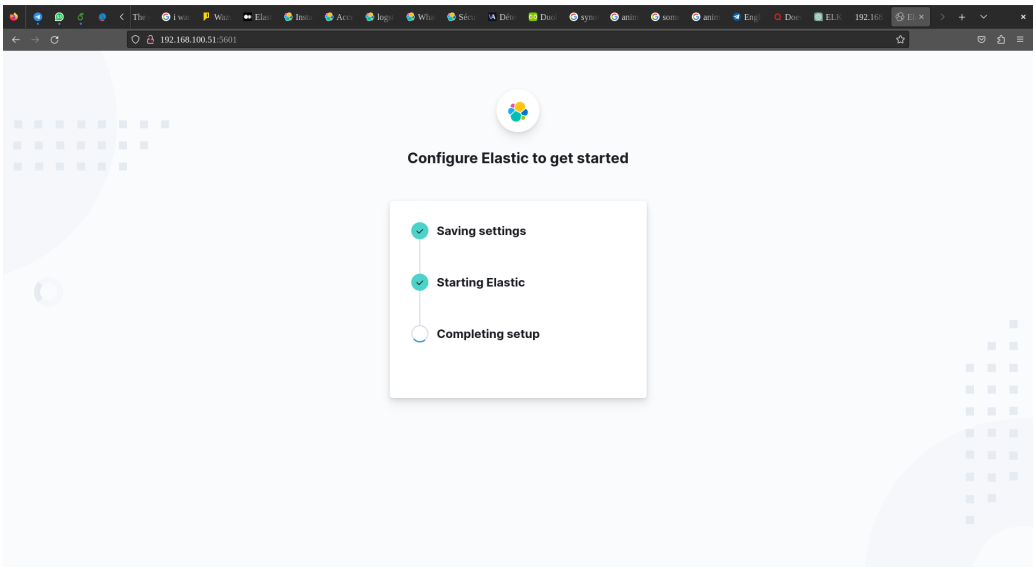


FIGURE 2.7 : Finalisation de la configuration de Kibana

Après renseignement des identifiants correspondant à l'utilisateur elastic, voici sur la figure 2.8 la page d'accueil de Kibana.

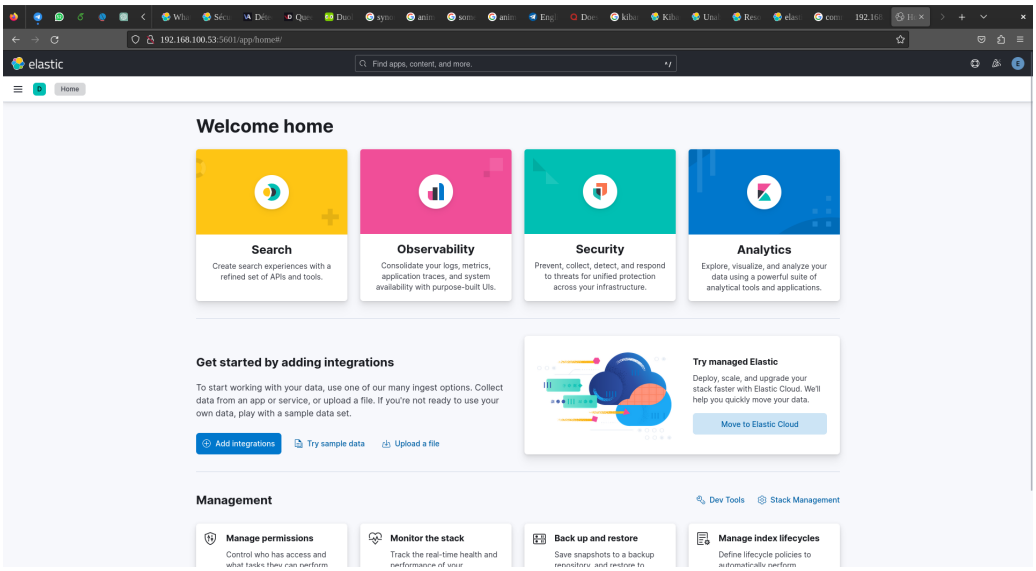


FIGURE 2.8 : Page d'accueil de Kibana

- Installation, activation et démarrage de Logstash

```
sudo apt-get install logstash
sudo systemctl enable logstash.service
sudo systemctl start logstash.service
```

La figure 2.9 montre l'état du service de Logstash.

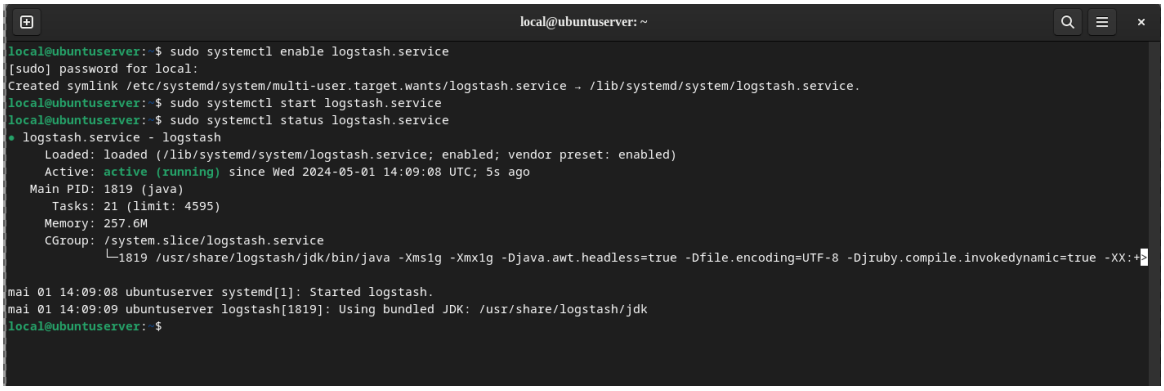


FIGURE 2.9 : Affichage du statut de Logstash

Comme indiquer Logstash est opérationnel.

- Installation des Beats
  Filebeat et **Auditbeat**, les beats qui seront utilisés dans le contexte actuel, seront installés sur chaque serveur en exécutant les commandes ci-après.

```
sudo apt-get install filebeat
sudo systemctl enable filebeat.service
sudo systemctl start filebeat.service
```

Les mêmes commandes seront utilisées pour Auditbeat en remplaçant filebeat par auditbeat.
A présent tous les éléments de la suite sont installés.

2.3.2 Configuration de la suite

La configuration de Filebeat consiste à définir les inputs (sources de données) et les outputs (destinations des données collectées). Un input gère les harvesters, qui sont des composants responsables de lire les données d'une source spécifique, comme un fichier de log, et de les transmettre au module d'input pour traitement.

- **Filestream** : conçu pour lire les fichiers de logs actifs, tels que ceux localisés dans le répertoire /var/log sous Linux.
- **Syslog** : utilisé pour récupérer les événements syslog via UDP ou TCP, en conformité aux formats RFC 3164 et RFC 5424. Cet input assurera l'intégration des journaux générés par le pare-feu, équipement sur lequel Filebeat ne peut pas être installé pour une lecture directe.
- **Netflow** : conçu pour lire les flux de données enregistrés par Netflow et IPFIX via UDP. Ce protocole, configuré sur pfSense, est dédié à l'observation et l'analyse du trafic réseau entre les différents réseaux. Les informations générées par Netflow seront collectées et traitées par cet input.

Les inputs Syslog et Netflow seront utilisés uniquement sur le serveur situé dans le réseau local. Les figures suivantes affichent les différentes configurations effectuées dans le fichier /etc/filebeat/filebeat.yml. La figure 2.10 montre la configuration du premier input de Filebeat.

```
# ============================== Filebeat inputs ===============================

filebeat.inputs:

# Each - is an input. Most options can be set at the input level, so
# you can use different inputs for various configurations.
# Below are the input-specific configurations.

# filestream is an input for collecting log messages from files.
- type: filestream

  # Unique ID among all inputs, an ID is required.
  id: my-filestream-id

  # Change to true to enable this input configuration.
  enabled: true

  # Paths that should be crawled and fetched. Glob based paths.
  paths:
    - /var/log/*.log
  scan_frequency: 1s
    #- c:\programdata\elasticsearch\logs\*
```

FIGURE 2.10 : Configuration de l'input filestream

On observe l'input filestream activé, configuré pour lire, chaque seconde, les fichiers portant l'extension .log dans le répertoire /var/log/ de la machine où il est déployé. La figure 2.11 montre les deux autres inputs configurés.



FIGURE 2.11 : Configuration des inputs syslog et netflow

Comme illustré sur la figure 2.11, les inputs syslog et netflow sont définis. Les événements syslog seront obtenus par UDP sur le port 9000 de la machine, conformément au format RFC 3164. Les flux Netflow, quant à eux, seront réceptionnés sur le port 2055. Une vérification sera effectuée sur chaque port toutes les secondes.
Pour finaliser la configuration de Filebeat, il est nécessaire de spécifier l'output ou la destination des données collectées, présenté sur la figure 2.12.



FIGURE 2.12 : Configuration de l'output de Filebeat

Dans la configuration, Logstash est défini comme la destination des données collectées par Filebeat, accessible via le port 5044 du serveur où il est installé.

Auditbeat fait appel à des modules, qui sont des composants logiciels conçus pour collecter, structurer et organiser des données spécifiques qui proviennent d'une source ou d'un type d'activité particulier. Ces modules sont configurés avec des paramètres par défaut adaptés à chaque source, tout en permettant la personnalisation selon les besoins. Trois modules seront utilisés ici :

- **Auditd** : Ce module a pour fonction de recevoir les événements d'audits générés par le framework d'audit du noyau de Linux. En se basant sur des règles définies, il fournit des informations sur les commandes exécutées, les changements des droits ou des privilèges, etc.
- **System** : Ce module est chargé de collecter des informations liées à la sécurité du système. Il envoie des données régulières sur l'état du système et des mises à jour en temps réel sur les changements, comme le démarrage ou l'arrêt d'un processus.
- **File Integrity** : Ce module surveille et rapporte les événements liés aux fichiers uniquement lorsqu'une modification est effectuée, comme la création, la suppression ou la modification des fichiers. Il fournit également des détails sur les fichiers concernés et les actions réalisées.

La figure 2.13 montre les règles établies pour le module auditd.

```
# ------------------------------ Modules configuration -----------------------------
auditbeat.modules:

- module: auditd
  # Load audit rules from separate files. Same format as audit.rules(7).
  audit_rule_files: [ '${path.config}/audit.rules.d/*.conf' ]
  audit_rules: |
    ## Define audit rules here.
    ## Create file watches (-w) or syscall audits (-a or -A). Uncomment these
    ## examples or add your own rules.

    ## If you are on a 64 bit platform, everything should be running
    ## in 64 bit mode. This rule will detect any use of the 32 bit syscalls
    ## because this might be a sign of someone exploiting a hole in the 32
    ## bit API.
    -a always,exit -F arch=b32 -S all -F key=32bit-abi

    ## Executions by users.
    -a always,exit -F arch=b64 -S execve -F exe=/usr/bin/python3 -F auid>=1000 -F key=script_exec
    -a always,exit -F arch=b64 -S execve -F exe=/bin/bash -F auid>=1000 -F key=bash_exec
    -a always,exit -F arch=b64 -S execve -F auid>=1000 -F key=user_exec

    ## External access (warning: these can be expensive to audit).
    -a always,exit -F arch=b64 -S accept,bind,connect -F key=external-access

    ## Identity changes.
    -w /etc/group -p wa -k identity
    -w /etc/passwd -p wa -k identity
    -w /etc/gshadow -p wa -k identity

    ## Unauthorized access attempts.
    -a always,exit -F arch=b64 -S open,creat,truncate,ftruncate,openat,open_by_handle_at -F exit=-EACCES -k access
```

FIGURE 2.13 : Configuration du module auditd

Voici une explication de chaque règle.

- la première règle surveille tous les appels système exécutés en utilisant une architecture 32 bits. Son objectif est de détecter une possible tentative d'exploitation d'une faille en relation avec l'architecture 32 bits, sachant que le système fonctionne sur 64 bits ;
- les trois règles suivantes surveillent tous les processus ou les commandes exécutés sur le système par un utilisateur réel, en utilisant un script bash ou python ;
- la cinquième règle surveille les connexions réseaux ;
- les sixième, septième et huitième règles permettent de surveiller les changements sur les identités des utilisateurs, comme le groupe auquel ils appartiennent ou le mot de passe.
- la dernière règle permet de surveiller les actions sur les fichiers non-accessibles par certains utilisateurs.

La figure 2.14 montre la configuration faite pour les deux derniers modules.

```
  GNU nano 4.8                          auditbeat.yml

- module: file_integrity
  paths:
  - /bin
  - /usr/bin
  - /sbin
  - /usr/sbin
  - /etc

- module: system
  datasets:
    - package # Installed, updated, and removed packages

  period: 2m # The frequency at which the datasets check for changes

- module: system
  datasets:
#    - host    # General host information, e.g. uptime, IPs
    - login   # User logins, logouts, and system boots.
#    - process # Started and stopped processes
    - socket  # Opened and closed sockets
#    - user    # User information
```

FIGURE 2.14 : Configuration des modules system et file integrity

Pour le module File Integrity, les répertoires surveillés sont les suivants :

- /bin : qui contient les fichiers exécutables essentiels, qui sont nécessaires au fonctionnement du système pour tous les utilisateurs (comme les commandes de base) ;
- /usr/bin : qui stocke les binaires non-essentiels utilisés par les utilisateurs et les administrateurs pour des tâches spécifiques ;
- /sbin : qui regroupe les binaires essentiels réservés aux tâches d'administration système ;
- /usr/sbin : qui contient les outils d'administration système non critiques, souvent utilisés pour des services ou des configurations avancées ;
- /etc : qui héberge les fichiers de configuration du système et des applications.

En ce qui concerne le module system, les datasets pris en compte sont :

- login : pour fournir des détails sur les connexions, les déconnexions et les démarrages du système ;
- socket : dans le but de suivre les événements liés aux connexions réseau ouvertes et fermées ;
- package : pour donner des informations sur les paquets installés, mis à jour et supprimés.

Le destinataire des données d'Auditbeat, tout comme pour Filebeat, est Logstash.
Le paramétrage des Beats étant maintenant terminé, passons à la configuration de Logstash. Il est composé de trois éléments : les inputs, les filtres et la sortie. Les inputs, comme pour les Beats, représentent la source des événements, c'est-à-dire ceux qui les génèrent. Les filtres servent à structurer

ou à modifier les données, qui sont ensuite transférées vers une destination pour analyse. Dans le contexte actuel, les Beats (Filebeat et Auditbeat) sont les sources de données, et toutes les configurations relatives au traitement de ces dernières seront faites dans le fichier "/etc/logstash/conf.d/beats.conf".
La figure 2.15 montre que Logstash recevra les données des Beats par le port 5044.

```
input {
  beats {
    port => 5044  # Le port où Filebeat envoient les données
  }
}
```

FIGURE 2.15 : Configuration de l'input de Logstash

Une fois analysées, les données traitées par Logstash seront envoyées à Elasticsearch, comme le montre la figure 2.16.

```
output {
  elasticsearch {
    hosts => ["https://192.168.0.1:9200"]
    index => "%{[@metadata][beat]}-%{+YYYY.MM.dd}"
    user => "elastic"
    password => "${password}"
#    cacert => "/etc/logstash/http_ca.crt"
    ssl_certificate_verification => false
  }
```

FIGURE 2.16 : Configuration de l'output de Logstash

On remarque que la sortie des données de Logstash est dirigée vers le port 9200 de la machine sur laquelle Elasticsearch est installé.
Un second output est configuré pour l'envoi d'e-mails à une adresse spécifiée lorsque certaines conditions sont remplies. L'output en question est "email", comme le montre la figure 2.17.

```
if ("ids" in [tags]) or ("vpn" in [tags]) {
   if "throttled" not in [tags] {
     email {
      to => 'cactitest13@gmail.com'
      from => 'elk@gmail.com'
      subject => "Alert"
      body => "
      Alert Details:
      ---------------
      Timestamp: %{[@timestamp]}
      Source: %{[src_ip]}:%{[src_port]}
      Destination: %{[dst_ip]}:%{[dst_port]}
      Message: %{[message]}"
     # body => "Détails de l'alerte : %{message}"
    #template_file => "/tmp/email_template.mustache"
    #domain => 'mail.example.com'
      via => smtp
      address => "smtp.gmail.com"
      port => 587
      username => "cactitest13@gmail.com"
      password => "${mail_password}"
      authentication => "plain"
      use_tls => true
     }
   }
 }
}
```

FIGURE 2.17 : Configuration pour l'envoi des mails

Le mail sera envoyé si l'un des tags "IDS" ou "VPN" est présent, à condition que le tag "throttled" ne soit pas associé à l'événement. Le message arrivera à l'adresse *cactitest13@gmail.com*, avec comme expéditeur *elk@gmail.com* avec pour objet "Alert". Le corps du mail contiendra les informations suivantes :

- la date ainsi que l'heure à laquelle l'événement s'est produit ;
- l'adresse source ainsi que le port source ;
- l'adresse de destination ainsi que le port de destination ;
- le message généré pour l'évènement.

Les conditions définies dépendent du tag de l'événement, lequel a été attribué lors du filtrage de Logstash. Par exemple, la figure 2.18 montre le paramètre permettant d'ajouter un tag.

```
# Pour les logs VPN
if [message] =~ /openvpn/ and ([event][severity] < 3) {
  mutate {
    add_tag => ["vpn"]
  }

  throttle {
    key => "[@metadata][throttle_key]"
    before_count => 3
    after_count => 5
    period =>3600  # Limite les emails à un toutes les 60 secondes pour les m>
    max_age => 7200
    add_tag => "throttled"
  }

}
```

FIGURE 2.18 : Exemple de configuration du plugin throttle

Si le mot "openvpn" est trouvé dans le champ "message" de l'événement et que le numéro de la priorité syslog est en dessous de 3 (indiquant une situation urgente), le tag "VPN" est ajouté grâce au plugin *mutate*. Par la suite le plugin *throttle* entre en jeu. Il sert à limiter la fréquence des actions répétitives, comme l'envoi de mails. Dans ce cas, si trois événements identiques se produisent dans un délai d'une heure (3600 secondes), cinq mails supplémentaires sont autorisés, ce qui porte le total à huit en une heure. Tous les événements identiques suivants sont marqués du tag "throttled". Ainsi, lors de l'envoi d'un mail, si le tag "throttled" est absent, l'email sera envoyé. Après une durée maximale de deux heures (7200 secondes), le compteur est réinitialisé, permettant de relancer le processus si de nouveaux événements similaires surviennent. Cette configuration permet de contrôler strictement le flux de mails tout en évitant de saturer la boîte mail.

Aucune configuration particulière n'a été effectuée pour Kibana et Elasticsearch. Cependant, pour permettre leur accès sur un navigateur autre que celui de la machine sur laquelle ils sont déployés, l'adresse de la machine a été spécifiée.
Pour Kibana, le fichier de configuration est kibana.yml, présent dans le répertoire /etc/kibana, fichier dans lequel l'adresse IP a été renseignée, comme le montre la figure 2.19.

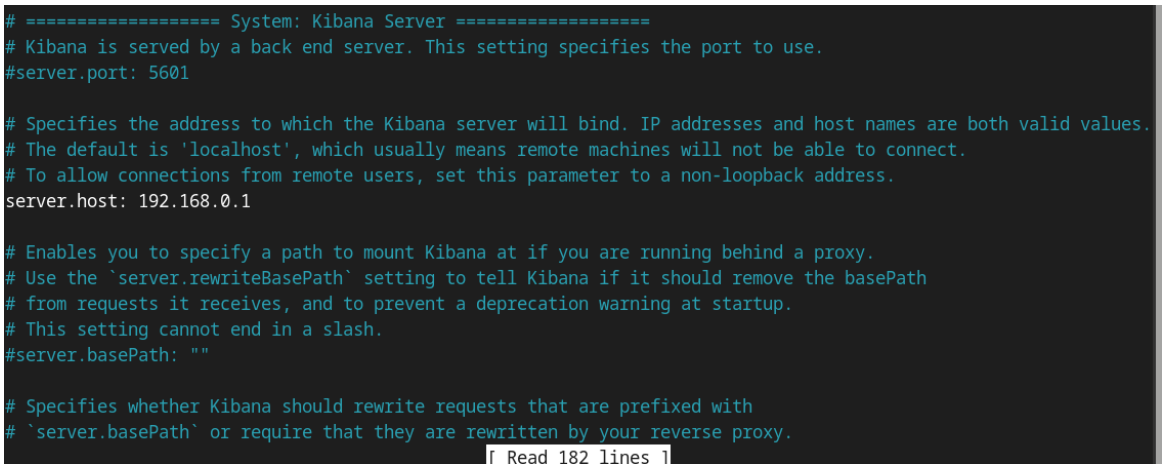


FIGURE 2.19 : Configuration de l'adresse IP de Kibana

Pour Elasticsearch, le fichier de configuration est elasticsearch.yml présent dans /etc/elasticsearch, et la configuration faite à ce niveau est présentée sur la figure 2.20.

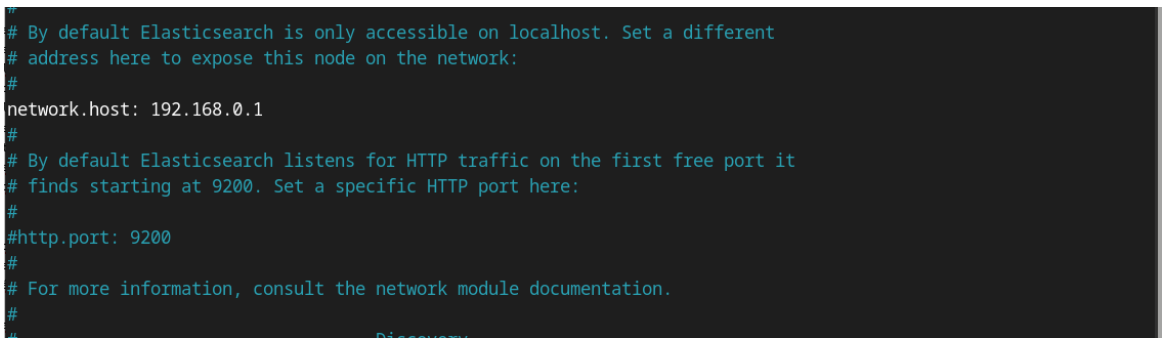


FIGURE 2.20 : Configuration de l'adresse IP de Elasticsearch

La suite ELK est désormais entièrement opérationnelle et prête à répondre efficacement aux besoins.

Dans l'optique de garantir la récupération des données collectées en cas de panne, de suppression accidentelle ou de migration vers un autre serveur. Elasticsearch permet la création des **"snapshots"** (instantanés) qui permettent de sauvegarder et de restaurer les données des index. Ils capturent l'état des index et des métadonnées dans un **repository** (référentiel) de stockage (local, Google cloud, Azure, etc) et peuvent être programmés pour des sauvegardes régulières. Bien que les instantanés enregistrent les données, ils ne prennent pas en compte les fichiers de configuration d'Elasticsearch, Logstash ou Kibana. Pour une reprise complète après un incident, il est donc recommandé de coupler les instantanés avec une sauvegarde manuelle des configurations. Cette action peut être effectuée au niveau de Kibana en se rendant dans l'onglet *"Stack Management"* et ensuite *"Snapshot et Restore"*. Les figures 2.21 et 2.22 suivantes montrent les onglets en question dans un cadre rouge.

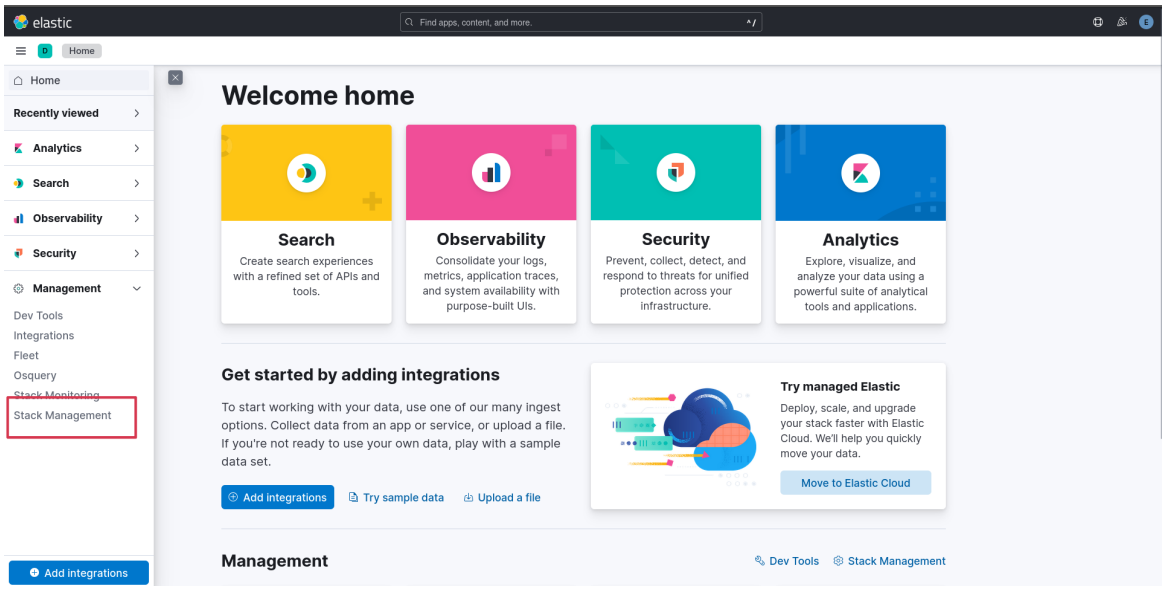


FIGURE 2.21 : Stack Management

Après avoir cliqué sur *"Stack Management"*, nous arrivons sur cette fenêtre (figure 2.22) où dans le cadre rouge on voit *"Snapshot and Restore"*, au niveau duquel sont effectuées les sauvegardes et les restorations.

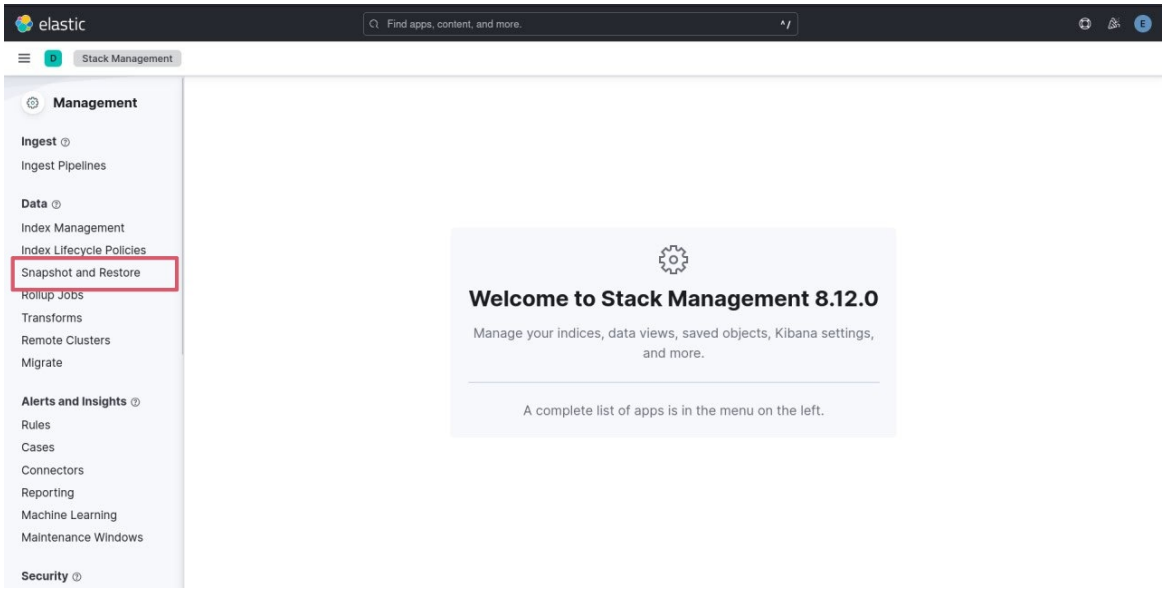


FIGURE 2.22 : Emplacement de "Snapshot et Restore"

Conclusion

Au cours de cette section, une étude comparative de quelques SIEM a été effectuée, afin d'identifier celui qui serait le plus approprié pour le travail. A l'issue de cette étude, l'Elastic Stack s'est révélé être l'outil de choix. Par la suite, l'outil a été déployé dans une architecture réseau définie au préalable. Le chapitre suivant est consacré à la présentation des tests ainsi que des résultats obtenus.

Chapitre 3

Tests et Résultats

Introduction

Dans ce chapitre, les tests réalisés sont présentés dans un contexte de mise en situation réelle avec un utilisateur externe au réseau d'entreprise. L'objectif est d'observer le comportement de la solution et d'évaluer son efficacité dans le cadre de la surveillance.

3.1 Test de validation

Une fois la configuration terminée, il est crucial de vérifier que toutes les étapes ont été correctement réalisées, en s'assurant que les données sont bien transmises des Beats vers Logstash, puis de Logstash vers Elasticsearch. Pour cela, il suffit de vérifier dans Kibana si les index des données présentes dans Elasticsearch sont bien reçus. Ces index peuvent être consultés en accédant à la page *"Index Management"*, disponible dans *"Stack Management"* dans le menu à gauche, affiché sur la figure 3.1.

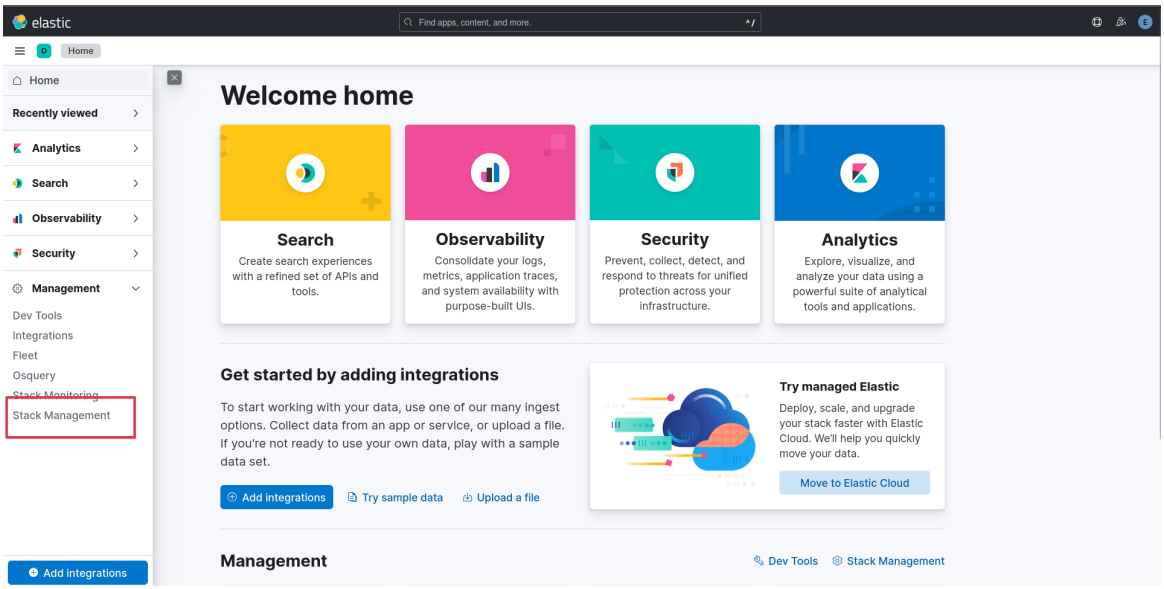


FIGURE 3.1 : Emplacement de l'index management

L'onglet "Index Management" se trouve dans le menu de gauche de *Management* dans la partie *Data*. En cliquant dessus, la page suivante s'affiche (figure 3.2).

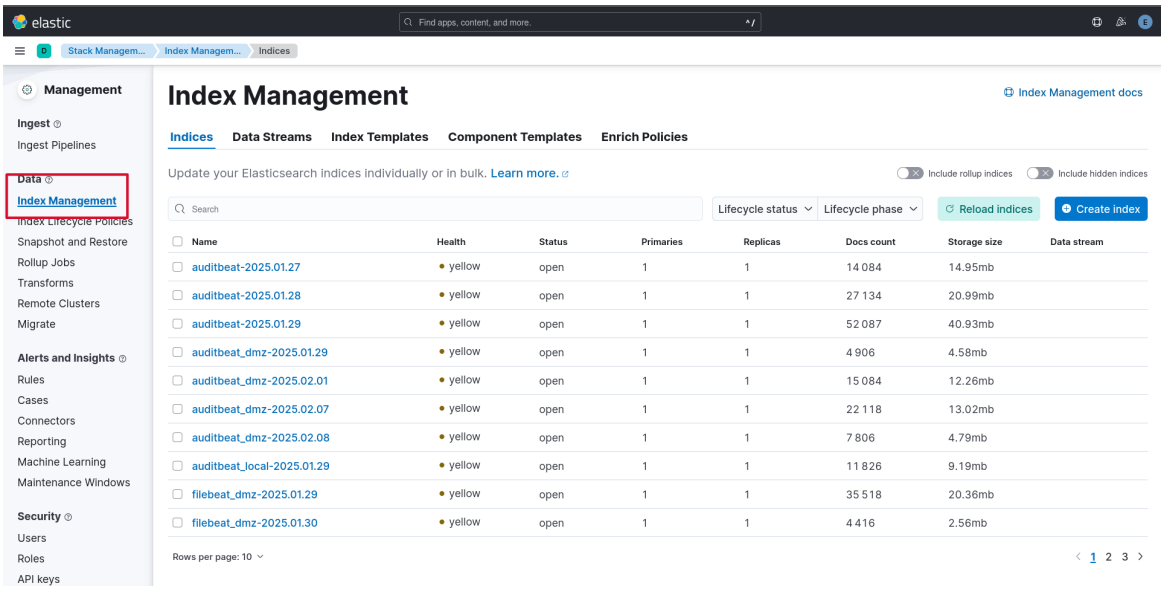


FIGURE 3.2 : Liste des indexs

L'ensemble des index envoyés à Elasticsearch par Logstash chaque jour sont affichés, ce qui confirme le bon déroulement du processus de collecte et de réception des données.

Pour visualiser les données dans *Discover*, il est nécessaire de créer des vues de données pour les indexs. Pour ce faire, il suffit de cliquer sur *'Data view'* dans le menu de gauche, comme illustré sur la figure 3.3.

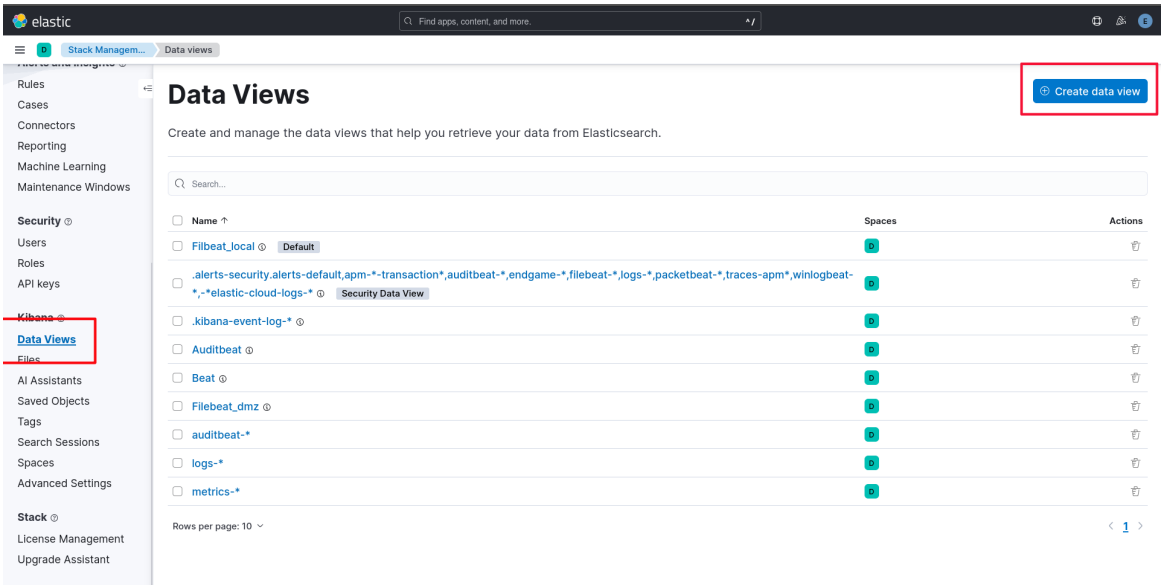


FIGURE 3.3 : Interface des vues de données

Sur cette image, le bouton *'Create data view'* est mis en évidence par un cadre bleu à droite. En cliquant dessus, l'interface montrée dans la figure 3.4 apparaît.

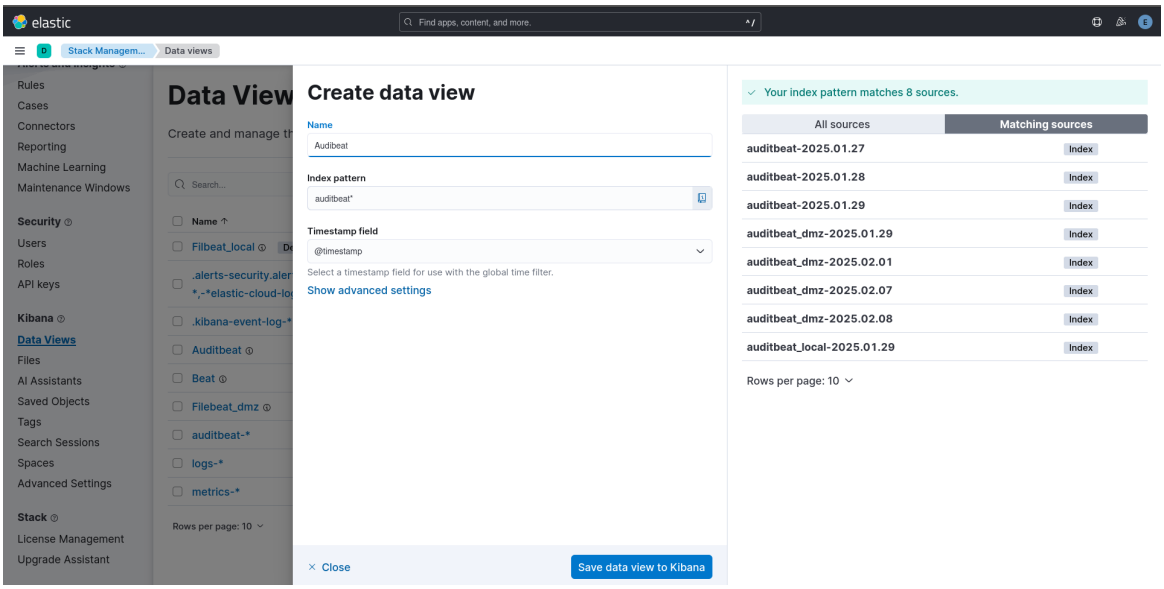


FIGURE 3.4 : Création d'une vue de données

À ce stade, il suffit de donner un nom au data view, de sélectionner l'index correspondant parmi ceux disponibles, puis de l'enregistrer. Une fois la création effectuée, les données associées à la vue de données sont affichées dans l'onglet *'Discover'* du menu de gauche, sur la page d'accueil de Kibana, comme le montre la figure 3.5. Elle présente également les données recueillies par Filebeat.

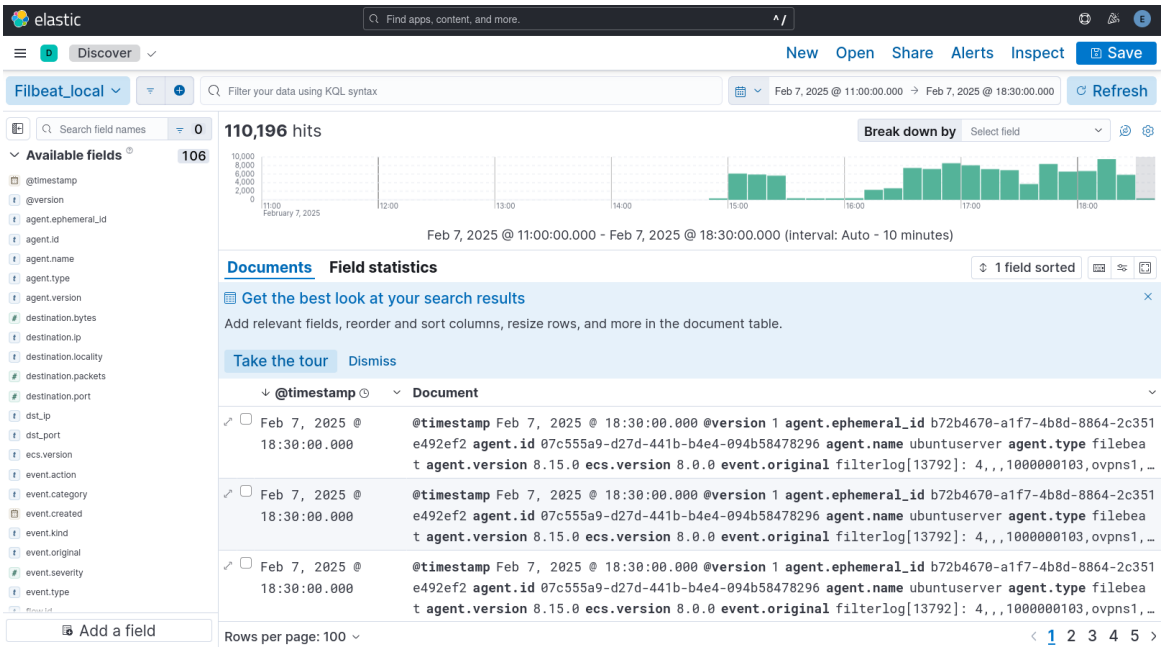


FIGURE 3.5 : Vue de données de Filebeat

La figure 3.6 affiche les données obtenues par Auditbeat.

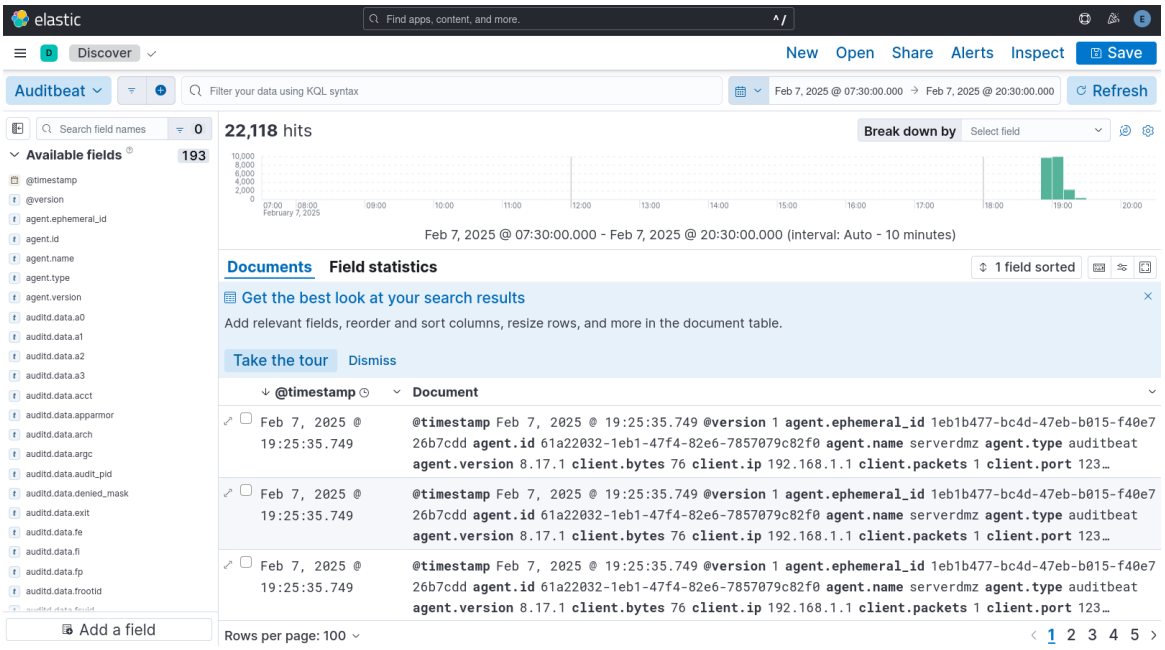


FIGURE 3.6 : Vue de données d'Auditbeat

L'un des objectifs visés, est d'avoir un suivi du flux réseau des utilisateurs distants. La figure 3.7 l'illustre parfaitement.

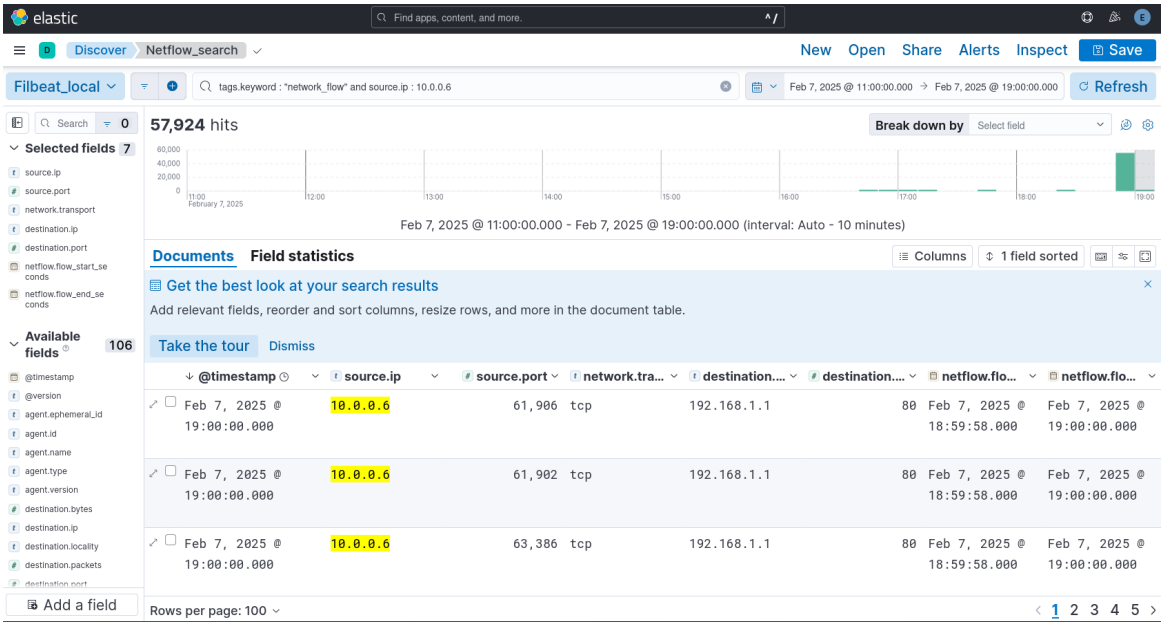


FIGURE 3.7 : Flux réseau d'un utilisateur distant

Dans ce travail, le réseau 10.0.0.0/24 est dédié au tunnel VPN. L'image nous montre l'adresse IP source, ici celle de l'utilisateur externe, la destination qui est la machine ayant l'adresse 192.168.1.1 sur le port 80. Cette recherche, élargit sur un temps plus grand, donnera le trafic réseau émis de l'utilisateur.

Concentrons-nous sur l'identification des utilisateurs connectés et leurs activités au sein du réseau. Il est essentiel de pouvoir retracer ces connexions. La figure 3.8 et la figure 3.9 illustrent la présence des utilisateurs actifs dans le réseau.

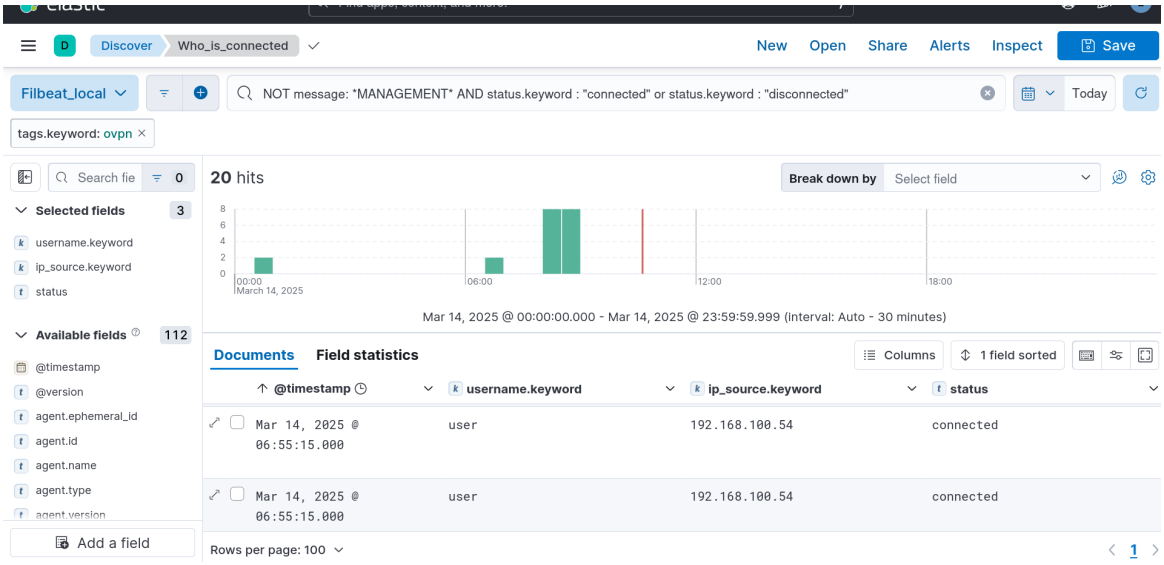


FIGURE 3.8 : Heures d'activité de l'utilisateur distant "*user*"

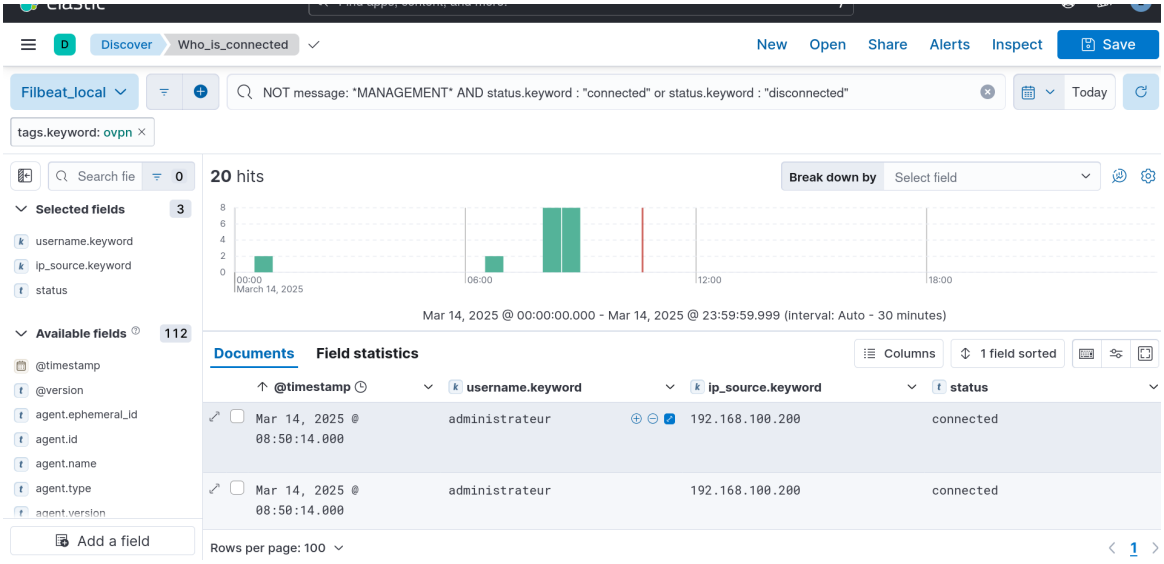


FIGURE 3.9 : Heures d'activité de l'utilisateur distant "*administrateur*"

D'après ces figures, deux utilisateurs se sont connectés : *user* à 06h55 et *administrateur* à 08h50.
Pour une analyse plus approfondie, nous pouvons examiner les actions effectuées sur un serveur suite à une connexion à distance. La figure 3.10 montre une partie du dahsboard créé à cet effet qui fournit des informations sur les utilisateurs distants ainsi que sur les serveurs internes auxquels ils ont accédé.

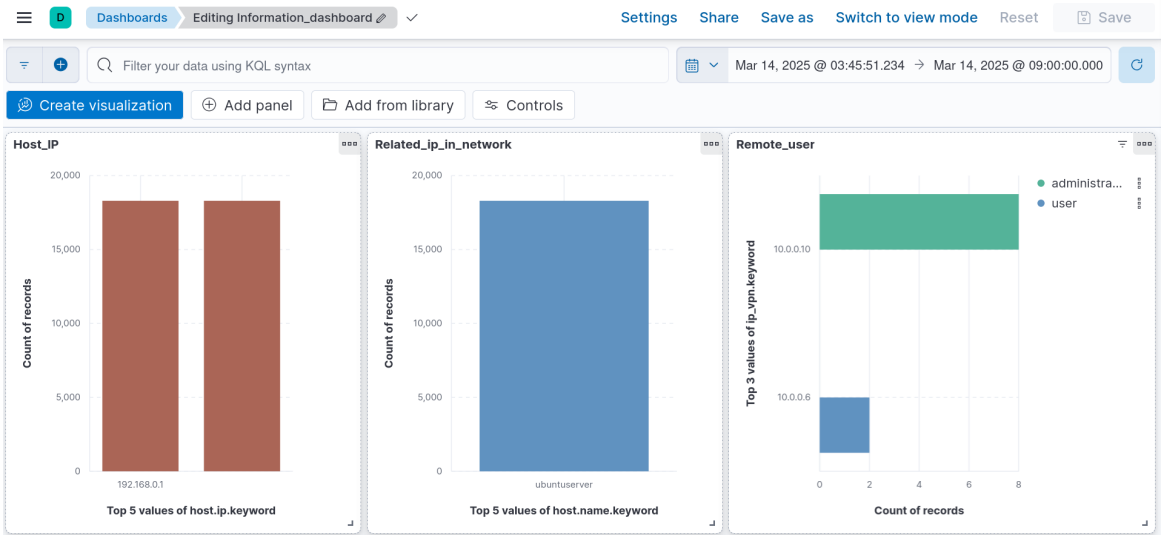


FIGURE 3.10 : Dashboard

Sur cette figure (3.10), on voit les utilisateurs distants connectés ainsi que l'adresse IP qui leur a été attribuée dans le tunnel. L'utilisateur *user* possède l'adresse 10.0.0.6, tandis que l'utilisateur *administrateur* a l'adresse 10.0.0.10. On retrouve également les adresses des machines du réseau qui ont été jointes, ainsi que le nom d'hôte desdites machines. La figure 3.11 montre les processus qui ont été exécutés.

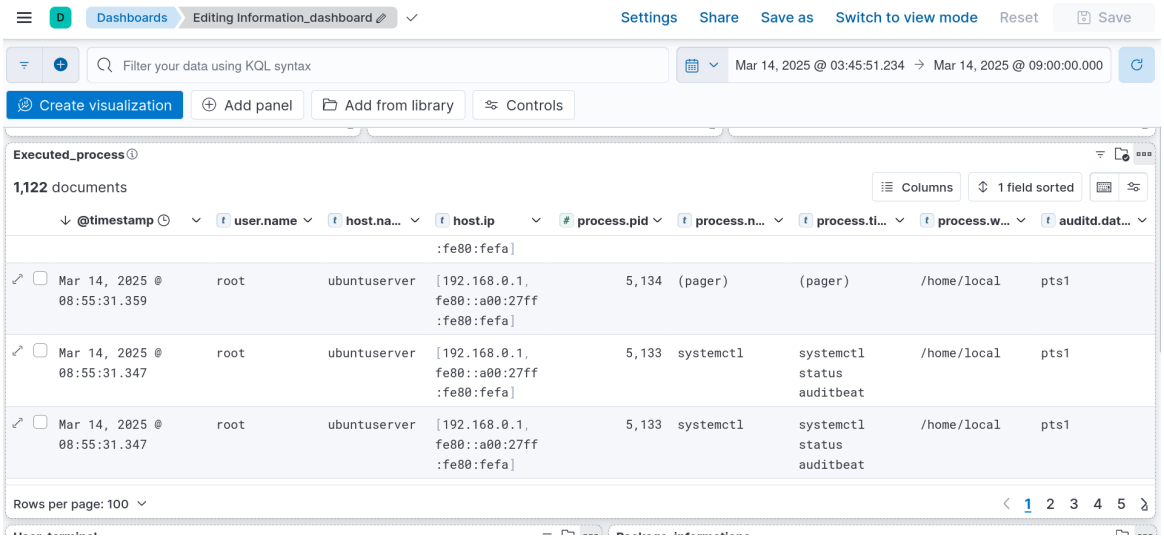
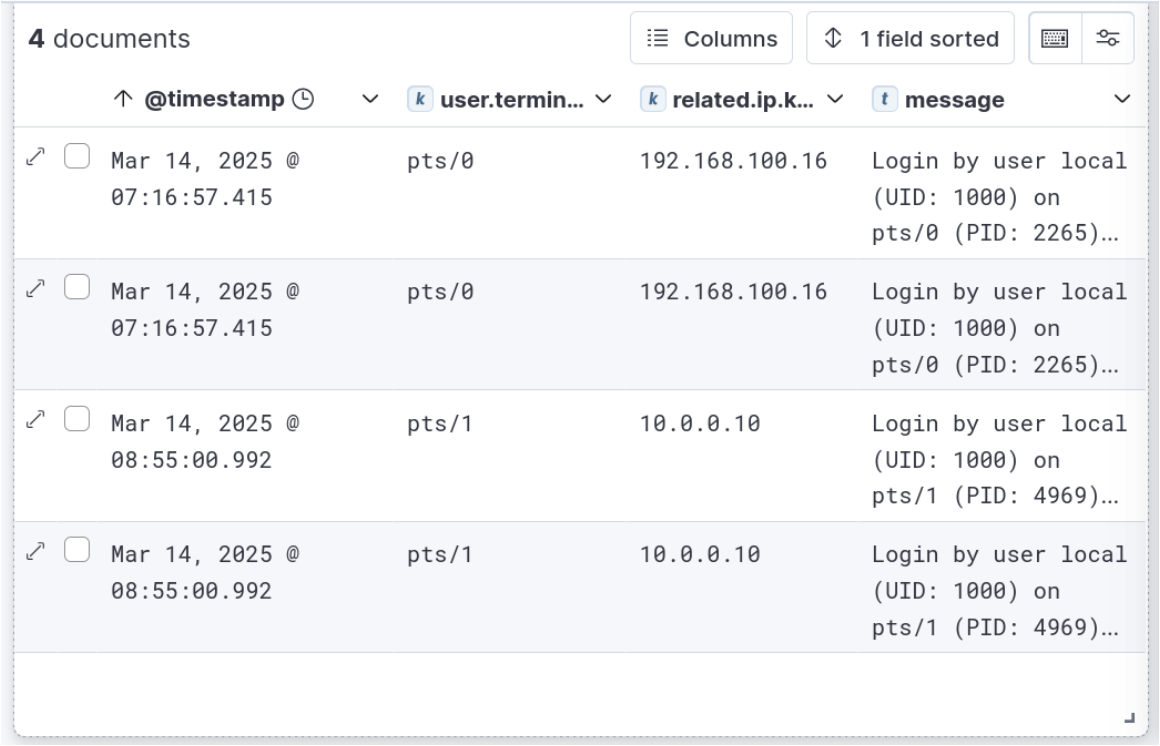


FIGURE 3.11 : Processus exécutés

On constate qu'à 09h34, le 14 mars 2025, la commande *systemctl status auditbeat* a été exécutée à distance depuis le terminal "pts1" sur une machine avec l'adresse IP 192.168.0.1.
Afin d'identifier précisément l'utilisateur responsable, nous pouvons analyser le tty attribué aux connexions distantes. La figure 3.12 met en évidence ces informations.



4 documents

Columns | 1 field sorted

@timestamp	user.termin...	related.ip.k...	message
Mar 14, 2025 @ 07:16:57.415	pts/0	192.168.100.16	Login by user local (UID: 1000) on pts/0 (PID: 2265)...
Mar 14, 2025 @ 07:16:57.415	pts/0	192.168.100.16	Login by user local (UID: 1000) on pts/0 (PID: 2265)...
Mar 14, 2025 @ 08:55:00.992	pts/1	10.0.0.10	Login by user local (UID: 1000) on pts/1 (PID: 4969)...
Mar 14, 2025 @ 08:55:00.992	pts/1	10.0.0.10	Login by user local (UID: 1000) on pts/1 (PID: 4969)...

FIGURE 3.12 : Terminal de l'utilisateur

À 08h55, *administrateur* s'est connecté par ssh, et le tty "pts1" lui a été attribué. Aucun log n'indique une déconnexion avant 09h34, ce qui signifie que c'est de sa machine que la commande *systemctl status auditbeat* a été exécutée.

Grâce à ELK, nous pouvons ainsi identifier les utilisateurs connectés, surveiller les services actifs dans le réseau et tracer les actions effectuées lors des connexions SSH. Cela permet d'assurer une traçabilité des accès distants et de détecter une activité suspecte. Malgré les précautions mises en place, un attaquant pourrait tenter d'accéder aux ressources internes via le tunnel VPN. Il est donc primordial de mettre en place des mécanismes d'alerte pour détecter toute activité anormale et identifier rapidement l'origine d'une éventuelle attaque. Pour illustrer cela, nous avons reproduit deux scénarios d'attaque dans la section suivante.

3.2 Test de réaction à une attaque par force brute

Des attaques ont été simulées, notamment une par brute force depuis un client VPN. Une attaque par force brute est technique qui sert à obtenir les identifiants d'accès (nom d'utilisateur, mot de passe) à un système. Elle consiste à effectuer un nombre très élevé de combinaisons possibles jusqu'à trouver les bons identifiants.
La figure 3.13 montre les alertes générées par l'IDS Snort suite à cette tentative, visualisées dans Kibana.

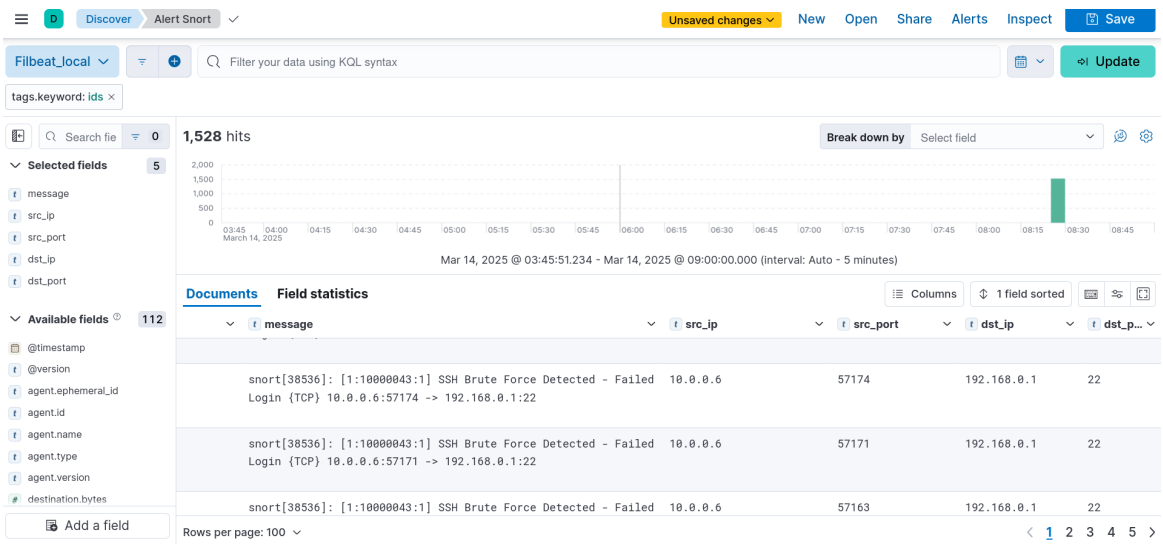


FIGURE 3.13 : Alertes de l'IDS Snort pour l'attaque par brute force

On identifie ici la machine avec comme adresse IP le 10.0.0.6 la source du brute force à l'encontre du serveur local (IP :192.168.0.1) sur le port 22 faisant référence au protocole SSH.

3.3 Test de réaction à une attaque par deni de service

Une autre attaque a été simulée. Cette fois-ci, il s'agit d'un deni de service pour affecter le service web sur le serveur de la DMZ. Une attaque par déni de service est une technique utilisée pour rendre un service indisponible en surchargeant un serveur avec un grand nombre de requêtes jusqu'à ce que le serveur ne puisse plus répondre aux utilisateurs légitimes. Ces requêtes peuvent provenir d'une seule machine qui envoie un flux massif de requêtes ou de plusieurs machines à la fois. Dans le contexte actuel, une seule machine a été utilisée. La figure 3.14 montre les alertes générées par l'IDS Snort suite à cette tentative, visualisées dans Kibana.

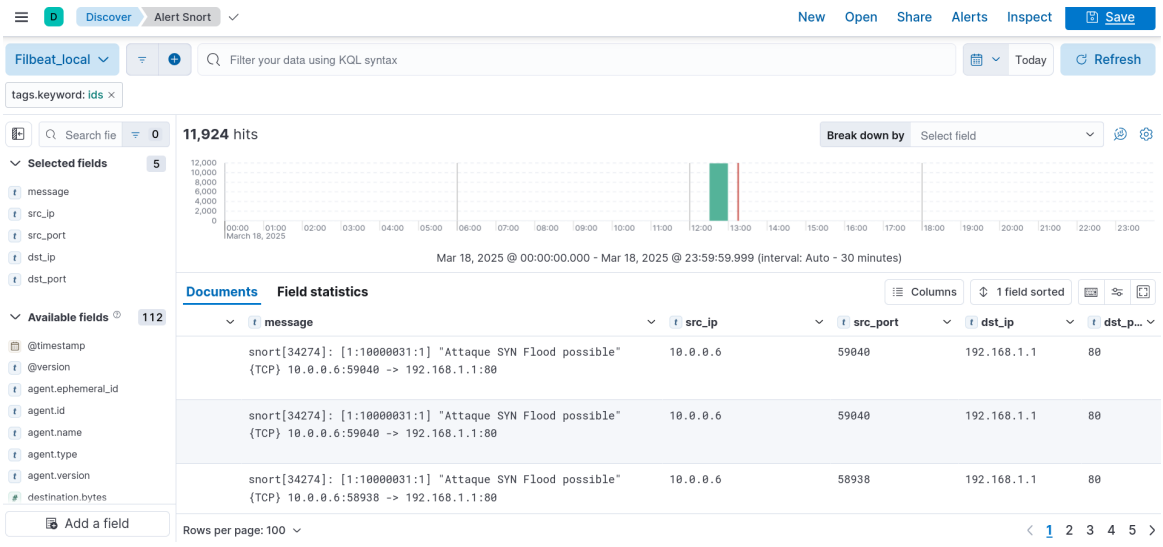


FIGURE 3.14 : Alertes de l'IDS Snort pour le deni de service

On identifie ici la machine avec comme adresse IP le 10.0.0.6 la source de l'attaque en direction du serveur qui a comme adresse IP le 192.168.1.1 sur le port 80.

Pour informer l'administrateur de la présence d'un éventuel problème, des mails ont été envoyés pour chacune des attaques. Voici sur la figure 3.15 le mail reçu pour l'alerte en rapport avec l'attaque par brute force.

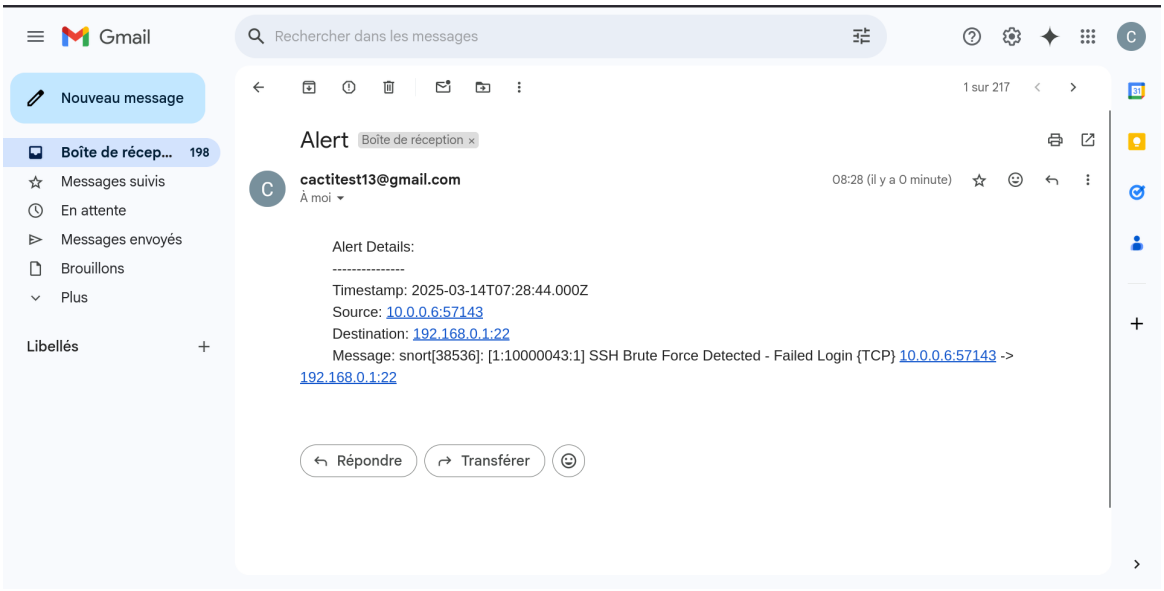


FIGURE 3.15 : Mail d'alerte sur le système

3.4 Discussion

L'implémentation de la solution Elastic Stack a été réalisée dans l'optique de répondre à la problématique du suivi des accès distants dans le cadre du télétravail. Elle a permis de collecter et de corréler les logs provenant de diverses sources, notamment les serveurs, le pare-feu et l'IDS. Grâce à l'intégration des données issues des logs VPN et des alertes de l'IDS, la solution a offert une visibilité sur les actions effectuées par les utilisateurs distants et a permis d'être informé en temps réel des anomalies détectées au sein de l'infrastructure.
Cependant, l'identification d'une menace constitue uniquement la première étape d'une réponse efficace aux incidents. Une fois le responsable de l'attaque identifié, plusieurs actions doivent être entreprises pour limiter les risques et protéger l'infrastructure comme :

- bloquer et restreindre l'accès de l'utilisateur concerné via le pare-feu ou le VPN ;
- faire une analyse approfondie et une investigation des événements corrélés pour comprendre l'étendue de l'attaque ;
- mettre en place des contre-mesures pour appliquer des correctifs de sécurité si l'attaque exploite une faille dans le système.

Bien que Elastic Stack fournisse une visibilité sur les événements et facilite la détection des menaces, il ne dispose pas nativement d'un mécanisme de réponse automatisée aux incidents. Une intégration avec un système de réponse aux incidents, pourrait être envisagée pour automatiser les réactions face aux attaques et améliorer la réactivité des équipes de sécurité.

Conclusion

Dans ce chapitre, les résultats des tests de simulation d'attaques ont été exposés, notamment le brute force et le déni de service. Après avoir vérifié la présence effective des données collectées sur les différentes machines dans Elasticsearch et leur visualisation dans Kibana, on peut remarquer que les différents objectifs fixés ont été atteints. Ces tests ont démontré l'efficacité d'Elastic Stack dans la surveillance des accès distants.

Conclusion Générale

Pour des raisons diverses, de nombreuses entreprises optent pour le travail à distance. Cependant, l'identité réelle et les intentions des utilisateurs demeurent questionnables. Il est donc important d'avoir des informations et une vision plus ou moins claire sur les actions de ces derniers. Une méthode adéquate à la situation est d'utiliser un SIEM pour centraliser et analyser les logs afin de recueillir des informations pertinentes sur les activités des employés. Le travail a donc été de déployer la suite ELK, qui à l'issue de l'évaluation s'est révélée être la solution appropriée. Avec cet outil, des données sur le trafic réseau ont pu être obtenues, de même que l'identification des différentes machines et des services accédés dans l'infrastructure ainsi que les commandes exécutées à distance sur les serveurs. Pour que l'administrateur puisse réagir rapidement en cas de problèmes, un processus d'envoi des mails avec un plugin de Logstash a été mis en place. Bien que la suite Elastic soit un élément clé dans ce système de surveillance, elle ne suffit pas en soi. La complémentarité avec d'autres outils est nécessaire pour garantir une traçabilité complète couvrant tous les utilisateurs, internes comme externes.

En perspectives, il est prévu :

- l'utilisation d'un pare-feu plus performant comme Fortigate que nous devrions mettre en place, mais qui pour des problèmes d'installation et de licence n'a pas été utilisé ;
- l'installation d'un système d'analyse des paquets afin de détecter efficacement une attaque avant qu'elle n'atteigne la cible ;
- l'intégration de l'outil Watcher qui permet de créer des alertes personnalisables ;
- la mise en place du NAT dans le but de masquer la réelle adresse IP des serveurs pour plus de sécurité ;
- la mise en place d'un système de réponse aux incidents en temps réel.

Bibliographie

[1] Harish Chandra Chandan, *Impact of Teleworking and Remote Work on Businesss*, IGI GLobal, 2024

[2] Jagdish Krishanlal Arora, *Cyber Security*, Jagdish Krishanlal Arora, 2024

Webographie

[1] Le robert, définition du télétravail, https://dictionnaire.lerobert.com/definition/teletravail ; consulté le 10/02/25.

[2] Opensourcing, avantages du télétravail, https://www.opensourcing.com/ressources/blog/le-teletravail-avantages-et-inconvenients ; consulté le 03/02/25.

[3] Nowteam, les avantages et les inconvénients du télétravail, https://www.nowteam.net/le-teletravail-en-entreprise-les-avantages-et-inconvenients/ ; consulté le 03/02/25.

[4] Splunk, qu'est-ce qu'un siem ?, https://www.splunk.com/fr_fr/blog/learn/siem-security-information-event-management.html?301=/fr_fr/data-insider/what-is-siem.html ; consulté le 05/02/25.

[5] Mca concept, les solutions open sources et propriétaires, https://mca-concept.com/open-source-proprietaire/ ; consulté le 05/02/25.

[6] Aws, qu'est-ce que la suite elk, https://aws.amazon.com/what-is/elk-stack/?utm ; consulté le 19/02/25.

[7] Wazuh, documentation de wazuh, https://documentation.wazuh.com/current/index.html ; consulté le 15/02/25.

[8] Graylog, communauté de graylog, https://community.graylog.org/t/graylog-query-using-time/21494?utm=true ; consulté le 19/02/25.

[9] Gartner, elastic vs graylog vs wazuh, https://www.gartner.com/reviews/market/security-information-event-management/compare/elastic-vs-graylog-vs-wazuh ; consulté le 19/02/25.

[10] Elastic, le flux de données dans elk, https://www.elastic.co/guide/en/beats/libbeat/8.17/beats-reference.html ; consulté le 03/02/25.

Table des matières

Dédicace ii

Remerciements iii

Résumé iv

Abstract v

Table des figures vi

Liste des tableaux viii

Liste des acronymes ix

Glossaire x

Introduction 2

1 Revue de littérature 4
Introduction 4
1.1 Télétravail 4
1.1.1 Définition 4
1.1.2 Avantages et inconvénients 4
1.1.3 Le facteur sécurité 5
1.2 Journalisation des activités 6
1.2.1 Définition et présentation d'un SIEM 7
1.2.2 Présentation de quelques SIEM 7
1.2.2.1 Elastic Stack 7
1.2.2.2 Graylog 7
1.2.2.3 Wazuh 8
Conclusion 8

2 Choix et mise en place de la solution 10
Introduction 10
2.1 Choix de la solution 10
2.1.1 Présentation des critères et de la méthode d'évaluation des solutions 10
2.1.2 Comparaison de quelques SIEM 11
2.2 Mode de fonctionnement de la solution 12
2.2.1 La suite Elastic 12
2.2.2 Présentation des différents beats 13

2.3 Déploiement de la solution . . . 14
2.3.1 Installation de la solution . . . 15
2.3.2 Configuration de la suite . . . 20
Conclusion . . . 28

3 Tests et Résultats . . . 29
Introduction . . . 29
3.1 Test de validation . . . 29
3.2 Test de réaction à une attaque par force brute . . . 35
3.3 Test de réaction à une attaque par deni de service . . . 36
3.4 Discussion . . . 37
Conclusion . . . 37

Conclusion . . . 39

Bibliographie . . . 40

Webographie . . . 41

Table des matières . . . 42