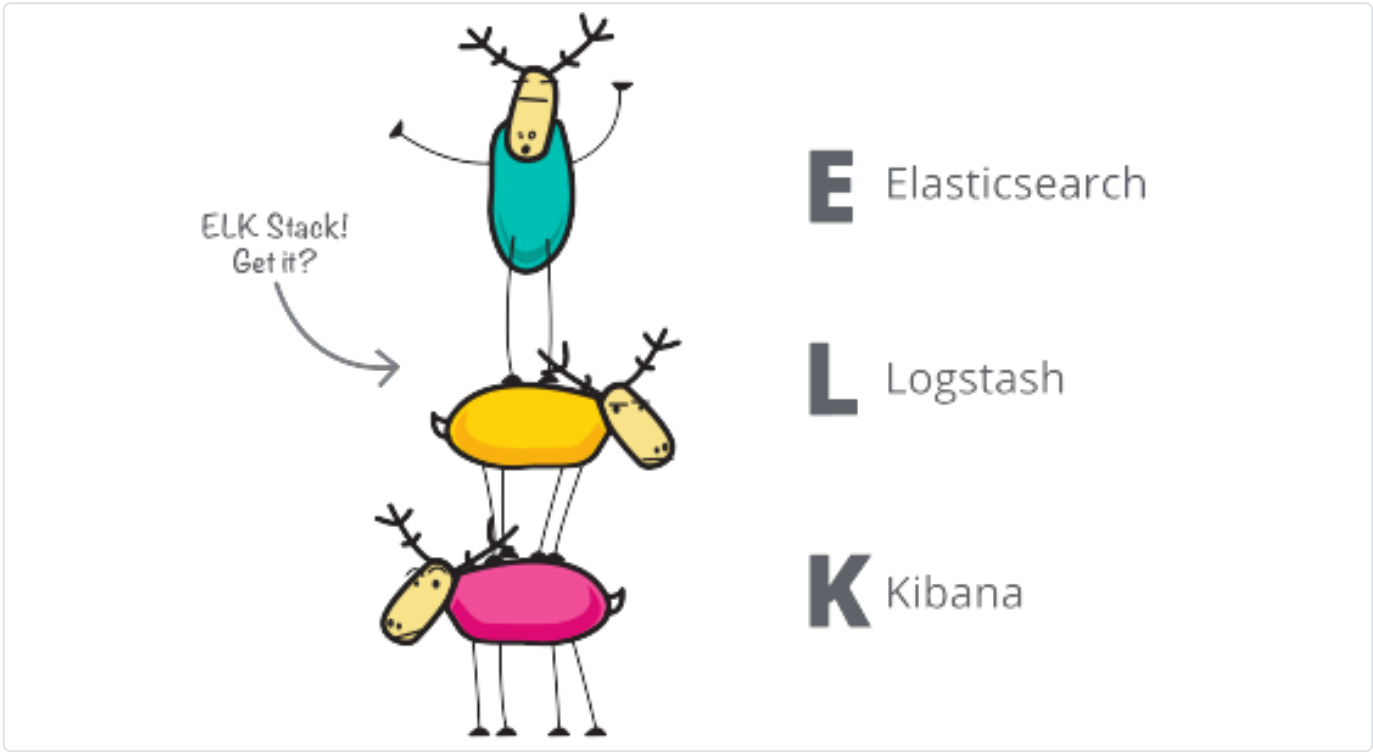
# INTRODUCTION AU COURS ELK

## Introduction

Hello les champions, ça vous tente de superviser vos logs de manières différentes ? Je suppose que la réponse est oui ! Dans ce guide, nous allons découvrir **pourquoi l'analyse de log connaît-il un tel intérêt** et **nous allons jeter un regard complet sur les différentes technologies composant la pile** ? Le but est de comprendre quel rôle ils jouent dans vos analyses de données. Enfin dans les futurs chapitres de ce cours nous verrons plus en détail comment les installer, les configurer et les utiliser et comment éviter au mieux certains pièges courants en cours de route.




## Public visé

Cette série d'articles est conçue pour les débutants ayant besoin de comprendre l'utilisation de la suite ELK à partir de zéro. Ce tutoriel vous donnera une

compréhension suffisante de la technologie, qui vous permettra plus tard d'atteindre des niveaux d'expertise beaucoup plus élevés.

## Prérequis

Ce cours ne demande pas forcément de prérequis que vous devriez au minimum avoir. Cependant si vous maîtrisez déjà ou que vous avez une compréhension générale d'une solution de supervision et que vous êtes à l'aise avec l'environnement Linux, alors il vous sera très facile de comprendre les différents concepts de la suite ELK et d'avancer rapidement sur votre piste d'apprentissage, mais ce n'est pas non plus indispensable !

Sans plus attendre Let's go !

# L'analyse de logs

Les logs sont l'une des informations les plus précieuses en matière de gestion et de surveillance des systèmes informatiques. Comme ils enregistrent toutes les actions qui ont eu lieu sur une machine, ils fournissent des informations dont vous avez besoin pour repérer les problèmes susceptibles d'avoir un impact sur les **performances**, la **conformité** et la **sécurité** des éléments de votre infrastructure. C'est pourquoi la gestion des logs devrait faire partie de toute infrastructure de surveillance.

Lorsqu'on parle d'analyse de logs, le plus important défi consiste à regrouper, normaliser, visualiser et analyser vos logs dans un emplacement unique et accessible, d'où l'utilisation de la suite ELK.

## Qu'est-ce que l'analyse des logs?

L'analyse de logs est le processus consistant à donner du sens aux messages de logs générés par votre système dans le but d'utiliser ces données pour améliorer ou résoudre des problèmes de performances au sein d'une application ou d'une infrastructure. Dans une vue d'ensemble, les entreprises analysent les logs pour atténuer les risques de manière proactive et réactive, se conformer aux politiques de sécurité, aux audits et aux réglementations, et permet également de mieux comprendre le comportement de vos utilisateurs utilisant vos applications.

## Pourquoi l'analyse des logs est-elle importante?

La plupart des entreprises sont tenues responsable d'effectuer l'archivage et l'analyse des logs dans le cadre de leurs réglementations de conformité. Ils doivent régulièrement surveiller et analyser les logs des systèmes pour rechercher des erreurs, des anomalies ou des activités suspectes ou non autorisées qui s'écartent de la norme. Cela leur permet de recréer la chaîne d'événements qui a conduit à un problème et de le résoudre efficacement.

Par exemple, vous pourrez rechercher des erreurs HTTP et comprendre où et pourquoi elles se sont produites ou détecter si les utilisateurs ne reçoivent pas les informations recherchées ou si le chargement de leurs demandes prend trop de temps, ou si certains microservices rencontrent des problèmes, etc ...

Lorsque vous tirez parti de l'analyse des logs, vous pouvez détecter les problèmes avant ou lorsqu'ils surviennent et éviter le gaspillage de temps, les retards inutiles et les coûts supplémentaires qui vont avec. Ainsi, les équipes peuvent intervenir et résoudre les problèmes plus rapidement, leurs permettant ainsi de se concentrer davantage sur l'amélioration des fonctionnalités existantes et sur l'ajout de nouvelles fonctionnalités aux produits et services qu'ils créent au lieu de passer du

temps à dépanner manuellement. Cela, à son tour, augmente la valeur des logiciels qu'ils construisent et conduit à des versions plus fréquentes (DevOps) et augmente la valeur globale pour l'entreprise.

## Meilleures pratiques d'analyse des logs

L'analyse des logs est un processus complexe qui doit suivre les bonnes pratiques suivantes :

- **La détection et la reconnaissance des patterns** : cela fait référence au filtrage des messages entrants sur la base d'un pattern souvent combiné avec l'utilisation des expressions régulières. Elle fait partie intégrante de l'analyse des logs car elle permet de détecter plus rapidement les anomalies.
- **La normalisation des logs** : ce sont le processus de conversion des éléments de logs tels que les adresses IP ou les horodatages, dans un format commun pour toutes vos équipes.
- **La classification et le balisage** : ce sont le processus de balisage (tag) des messages avec des mots-clés et de leur catégorisation en classes. Cela vous permet de filtrer et de personnaliser la façon dont vous visualisez les données.
- **L'analyse de corrélation** : fait référence à la collecte de données provenant de différentes sources et à la recherche de messages appartenant à un événement spécifique. Par exemple, en cas d'activité malveillante, il vous permet de filtrer et de corréler les logs provenant de vos périphériques réseaux, pare-feu, serveurs et autres sources afin de très rapidement détecter la source du problème.

- **Alerte** : L'analyse de corrélation est généralement associée aux systèmes d'alerte, en fonction du pattern que vous avez identifié, vous pouvez créer des alertes lorsque votre analyseur de logs détecte une activité anormale.

# ELK

## C'est quoi ELK ?

La stack ELK est un acronyme utilisé pour décrire une pile qui comprend trois projets open sources populaires : Elasticsearch, Logstash et Kibana. Elle est la principale **solution open source de gestion des logs pour les entreprises** qui souhaitent bénéficier des avantages d'une **solution de journalisation centralisée**.

En effet, les outils : Elasticsearch, Logstash et Kibana, lorsqu'ils sont utilisés ensemble (car oui on peut les utiliser séparément ou avec d'autres technologies), forment une pile de bout en bout offrant une analyse de données dont des logs en temps réel afin de fournir des informations exploitables à partir de presque tout type de source de données structurée et non structurée. Vous comprendrez plus tard que chacun de ces produits joue un rôle différent dans vos analyses. Ils peuvent être utilisés pour des projets simples ou complexe car ils prennent en charge des opérations simples et avancées.

## Les rôles des composants

Il est important de dissocier le rôle de ces 3 outils afin de mieux comprendre le rôle et la communication de l'ensemble de ces composants :

- **Elasticsearch** : en charge de l'indexation et du stockage de vos informations sur une base de données NoSQL qui est basé sur le moteur de recherche Apache Lucene et il est construit pour fournir des APIS rest. Il offre un déploiement simple, une fiabilité maximale et une gestion facile. Il propose également des requêtes avancées pour effectuer une analyse détaillée et stocke toutes les données de manière centralisée. Il est également utilisé sur de nombreux projets hors la suite ELK car il permet d'exécuter une recherche rapide des documents.
- **Logstash** : outil d'intégration de données open source qui vous permet de collecter des données à partir d'une variété de sources, de les filtrer, les transformer et de les envoyer à la destination souhaitée (ex: Elasticsearch). Son but principal est de rassembler et normaliser tous les types de données provenant de différentes sources et de les rendre disponibles pour une utilisation ultérieure.
- **Kibana** : outil de visualisation de données qui complète la pile ELK , c'est une couche de visualisation qui fonctionne au-dessus d'Elasticsearch, offrant aux utilisateurs la possibilité d'analyser et de visualiser les données récupérées Elasticsearch. C'est un outil puissant offrant pour vos tableaux de bord divers diagrammes interactifs, données géographiques et graphiques pour visualiser les données les plus complexes.

### Information

Récemment, un nouvel outil a été rajouté à la Stack Elastic nommé `Beats` il est très utile si vous utilisez plusieurs machines car ça reste un agent léger et réservé

au transfert de données, il se charge de transférer les données vers Logstash et Elasticsearch. Nous l'utiliserons dans un futur chapitre.

## Comment fonctionnent-ils ensemble ?

Comme il l'est mentionné ci-dessus, prit ensemble, les différents composants de la pile ELK fournissent une solution simple mais puissante pour la gestion et l'analyse des logs.


Logstash sera utilisé pour récupérer, filter et normaliser des informations liées à vos logs issues de différents sources. Une fois cela fait, Logstash envoie ces informations dans un système de stockage ici Elasticsearch qui quant à lui se chargera d'indéxer, stocker et effectuer une recherche et une analyse en temps réel de vos données. Enfin, Kibana récupère ces données afin de fournir un système de visualisation et d'exploration en plus de Logstash et Elasticsearch afin que vous puissiez facilement comprendre vos données sous forme de tableaux et de graphiques.

**Information**

Cependant, pour gérer des pipelines plus complexes conçus pour gérer de grandes quantités de données en production, des composants supplémentaires sont susceptibles d'être ajoutés à votre architecture de journalisation comme par ex : Kafka, RabbitMQ, Redis.

## Études de cas

Voici quelques entreprises populaires qui utilisent la suite ELK :

- NetFlix s'appuie fortement sur la pile ELK. L'entreprise utilise la pile ELK pour surveiller et analyser les logs de sécurité des opérations du service client. Il leur permet d'indexer, de stocker et de rechercher des documents à partir de plus de quinze clusters qui comprennent près de 800 nœuds.
- LinkedIn utilise la pile ELK pour surveiller leurs performances et leur sécurité. L'équipe informatique a intégré ELK à Kafka pour comprendre leur charge en temps réel. Leur opération ELK comprend plus de 100 clusters dans six datacenters différents.
- Medium est une célèbre plateforme de publication de blogs. Ils utilisent la pile ELK pour déboguer leurs problèmes de production. De plus, en utilisant cette pile, la société peut prendre en charge 25 millions de lecteurs uniques ainsi que des milliers de publications publiées chaque semaine.

# Conclusion

Retrouvons-nous dans le prochain chapitre afin d'installer et configurer notre environnement ELK.