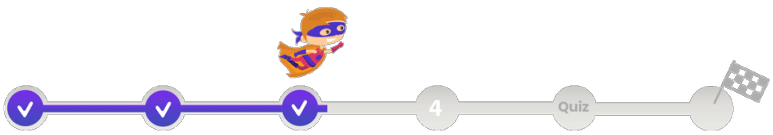
4 - Smart Logging



4. Smart Logging

La collecte de log applicatif est un élément essentiel de toute approche DevOps.
Il existe une **différence fondamentale** entre la **protection** et la **détection**.

Cependant, pour réaliser votre voyage vers le DevOps en toute sécurité, vous aurez besoin d'aborder les deux aspects.

La collecte systématique d'information joue un rôle important dans la détection d'erreurs, du moins, elle fournit une piste d'indices qui même après un incident va vous aider à établir la cause du problème.

Dans le meilleur des cas, pour les **outils proactifs**, cette **collecte** peut vous aider à **prévenir les incidents avant qu'il n'apparaissent**.

Il faut cependant se méfier de l'exploitation massive de données, car en cas d'incident, trouver la cause du problème peut s'avérer aussi périlleux que de rechercher une aiguille dans une botte de foin. Il faut donc **bien choisir les informations** que vous allez remonter dans votre outil de gestion des logs afin d'obtenir le **meilleur rendement pour la détection d'erreurs**.

Peu importe l'outil que vous choisirez dans la collecte et l'analyse de vos logs, il y a **trois notions principales** que vous devriez enregistrer :

- Les **erreurs** qui se produisent sur votre système. Vous aurez ainsi la possibilité de revenir en arrière et de corriger les erreurs dans votre application.
- Les **actions sensibles,** comme l'ouverture de session, l'exportation de fichiers, la suppression d'enregistrements de base de données. Toutes les actions d'authentification, d'exportation et d'intégrité des données qu'il faut monitorer au fur et à mesure pour identifier des comportements anormaux.
- Les **attaques possibles**, vous pouvez vous référer au guide de l'OWASP pour connaitre les dix attaques les plus utilisées sur les applications web. Il faut noter tout ce qui semble suspect.

4.1 Elastic Stack

Ce projet est anciennement connu sous le nom d'**ELK**, un acronyme désignant trois projets Open Source :

- **Elasticsearch**, un moteur de recherche et d'analyse de données.
- **Logstash,** un pipeline de traitement de données côté serveur qui ingère des données provenant de plusieurs sources simultanément, les transforme, puis les envoie à une "stash" comme *Elasticsearch*.
- **Kibana**, qui permet aux utilisateurs de visualiser les données d'*Elasticsearch* à l'aide de tableaux et de graphiques.

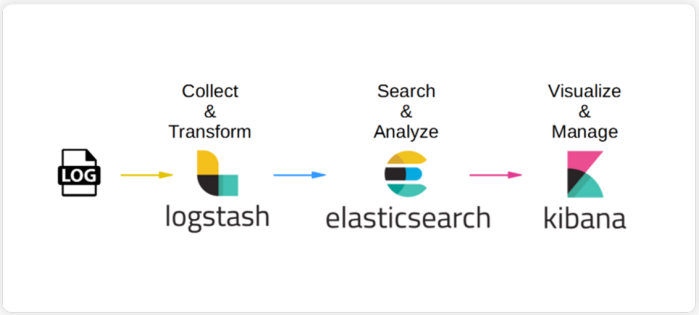


Figure 1 : Fonctionnement d'Elastic Stack

Ces outils **libres** sont développés par la société Elastic, qui encadre le développement de la communauté tout en proposant des services complémentaires tels que le support, la formation, l'intégration et l'hébergement dans un cloud avec une version SaaS.

Il existe en réalité 4 outils développés par la société Elastic. Le plus récent porte le nom de **"Beats"** et s'occupe principalement de la collecte de métriques systèmes, de réseaux et même de statistiques temps réel sur des fichiers. C'est l'arrivée de ce dernier outil qui a entrainé un renommage et c'est pour cela que la stack d'application initialement connu sous « ELK » porte désormais le nom « d'Elastic Stack ».

Figure 2 : Exemple de Dashboard avec Kibana

Dans sa version initiale, cette solution permet de répondre facilement et gratuitement à une problématique de collecte de **log centralisé**. Elle ne vous aidera cependant pas à implémenter des mécanismes de recherches proactives vous permettant de **prévoir des comportements anormaux**.
Pour cela, il vaut mieux se pencher sur un outil payant tel que **« Splunk »**.

4.2 Splunk

Splunk est **l'outil de gestion des logs le plus évolué du marché**.

Sa mission principale est de rendre les données générées par des machines compréhensibles de tous dans une même entreprise, en identifiant des tendances de données, en donnant des outils de mesure, en diagnostiquant les éventuels problèmes et en fournissant des informations relatives à l'activité commerciale de l'entreprise.

Figure 3 : Logo de Splunk

Splunk est en réalité le nom de l'entreprise qui développe et vend une suite de logiciels permettant le :

- Reporting pour les métiers,
- Suivi de la performance et du respect des SLA,
- Évaluation de la qualité d'une release ou d'un code,
- Surveillance opérationnelle en 24/7 des infrastructures et de l'utilisation des ressources,
- Monitoring de la performance et de la montée en charge des applications,
- Monitoring de tests et de déploiements,
- Monitoring des applications mobiles,
- Analyses et monitoring des IoT et des distributeurs automatiques,
- Supervision des systèmes industriels,
- Analyses de tendance & planification des capacités,
- Surveillance des matériels/OS/processus,
- Monitoring de la sécurité,
- Investigation d'incidents,
- Journaux d'Audit et respect des règles de conformité.

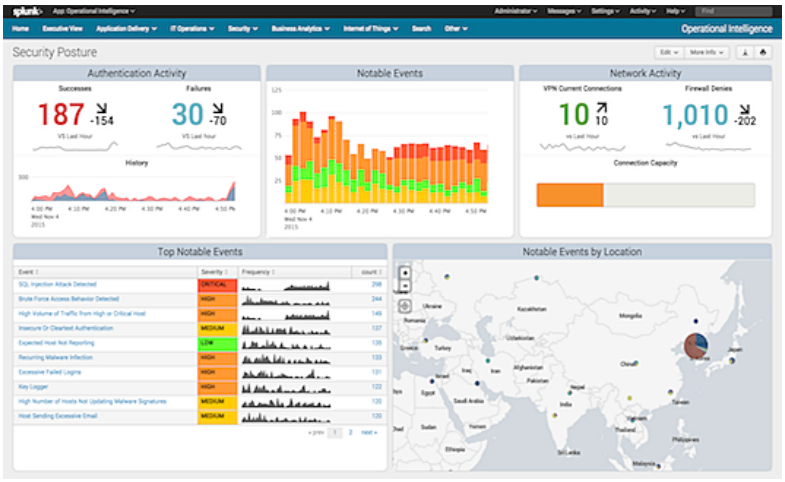


Figure 3 : Exemple de dashboard avec Splunk

Vous l'aurez compris, leurs outils ne se limitent pas à de la **collecte de log** : ils permettent d'effectuer une **analyse proactive** de **menace informatiqu**e, de **détection** et de **prévention de panne**.

La société utilise principalement des technologies de **Big Data** et de **Machine Learning** pour arriver à traiter d'immenses quantités d'informations.

4.3 Le Big Data

Également appelé **méga-données** ou **données massives**, il désigne des ensembles de données devenues si volumineuses qu'elles dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques de gestion de base de données.

4.4 Le Machine Learning

L'**apprentissage automatique**, ou **apprentissage statistique**, est un champ d'étude de l'intelligence artificielle, qui

concerne la conception, l'analyse, le développement et l'implémentation de méthodes permettant à une machine d'évoluer par un processus systématique et ainsi de remplir des tâches difficiles ou problématiques par des moyens algorithmiques plus classiques.