Project : Manipulating ELK Stack

Contexte orienté fonctionnel :

ELK veut faciliter et accélérer la recherche et l'analyse de grands ensembles de données.

Elasticsearch va permettre d'extraire les données, **Logstash** normalise toutes sortes de données temporelles et **Kibana** apporte un insight.

Bien qu'**Elasticsearch**, **Logstash** et **Kibana** aient été conçus pour fonctionner ensemble, chacun d'entre eux est un outil distinct.

ElasticSearch est un moteur de recherche et d'analyse qui utilise le format JSON. Son objectif est d'extraire efficacement les données à partir de sources de données structurées ou non structurées en temps réel. **Elasticsearch** utilise **Lucene** pour fournir les capacités de recherche en texte intégral les plus puissantes disponibles dans n'importe quel produit open-source.

Logstash est un outil pour la saisie, le traitement et la sortie des données logs. Sa fonction est d'analyser, filtrer et découper les logs pour les transformer en documents formatés à destination d'**Elasticsearch**.

Kibana est un tableau de bord interactif et paramétrable qui permet de visualiser les données stockées dans **ElasticSearch**. **Kibana** apporte un insight sur les tendances et les modèles sous toutes formes de diagrammes et courbes. Ce **dashboard** peut être partagé et associé à des visualisations de données pour une communication rapide et intelligente.

Beats :
Filebeat, Metricbeat, Packetbeat, Winlogbeat, Auditbeat, Heatbeat, Functionbeat, Packetbeat
Beats est la plateforme qui regroupe des agents légers réservés au transfert de données. Que vous exploitiez des centaines ou des milliers de machines, les agents Beats se chargent de transférer vos données vers Logstash et Elasticsearch.

Beats Communautaire :
dockerbeat, execbeat, flowbeat, mysqlbeat,

- https://www.elastic.co/
  (Elasticsearch, Kibana, Logstash, Beats, PacketBeat)

Contexte orienté infrastructure

Virtualisation :
La virtualisation consiste, en informatique, à exécuter sur une machine hôte, dans un environnement isolé, des systèmes d'exploitation — on parle alors de virtualisation système — ou des applications — on parle alors de virtualisation applicative (conteneur).

- https://fr.wikipedia.org/wiki/Virtualisation

Conteneur :
Un conteneur est une enveloppe virtuelle qui permet de distribuer une application avec tous les éléments dont elle a besoin pour fonctionner : fichiers source, environnement d'exécution, librairies, outils et fichiers. Ils sont assemblés en un ensemble cohérent et prêt à être déployé sur un serveur et son système d'exploitation (OS). Contrairement à la virtualisation de serveurs et à une machine virtuelle, le conteneur n'intègre pas de noyau, il s'appuie directement sur le noyau de l'ordinateur sur lequel il est déployé.

- https://fr.wikipedia.org/wiki/Conteneur_(informatique)

Virtualbox :
est un logiciel libre de virtualisation publié par Oracle.

- https://www.virtualbox.org/
- http://www.if-not-true-then-false.com/2010/install-virtualbox-with-yum-on-fedora-centos-red-hat-rhel/
- http://www.itzgeek.com/how-tos/mini-howtos/how-to-install-virtualbox-extension.html

Vagrant :
Vagrant est un logiciel libre et open-source pour la création et la configuration des environnements de développement virtuel. Il peut être considéré comme un wrapper autour de logiciels de virtualisation comme VirtualBox.

- https://fr.wikipedia.org/wiki/Vagrant
- https://computingforgeeks.com/how-to-install-vagrant-and-virtualbox-on-fedora/

L'objectif étant de construire la stack ELK et la manipuler

- Construction de l'environnement (machine virtuelle + stack)
- Réaliser une collecte de données (beats/packetbeat)
  - Source des PCAP : https://www.netresec.com/?page=pcapfiles
    #SCADA/ICS Network Captures
- Réaliser une indexation des données (packetbeat+elasticsearch)
- Construire un dashboard et présenter les données (dashboard kibana)
  - les données et features à afficher dans le dashboard seront à votre appréciation. L'objectif étant de déployer plusieurs widgtets utiles à l'analyse des données.

Pour réaliser ce projet, vous pouvez

- Utiliser une machine virtuelle (virtualbox, vagrant, ...) avec un OS type linux (fedora, centos, ubuntu, debian, ...) pour installer la Stack
- ou Utiliser un conteneur docker contenant déja toute la stack
- Installer la dernière version de la stack ELK + PacketBeat (tout sur la même machine virtuelle)
- Utiliser les agents beats d'elasticsearch pour parser et indexer le flux réseau disponible sous forme de PCAP (PacketBeat)
  - Choisir un PCAP à faire analyser par PacketBeat pour indexation dans Elasticsearch
- Créer un Dashboard pour présenter les données du PCAP Indexé
- Exporter votre Dashboard

Les livrables du dossier seront

- Un document PDF contenant :
  - Définissez et décrivez l'environnement que vous avez construit
    (technologie de virtualisation, machine hôte, machine invitée, ...)
  - Définissez et décrivez le processus d'installation de la stack et de votre environnement de travail *(commandes bash utilisées pour tout réaliser)*
  - Définissez et décrivez le processus de parsing et d'indexation du flux réseau (le fichier PCAP utilisé par PacketBeat)
    (commandes bash utilisées pour tout réaliser)
  - Insérez des schémas de type :
    - Workflow des traitements
    - Architecture de votre environnement
  - Précisez sur quel PCAP vous avez travaillés
  - Présentez les résultats de votre Dashboard Kibana et explicitez les données et widgets
- Votre dashboard kibana
  (vous devriez pouvoir exporter votre dashboard en javascript (.js) pour que je puisses le recharger dans mon kibana)

Consignes

- Le travail est individuel
- Format du livrable : PDF,
  « Je ne lirais pas les documents words, odt, powerpoint, uniquement la version PDF de votre document »
- Le choix de la langue de votre rapport est à votre appréciation (Français ou Anglais)
- ***M'envoyer par mail votre étude avant le 16/02 :*** julien.dreano@gmail.com.

Références

- https://www.elastic.co/
  (Elasticsearch, Kibana, Logstash, Beats, PacketBeat)
- https://www.elastic.co/fr/beats/packetbeat
- https://www.netresec.com/?page=pcapfiles
  #SCADA/ICS Network Captures
- https://fr.wikipedia.org/wiki/Conteneur_(informatique)
- https://fr.wikipedia.org/wiki/Virtualisation
- https://fr.wikipedia.org/wiki/Pcap
- https://www.virtualbox.org/
- http://www.if-not-true-then-false.com/2010/install-virtualbox-with-yum-on-fedora-centos-red-hat-rhel/
- http://www.itzgeek.com/how-tos/mini-howtos/how-to-install-virtualbox-extension.html
- https://fr.wikipedia.org/wiki/Vagrant
- https://computingforgeeks.com/how-to-install-vagrant-and-virtualbox-on-fedora/

PCAP	pcap (« packet capture ») est une interface de programmation permettant de capturer un trafic réseau. Elle est implémentée sous les systèmes GNU/Linux, FreeBSD, NetBSD, OpenBSD et Mac OS X par la bibliothèque libpcap. WinPcap est le portage sous Windows de libpcap. • https://fr.wikipedia.org/wiki/Pcap
BASH	Bourne-Again shell est un interpréteur en ligne de commande de type script. C'est le shell Unix du projet GNU.