SIEM OpenSearch

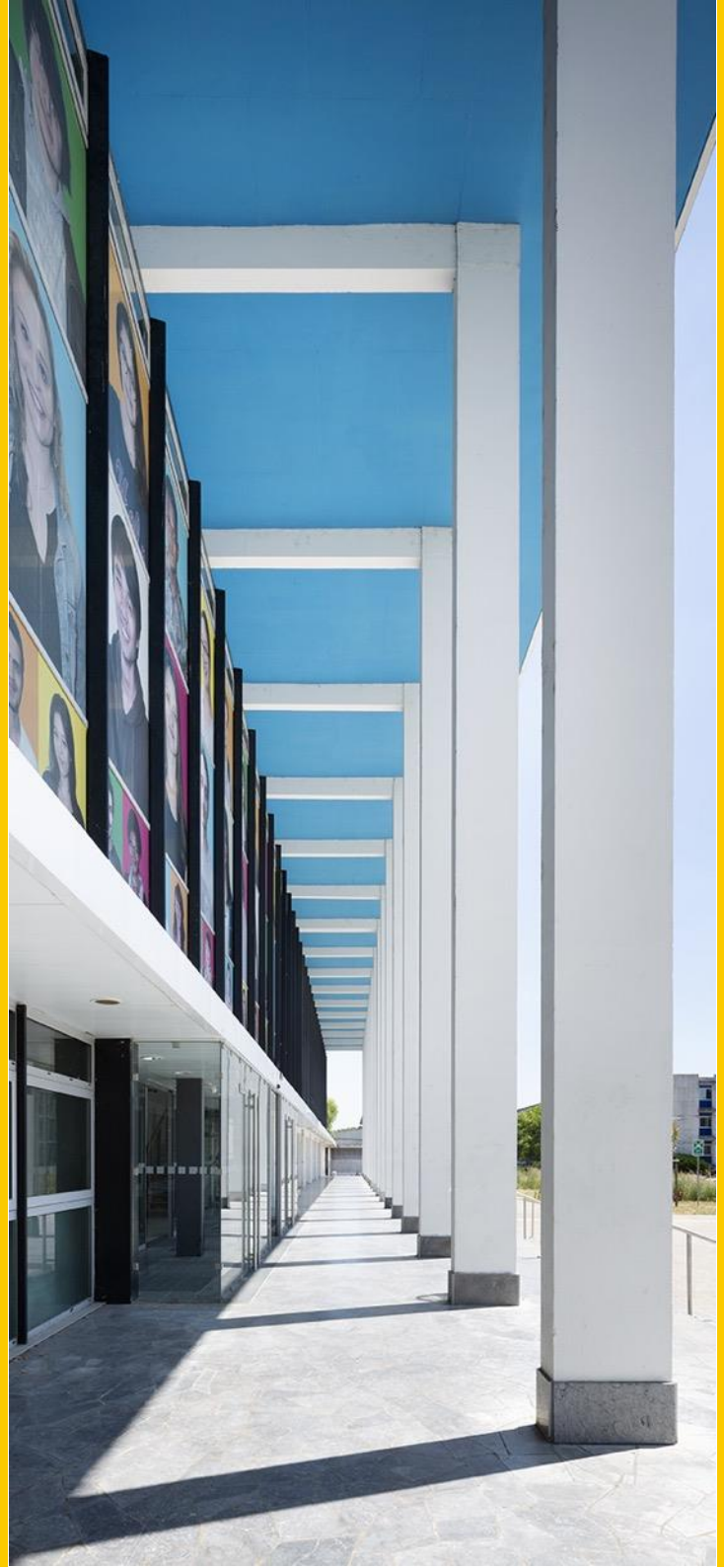
DSI
Université Toulouse 3
Paul Sabatier

-

CAPITOUL

-

16 février 2023
Antoine Madeline



Sommaire

- Contexte et contraintes
- Un SIEM ?
- Registre
- OpenSearch
- Sources de logs
- Architecture
- Stockage
- Docker
- Collecte avec Filebeat
- Traitement et enrichement avec Logstash
- Elastalert
- Monitoring
- Difficultés
- Questions

Contexte et contraintes

- Au sein de la DSI :
  - Utilisé durant les permanences
  - En lien avec la supervision check_mk
- Remplacement de la solution anti-spam
- Utilisation de docker
- Le siem ne remplace pas les serveurs syslog existants

Un SIEM ?

- Security Information And Event Management
- Permet de collecter, traiter, enrichir, stocker, filtrer, présenter, rechercher, alerter parmi les logs issus de multiples serveurs.
- Blocage de comptes messagerie compromis
- Analyse de compte de messagerie douteux
- Détection de session windows à privilèges
- Analyse des sessions à problèmes lors d’examens en ligne

Registre 1/3

Activité “Corrélation de journaux informatiques” déclarée au registre des activités de traitement de données de l’Université

Finalités positives :

Effectuer une analyse technique d’évènements :

- pour permettre de détecter et/ou aider à l’analyse d’un dysfonctionnement ou une compromission,
- à la demande d’un utilisateur et uniquement sur les données qui le concernent.

Registre 2/3

Finalités négatives :

Cette activité ne peut pas être réalisée pour :

- effectuer un suivi des activités des personnels, des usagers (ex : messagerie, connexions, consultation web, violation du secret de la correspondance) en dehors des finalités positives,
- exporter des données à des personnes qui n'ont pas fait l'objet d'une autorisation à traiter les données, en dehors des finalités positives décrites.

Registre 3/3

- Le co-dir décide des habilitations.

- Information de la dsi :

Une information sur le traitement des données dans le cadre de cette activité doit être faite auprès :

- des personnes autorisées à traiter les données,
- des personnes concernées par le traitement de leurs données.

OpenSearch

- Créé suite changement licence elastic
- Septembre 2019 opendistro
- Juillet 2021 OpenSearch 1.0.0
- Fonctionnement comparable à ELK
  - Noeuds: OpenSearch
  - Affichage: OpenSearch-dashboards
  - Collecte: logstash-oss-with-opensearch-output-plugin
- Authentification et autorisation via annuaire ldap/ad

Architecture physique

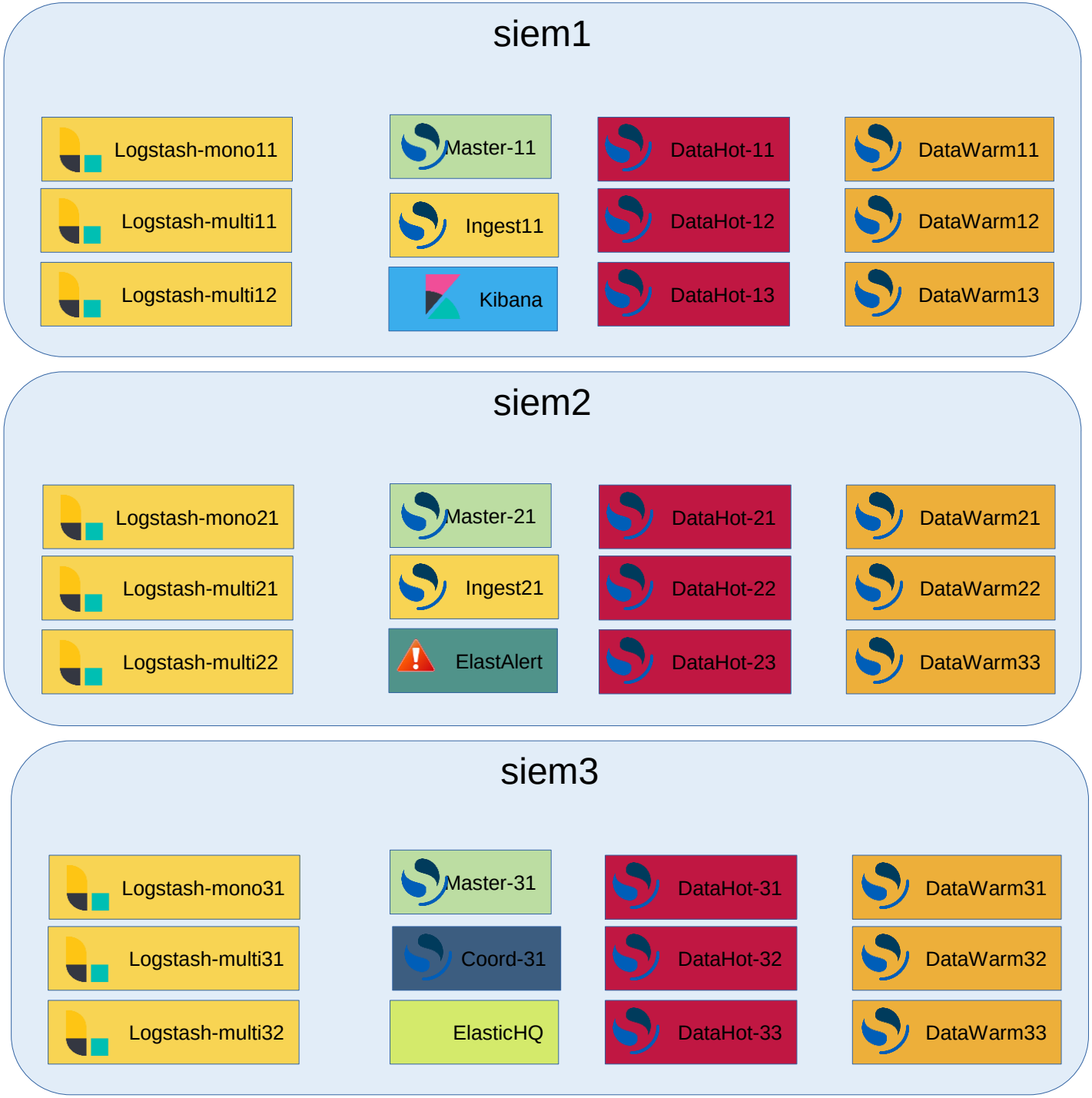
- 3 serveurs recyclés
  - Type serveur : PowerEdge R730xd
  - RAM : 384 Go
  - Processeurs : 48 cœurs
  - Disques
    - 9 disques ssd : 3 volumes raid 0 d’environ 1 To
    - 4 disques hdd : 1 volume raid 1 200 Go, 1 volume raid 0 200 Go
  - Réseau : bonding 2*10 GB

Architecture logique
1 noeud OpenSearch c’est 1 conteneur

- 9 nœuds data hot (sur ssh locaux)
- 9 nœuds data warm (sur nfs distant)
- 3 nœuds master – cluster manager
- 2 nœuds ingest
- 1 nœud coordinateur

- 6 logstash multithread
- 3 logstash monothread
- elastalert
- geoipupdate

Architecture schéma



Architecture : Calcul nbre de nœuds HOT

https://www.elastic.co/fr/blog/benchmarking-and-sizing-your-elasticsearch-cluster-for-logs-and-metrics

Volume journalier de logs palo 120 Gb

Total Data : logs par jours * Nombre jours * (Nbre replicat +1)

conservation 15 jours

1 replicat

TD = 150 * 15 * 2 = 4500 Gb

Total Storage :Total data (GB) * (1 + 0.15 disk Watermark threshold + 0.1 Margin of error)

TS = 4500 * 1.25 = 5 625 Gb

Total Data Nodes = ROUNDUP(Total storage (GB) / Memory per data node / Memory:Data ratio)

Memory:Data ratio : 30 pour hot

Memory per data node 24 Gb

TDN = 5625 /24/30 = 8 nœuds hot

ajouter un noeud pour redondance → 9 nœuds HOT

- "OPENSEARCH_JAVA_OPTS=-Xms25g -Xmx25g"

Stockage : GET cat/shards/palo-000002

indice	shard	primaire/replicas	Nbre documents	taille	nœud
palo-000002	0	p	49277993	12.8gb	hot-23
palo-000002	0	r	49229204	12.8gb	hot-21
palo-000002	1	p	49241830	13.8gb	hot-33
palo-000002	1	r	49290153	13.5gb	hot-22
palo-000002	2	p	49266475	16.8gb	hot-11
palo-000002	2	r	49216259	14.1gb	hot-13
palo-000002	3	p	49267912	12.7gb	hot-32
palo-000002	3	r	49126016	13.1gb	hot-31

Stockage

- Définition de la température des données par stratégies ISM (Index State Management)
  - Hot 5 jours
  - Warm 30 jours
  - Ensuite suppression

- Un seul réplica configuré : permet la redondance et accèlere les requetes
- 4 shards (segmentation sur plusieurs nœuds)
- Roolover 60 Gb : taille max d’un indice

Docker

- Contrainte du projet : montée en compétences docker
- Docker compose sur chacuns des serveurs

Docker : Extrait docker-compose hot-11

```
hot-11:
  container_name: hot-11
  image: opensearchproject/opensearch:2.5.0
  restart: unless-stopped
  depends_on:
    master-11:
      condition: service_healthy
  environment:
    - cluster.name=opensearch-cluster
    - node.name=hot-11
    - network.publish_host=IP
    - network.host=0.0.0.0
    - transport.port=9311
    - http.port=9211
    - node.roles=data
    - node.attr.temp=hot # - node.attr.box_type=hot
    - discovery.seed_hosts=master1:9300,master2:9300,master3:9300
    - cluster.initial_master_nodes=master-11, master-21, master-31
    - bootstrap.memory_lock=true # along with the memlock settings below, disables swapping
    - "OPENSEARCH_JAVA_OPTS=-Xms25g -Xmx25g" # minimum and maximum Java heap size, recommend setting both to 50% of system RAM
  ulimits:
    memlock:
      soft: -1
      hard: -1
    nofile:
      soft: 65536 # maximum number of open files for the Elasticsearch user, set to at least 65536 on modern systems
      hard: 65536
  volumes:
    - /data_ssd1:/usr/share/opensearch/data
    - ./../certificates/opensearch/hot-11.pem:/usr/share/opensearch/config/node.pem
    - ./../certificates/opensearch/hot-11-key.pem:/usr/share/opensearch/config/node-key.pem
    - ./../certificates/opensearch/root-ca.pem:/usr/share/opensearch/config/root-ca.pem
    - ./../certificates/opensearch/admin.pem:/usr/share/opensearch/config/admin.pem
    - ./../certificates/opensearch/admin-key.pem:/usr/share/opensearch/config/admin-key.pem
    - ./opensearch.yml:/usr/share/opensearch/config/opensearch.yml
  ports:
    - "9211:9211"
    - "9311:9311"
  networks:
    - opensearch-net
```

Sources de logs

Eco-système messagerie :

- Zimbra, passerelles postfix, anti-spam rspamd

Eco-système authentification :

- cas
- active directory

Autres :

- Moodle
- Pare-feux

Collecte avec Filebeat

- Filebeat installé sur serveurs producteur de logs et sur serveurs syslog
- Filebeat sous docker pour équipements émetteur au format syslog
- Nxlog pour windows

- Toutes les soureces envoient vers logstash

- Redondance / Equilibrage vers plusieurs logstash

Collecte : Exemple

```
#========================== Filebeat inputs =============================
filebeat.inputs:
- type: log
  enabled: true
  paths:
    - /data/log/syslog.firewall1
    - /data/log/syslog.firewall2
  exclude_files: ['.gz2$']
  fields:
    log_filter: palo
    service_type: prod #debug
- type: log
  # Change to true to enable this input configuration.
  enabled: true
    - /data/log/test_redo/*.log
  exclude_files: ['.gz2$']
  fields:
    log_filter: palo
    service_type: debug
  # Defines the buffer size every harvester uses when fetching the file
  # harvester_buffer_size: 16384
  harvester_buffer_size: 4096000 #16384 * 250

#----------------------------- Logstash output --------------------------------
output.logstash:
#   # The Logstash hosts
  hosts: ["logstash1","logstash2","logstash3.fr","logstash4","logstash5","logstash6"]
  loadbalance: true
  bulk_max_size: 10240 # 1024*5
  # Number of workers per Logstash host.
  worker: 5

# ================================== General ===================================
queue:
  # Queue type by name (default 'mem')
  mem:
    # Max number of events the queue can buffer.
    #events: 4096
    events: 96000
```

Traitement avec Logstash

- Limiter le nombre de champs
- Uniformiser les champs entre les diffrentes sources
- Elastic Common Schema

- Utiliser le filtre adapté

Logstash : Filtre dissect
Texte formaté avec séparateurs - mailbox

Log :

2021-09-01 13:47:44,523 INFO [Pop3Server-2286] [name=jacques.toot@univ-tlse3.fr;ip=aaa.bbb.ccc.ddd;oip=AAA.BBB.CCC.DDD ;cid=2913737;] pop - QUIT elapsed=0

Filtre :

%{date} %{time} %{[log][level]} %{}[%{part1}] [%{part2}] %{module} - %{module_args}

Champs :

"date": "2021-09-01",

"time": "13:47:44,523",

"[log][level]": "INFO",

"part1": "Pop3Server-2286",

"part2": "name=jacques.toot@univ-tlse3.fr;ip=aaa.bbb.ccc.ddd;oip=AAA.BBB.CCC.DDD ;cid=2913737;",

"module": "pop",

"module_args": "QUIT elapsed=0"

Logstash : Filtre grok cas apereo Expressions régulieres et templates

```
grok
{
  match =>
    { "message" =>
      [
        "%{APEREO_HEADER}Audit trail record BEGIN%{DATA}\nWHO: %{DATA:[cas][who]}\nWHAT:
%{DATA:[cas][what][raw]}\nACTION: %{DATA:[event][action]}\nAPPLICATION: %{DATA:[cas][application]}?\nWHEN:
%{APEREO_DATE}\nCLIENT IP ADDRESS: %{IP:[source][ip]}\nSERVER IP ADDRESS:
%{IP:[cas][serveripaddress]}\n%{GREEDYDATA:fin}$",
        "%{APEREO_HEADER}%{GREEDYDATA:[cas][data]}>$"
      ]
    }
    pattern_definitions =>
    {
      "TOMCAT8_DATE" => "^20%{YEAR}-%{MONTHNUM}-%{MONTHDAY} %{HOUR}:?%{MINUTE}(?::?%{SECOND})"
      "APEREO_HEADER" => "%{TOMCAT8_DATE:[timestampstr]} %{LOGLEVEL:[log][level]}
\[%{NOTSPACE:[cas][category]}\] - <"
      "APEREO_DATE" => "%{DAY} %{MONTH} %{MONTHDAY} %{TIME} %{WORD:tz} %{YEAR}"
    }
}
```

Logstash : Filtre aggregate - postfix

- Permet d'agréger des champs de plusieurs lignes logs différentes dans un seul document
- Il faut un id unique postfix_queueid
- Logstash filter workers to 1

Feb 7 15:56:06 gw1-out postfix/**qmgr**[1011]: **5F5FC418FC**: **from=<xxx@univ-tlse3.fr>**, size=23937, nrcpt=1 (queue active)

Feb 7 15:56:07 gw1-out postfix/**smtp**[1784071]: **5F5FC418FC**: to=<toto@univ-tlse3.fr>, relay=prod-zextras-mta-in.univ-tlse3.fr[195.220.57.8]:25, delay=0.08, delays=0.06/0/0.01/0.01, dsn=2.0.0, status=sent (250 2.0.0 Ok: queued as 71A4E185977)

Traitement première ligne **qmgr** : **mise en mémoire du champ postfix_from et suppression**

```
task_id => "%{postfix_queueid}"
map['postfix_from'] = event.get('postfix_from')
    drop
```

Traitement seconde ligne **smtp** : **ajout de tous les champs mémorisés**

```
 task_id => "%{postfix_queueid}"
map.each do |key, value|
event.set(key, value)
```

Logstash : Filtre transalte

Permet d'enrichir avec un dictionnaire yaml : attributs ldap à partir d'un email

Yaml :

```
antoine.madeline@univ-tlse3.fr:
  eduPersonAffiliation: [employee, member]
  eduPersonPrimaryAffiliation: [employee]
  email: [antoine.madeline@univ-tlse3.fr]
  givenName: [antoine]
  sn: [madeline]
  uid: [xxxxx]
```

Translate :

```
field  => "[user][email]"
destination  => "[user]"
override => true
dictionary_path => "/usr/share/logstash/dictionaries/ldap-email.yaml"
```

Objet :

```
"user": {
        "sn": [ "madeline" ],
        "email": [ "antoine.madeline@univ-tlse3.fr" ],
        "eduPersonAffiliation": [ "employee", "member" ],
        "givenName": [ "antoine" ],
        "eduPersonPrimaryAffiliation": [ "employee" ],
        "uid": [ "xxxxx" ]
      },
```

Logstash : geoip

- Permet d'enrichir les IP publiques d'information géographiques (ville, pays, lon, lat, ...)
- Mise à jour 2 fois par semaine chez maxminddb, par conteneur geoipupdate
- Enrichissement de la base mmdb avec infos IP ut3, contenues dans 2 yaml.
- https://github.com/maxmind/mmdb-from-go-blogpost Enriching MMDB files with your own data using Go

- **Par subnets :**

192.bbb.ccc.0/zz:
- utilisation_ut3: 4R3 Rech
- vlan_ut3: '12'
- zone_ut3: RECH

Par ip :

192.bbb.ccc.ddd:
- dns_name_ut3: xxxx

Le champ `region_name` est récupéré classiquement par le filtre geoip, sous forme de json, puis décomposé par ruby

```
{"region_name":"", "utilisation_ut3":"4R3
Rech","vlan_ut3":« 12","zone_ut3":"RECH","dns_name_ut3":"xxx"}
```

```
a =  event.get("[destination][geo][region_name]")
a_json = JSON.parse(a)
a_json.each { |k, v| event.set("[destination][geo][#{k.to_s}]", v.to_s) }
```

Elastalert : Principe

Projet indépendant d'Elastic et d'OpenSearch

Alertes envoyées par exécution de règles

Ajout de fonctionnalités en python, sur les régles et les alertes

Régles de type :

Any, frequency, flat, new, spike, ...

Alertes :

Mails

Event syslog (vers la supervision)

Script de blocage

Autre règle

...

Elastalert : Messagerie

Détection de compte compromis avec prise en compte de multiples critères :

- Localisation géographique
- Envoi massif
- Comportement suspect : modification préférence dans le webmail

Alerte et blocage de compte

Elastalert : cas3_rspamd-utilisateur_composante_douteux 1/2

```
# X Dépassements du burst depuis une composante en Y heures
type: frequency
index: rspamd*

filter:
- exists:
    field: "bucket_burst_value"
- exists:
    field: "ut3_mta"
- query:
   query_string:
    query: "NOT ut3_mta: WEBMAILDSI AND NOT ut3_mta: SMTPSDSI"

num_events: X
timeframe:
  hours: Y

# The maximum number of documents that will be downloaded from Elasticsearch in a single query.
max_query_size: 1
```

Elastalert : cas3_rspamd-utilisateur_composante_douteux 2/2

```
realert:
    minutes: 60
query_key: user.email

alert:
- "email"
email: "xxxx@univ-tlse3.fr"
alert_subject:  "Compte utilisateur composante douteux : {{user.email}}"
alert_text_type: alert_text_jinja
alert_text: |
  Bonjour,<br><br>
  le compte {{user.email}} est douteux <br>
  X dépassements de seuil en moins de Y heures <br>
  Les destinataires sont du type : {{rcpts}} <br>
  Le sujet du message est : {{msg_subject}}<br>
  Caractéristiques :
    - Adresse IP : {{source.ip}} <br>
    - Pays : {{source.geo.country_name}} <br>
    - Ville : {{source.geo.city_name}} <br>
```

Monitoring

- Dans le dispositif de supervision existant :
  - Etat du cluster
  - Santé des conteneurs
  - Surveillance des index
  - Surveillance des logstash
  - Metrics : filebeat, logstash

Difficultés

- Trier les informations :
  - Explosion du nbre de champs
  - Choisir de ne pas traiter certaines lignes
- Interpréter les logs des applications
- Performances d’ingestion
  - Plus de 10 000 événements seconde
- Limiter le nombre d’index : Réunir des données issues d’appli différentes

Questions ?