



SOC LEVEL 1 ESSENTIALS

– PART B: SIEM & LOG ANALYSIS (POST 11-20)

Mastering SIEM Fundamentals for Modern SOC Analysts



JUNE 25, 2025

INAAM KABBARA

Security Analyst | Cybersecurity Content Creator | EPITA MSc Graduate

Table of Contents

1- What Is a SIEM and Why Do SOCs Use It?.....	2
2- SIEM Examples: What Tools Do SOCs Actually Use?	4
3- What Are Log Sources in a SOC?	5
4- How to Read an Authentication Log	7
5- Event vs Alert vs Incident – What's the Difference?.....	9
6- Reducing False Positives in a SIEM.....	11
7- Correlating Logs – Why and How	13
8- Log Enrichment Using Threat Intelligence.....	15
9- SIEM Use Case: Detecting Anomalous Logins.....	17
10- SIEM Query Examples: KQL, Lucene, DSL.....	19

1- What Is a SIEM and Why Do SOCs Use It?

A SIEM (Security Information and Event Management) is a critical tool used in Security Operations Centers to collect, analyze, and correlate log data from across an organization's infrastructure.

🧠 Why it matters in a SOC:

SIEMs are the backbone of modern SOCs. They centralize visibility, detect potential threats, generate alerts, and support incident investigation and response — all from a single platform.

💡 Key SIEM capabilities include:

- Centralized log aggregation
- Real-time correlation & alerting
- Historical research & forensics
- Dashboards for threat visibility
- Compliance and audit reporting

⚙️ Popular SIEMs: Splunk, ELK Stack, Microsoft Sentinel, IBM QRadar, ArcSight

🔒 Pro Tip: Mastering your SIEM's detection rules and understanding its log ingestion process will make you far more effective as a SOC analyst.

❓ Which SIEM have you worked with, or are planning to learn first?

Qu'est-ce qu'un SIEM et pourquoi est-il utilisé dans un SOC ?

Un SIEM (Security Information and Event Management) est un outil indispensable utilisé dans les centres d'opérations de sécurité pour collecter, analyser et corrélérer les données de journaux provenant de l'ensemble de l'infrastructure informatique.

🧠 Pourquoi c'est essentiel dans un SOC :

Le SIEM est la colonne vertébrale du SOC moderne. Il centralise la visibilité, détecte les menaces potentielles, génère des alertes et facilite l'investigation et la réponse aux incidents — le tout à partir d'une seule plateforme.

💡 Fonctionnalités principales d'un SIEM :

- Agrégation centralisée des logs
- Corrélation en temps réel & alertes
- Recherche historique et forensic
- Dashboards pour la visibilité des menaces
- Rapports d'audit & conformité

⚙️ Exemples populaires : Splunk, ELK Stack, Microsoft Sentinel, IBM QRadar, ArcSight

 Conseil pro : Comprendre comment votre SIEM ingère les logs et applique les règles de détection vous rendra bien plus efficace en tant qu'analyste SOC.

 Quel SIEM avez-vous déjà utilisé ou souhaitez-vous apprendre ?

2- SIEM Examples: What Tools Do SOCs Actually Use?

In real SOC environments, you'll find several types of SIEM platforms — each with its own strengths, interfaces, and use cases. Here are five widely-used SIEM tools every analyst should know:

 Top SIEM Solutions Used in SOCs:

- 1 Splunk: Very powerful, scalable, supports advanced detection logic and dashboards
- 2 ELK Stack (Elasticsearch, Logstash, Kibana): Open-source, flexible, widely used in custom setups
- 3 Microsoft Sentinel: Cloud-native, integrates deeply with Azure services
- 4 IBM QRadar: Enterprise-grade, known for correlation and threat intelligence integration
- 5 ArcSight: Strong in compliance and large-scale event management

 Pro Tip: As a beginner, start with ELK or Splunk — both are great for learning log ingestion, parsing, detection, and dashboards.

 ? Which SIEM would you want to master first — and why?

Exemples de SIEM : Quels outils sont utilisés en SOC ?

Dans les environnements SOC réels, on retrouve plusieurs types de SIEM — chacun avec ses propres fonctionnalités, interfaces et cas d'usage. Voici cinq SIEM très utilisés que tout analyste devrait connaître :

 Principaux SIEM utilisés dans les SOC :

- 1- Splunk : Très puissant, évolutif, avec des règles avancées et de superbes dashboards
- 2- ELK Stack (Elasticsearch, Logstash, Kibana) : Open-source, flexible, idéal pour les architectures personnalisées
- 3- Microsoft Sentinel : Cloud-native, bien intégré à l'écosystème Azure
- 4- IBM QRadar : Solution entreprise, forte en corrélation et intégration avec la CTI
- 5- ArcSight : Réputé pour la gestion d'événements à grande échelle et la conformité

 Conseil pro: Pour débuter, commencez avec ELK ou Splunk — parfaits pour apprendre à gérer l'ingestion, la corrélation, et les dashboards.

 ? Quel SIEM souhaitez-vous maîtriser en premier — et pourquoi ?

3- What Are Log Sources in a SOC?

In a SOC, log sources are the various systems and tools that generate security-relevant data. These logs are ingested by the SIEM to help analysts detect suspicious or malicious activity.

🧠 Why it matters in a SOC:

Without diverse log sources, a SIEM can't see the full picture. The more complete your visibility, the more accurate your threat detection.

- 💡 Common SOC log sources include:
 - 🔒 Authentication systems: e.g. /var/log/auth.log, Active Directory
 - 🌐 Network devices: Firewalls, routers, switches (e.g. Cisco ASA, Fortinet)
 - 💻 Endpoints: Windows Event Logs, Sysmon, Linux syslog
 - 🧠 EDR/AV: Microsoft Defender for Endpoint, CrowdStrike, HarfangLab
 - ✳️ DNS and DHCP logs: Help trace lateral movement or C2 traffic
 - 💼 Applications: Web servers (Apache, Nginx), VPNs, mail servers
 - 📦 Security tools: Zeek, Suricata, IDS/IPS, SOAR platforms
 - ☁️ Cloud services: Azure AD, AWS CloudTrail, GCP audit logs

🔒 Pro Tip: Always map each log source to the MITRE ATT&CK framework to understand what attack techniques you're capable of detecting.

❓ Which log source do you think provides the richest data in your environment?

Qu'est-ce qu'une source de logs dans un SOC ?

Dans un SOC, les sources de logs sont les systèmes et outils qui génèrent des données liées à la sécurité. Ces journaux sont collectés par le SIEM pour aider les analystes à détecter les comportements suspects ou malveillants.

🧠 Pourquoi c'est important dans un SOC :

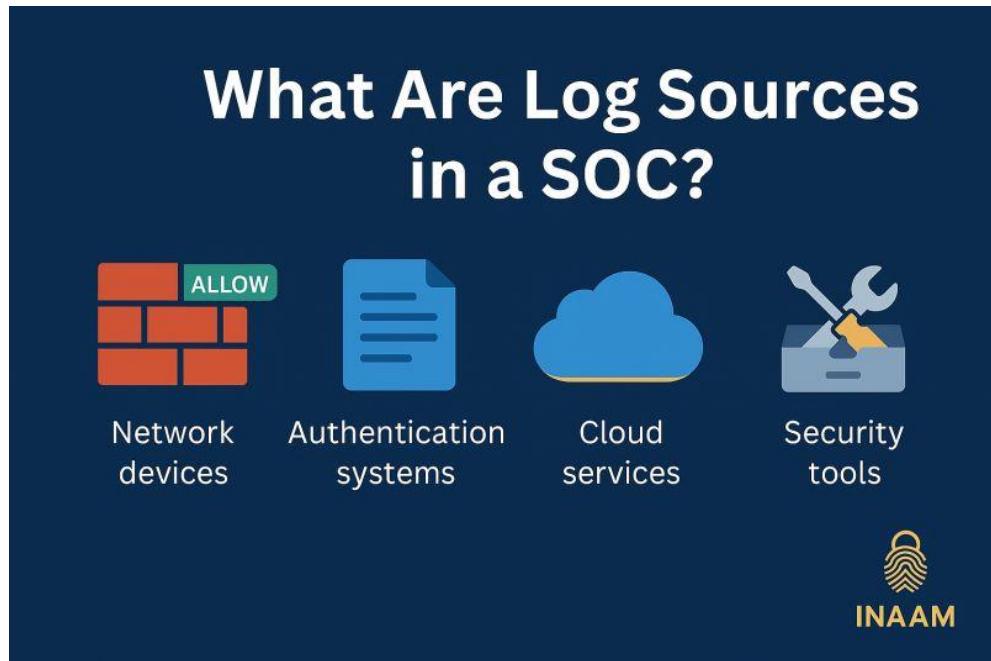
Sans diversité dans les sources de logs, le SIEM ne voit qu'une partie de l'activité. Plus votre visibilité est complète, plus vos détections sont fiables.

- 💡 Exemples courants de sources de logs :
 - 🔒 Systèmes d'authentification : /var/log/auth.log, Active Directory
 - 🌐 Équipements réseau : Pare-feux, routeurs, switches (Cisco ASA, Fortinet...)
 - 💻 Postes de travail et serveurs : Journaux Windows, Sysmon, syslog Linux
 - 🧠 EDR/antivirus : Microsoft Defender, CrowdStrike, HarfangLab
 - ✳️ Logs DNS/DHCP : Traçabilité des déplacements latéraux ou du trafic C2
 - 💼 Applications : Serveurs web (Apache, Nginx), VPN, messagerie
 - 📦 Outils de sécurité : Zeek, Suricata, IDS/IPS, plateformes SOAR

- ⌚ Services cloud : Azure AD, AWS CloudTrail, journaux GCP

⚠ Conseil pro: Cartographiez chaque source de logs au framework MITRE ATT&CK pour savoir quelles techniques vous pouvez détecter.

❓ Selon vous, quelle source de logs fournit les données les plus riches ?



4- How to Read an Authentication Log

Authentication logs track login attempts to systems and services. In a SOC, they're essential for spotting brute-force attacks, unauthorized access, and privilege escalation.

Why it matters in a SOC:

Authentication logs show who accessed what, when, and whether it was successful. Analysts use these logs to reconstruct the attack path and detect anomalies like impossible travel or account abuse.

How to read Linux auth logs (e.g. /var/log/auth.log):

A typical log line:

May 30 21:04:53 server sshd[6202]: Failed password for invalid user admin from 192.168.1.10
port 44522 ssh2

Key fields to interpret:

-  Timestamp: When the event occurred
-  Hostname: Which system it happened on
-  Service: sshd means SSH
-  Action: Success or failure
-  Username: Targeted account
-  Source IP: Where the request came from
-  Port/Protocol: Useful for correlating with firewall or IDS logs

 Pro Tip: Correlate failed login attempts across multiple hosts with the same source IP to spot distributed attacks.

 Have you ever traced an incident using auth logs? What did you discover?

Comment lire un journal d'authentification

Les journaux d'authentification enregistrent les tentatives de connexion aux systèmes et services. En SOC, ils sont cruciaux pour repérer les attaques par force brute, les accès non autorisés ou l'élévation de priviléges.

Pourquoi c'est important en SOC :

Ils permettent de savoir qui s'est connecté, quand, où et si la tentative a réussi. Les analystes s'en servent pour reconstituer le parcours d'une attaque ou détecter des anomalies (ex. : déplacement impossible, compte compromis).

Exemple d'analyse d'un log Linux (/var/log/auth.log) :

May 30 21:04:53 server sshd[6202]: Failed password for invalid user admin from 192.168.1.10
port 44522 ssh2

Champs clés à lire :

- ⌚ Horodatage : Date et heure de l'événement
- 💻 Hôte : Système concerné
- 🔒 Service : sshd signifie SSH
- ⚠ Action : Succès ou échec
- 👤 Nom d'utilisateur : Compte ciblé
- 🌐 Adresse IP source : Origine de la tentative
- 📦 Port/Protocole : À croiser avec les logs IDS ou pare-feu

🔒 Conseil pro: Corrélez les échecs de connexion sur plusieurs machines depuis une même IP pour détecter une attaque distribuée.

❓ Avez-vous déjà analysé un incident à partir d'un auth.log ? Qu'avez-vous trouvé ?



5- Event vs Alert vs Incident – What's the Difference?

In a SOC, these three terms often get confused — but each plays a different role in the detection and response pipeline.

🧠 Why it matters in a SOC:

Understanding the difference between an event, an alert, and an incident is critical for proper triage, escalation, and reporting.

💡 Simple breakdown:

✓ Event:

Any log or activity recorded by a device or system. It could be harmless — like a user login or system reboot. Most events are just raw data.

⚠ Alert:

A notification triggered when a rule or threshold is matched — e.g., 10 failed logins in 1 minute. Alerts are generated by SIEMs and require attention but aren't always malicious.

❗ Incident:

A confirmed security issue — validated by an analyst. It could be a successful brute-force attack, malware infection, or unauthorized access. Incidents are escalated, documented, and often followed by a response plan.

🔒 Pro Tip: Always remember: All incidents start as alerts, and all alerts originate from events. But not every event is bad, and not every alert becomes an incident.

❓ Have you ever had to escalate an alert into a confirmed incident? How did you confirm it?

Événement vs Alerte vs Incident – Quelle est la différence ?

Dans un SOC, ces trois notions sont souvent confondues — mais elles jouent chacune un rôle précis dans le cycle de détection et de réponse.

🧠 Pourquoi c'est important en SOC :

Bien faire la distinction permet de hiérarchiser les risques, d'escalader correctement et de mieux documenter les menaces.

💡 Explication simple:

✓ Événement :

Toute activité enregistrée — une connexion réussie, un redémarrage, un accès réseau. La

plupart sont bénins : ce sont des données brutes.

⚠️ Alerte :

Signal déclenché quand une règle ou condition est atteinte (ex. : 10 échecs de connexion en 1 minute). Les SIEM génèrent les alertes — elles attirent l'attention, mais ne sont pas toujours malveillantes.

❗ Incident :

Menace confirmée par un analyste — ex. : malware détecté, accès non autorisé, attaque réussie. Nécessite une réponse, une documentation, et souvent une analyse post-incident.

🔒 Conseil pro: Souvenez-vous : tout incident commence par une alerte, et toute alerte vient d'un événement. Mais tous les événements ne sont pas des menaces, et toutes les alertes ne deviennent pas des incidents.

❓ Avez-vous déjà transformé une alerte en incident ? Comment l'avez-vous confirmé ?



6- Reducing False Positives in a SIEM

False positives are one of the biggest challenges in a SOC. When your SIEM generates too many alerts that aren't actual threats, it leads to alert fatigue, slower response, and missed real attacks.

Why it matters in a SOC:

Reducing false positives improves detection accuracy, saves analysts' time, and makes the SOC more efficient.

How to reduce false positives in a SIEM:

-  Tune correlation rules: Adjust thresholds (e.g., raise brute-force attempts from 5 to 10)
-  Add context: Enrich alerts with threat intel, asset criticality, or known baselines
-  Whitelist known behavior: Exclude internal scanners, admin scripts, or test accounts
-  Use anomaly-based detection: Compare events to user or host behavior patterns
-  Test and simulate: Validate rules with red team scenarios or replayed attack data
-  Leverage SOAR: Automate triage and enrich alerts with contextual data before escalation

 Pro Tip: Always measure alert quality, not just volume. It's better to have 5 accurate alerts than 500 noisy ones.

 What's one technique you've used to reduce false positives in your environment?

Réduire les faux positifs dans un SIEM

Les faux positifs sont l'un des plus grands défis en SOC. Trop d'alertes non pertinentes entraînent de la fatigue, un triage lent, et parfois le manque d'identification des vraies attaques.

Pourquoi c'est important :

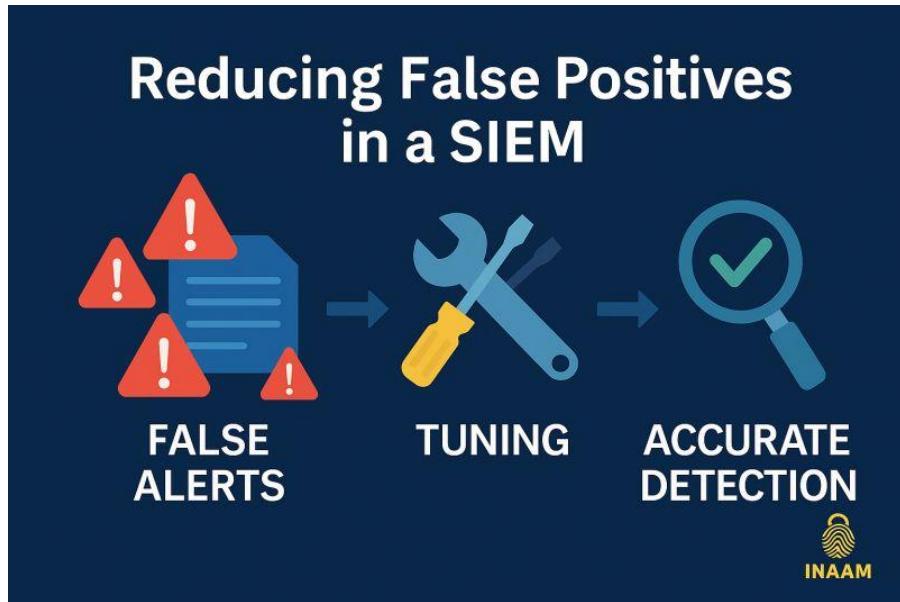
Réduire les faux positifs améliore la précision des détections, économise le temps des analystes et renforce l'efficacité du SOC.

Comment réduire les faux positifs dans un SIEM :

-  Affiner les règles de corrélation : Ajuster les seuils (ex. : 10 tentatives au lieu de 5)
-  Ajouter du contexte : Criticité des assets, renseignement sur la menace, comportement normal
-  Lister les comportements légitimes : Scripts internes, scanners autorisés, comptes de test
-  Détection basée sur l'anomalie : Comparer à des profils comportementaux
-  Tester avec des scénarios : Simulation d'attaques, données de red team
-  Utiliser un SOAR : Automatiser l'enrichissement des alertes avant leur traitement

 Conseil pro: Ne mesurez pas la quantité d'alertes, mais leur qualité. Mieux vaut 5 vraies alertes que 500 fausses.

❓ Quelle technique utilisez-vous pour réduire les faux positifs dans vos alertes ?



7- Correlating Logs – Why and How

Understanding Log Correlation in a SOC Environment

🧠 What is a log correlation?

It's the process of linking related events from multiple systems to uncover patterns that may indicate security threats. Individually, logs may seem benign, but together, they often reveal suspicious activity.

💡 Why it matters in a SOC:

SOC analysts use correlation to detect advanced attacks like lateral movement, privilege escalation, or multi-stage intrusions. It helps identify coordinated threats beyond isolated alerts.

⚙️ How it's done:

SIEM tools like Splunk, ELK Stack, or Microsoft Sentinel apply correlation rules to match attributes (IP, user ID, timestamp, event type) across multiple sources (e.g., firewall, endpoint, Active Directory).

🔒 Pro Tip: Tailored correlation rules reduce false positives and improve threat visibility.

❓ Question: Have you customized correlation rules in your SIEM? Which log sources give you the most insight?

Corrélation des journaux – Pourquoi et comment

🧠 Qu'est-ce que la corrélation de logs ?

C'est le fait de relier des événements de sources diverses pour détecter des schémas indiquant une menace. Pris seuls, les logs paraissent anodins, mais ensemble, ils révèlent des comportements suspects.

💡 Pourquoi c'est important dans un SOC :

Elle permet de détecter des attaques complexes comme les mouvements latéraux, les élévations de privilège ou les intrusions en plusieurs étapes.

⚙️ Comment on le fait :

Des outils comme Splunk, ELK ou Sentinel appliquent des règles de corrélation pour relier des attributs communs (IP, utilisateur, horodatage, type d'événement) entre les sources (pare-feu, endpoints, AD...).

🔒 Astuce : Des règles bien conçues réduisent les faux positifs et augmentent la visibilité sur les menaces.

❓ Question : Avez-vous déjà personnalisé des règles de corrélation dans votre SIEM ? Quelles sources de logs trouvez-vous les plus utiles ?

Correlating Logs – Why and How



8- Log Enrichment Using Threat Intelligence

Enriching Logs with Threat Intelligence – Why and How

🧠 What is log enrichment?

Log enrichment means adding contextual or external data to raw log entries. When combined with threat intelligence, it helps analysts understand the who, what, and why behind suspicious indicators like IPs or hashes.

💡 Why it matters in a SOC:

Raw logs alone rarely tell the full story. Enriching logs with threat intel can reveal whether an IP is a known C2 server, a hash matches known malware, or a domain is linked to phishing. This makes alerts more meaningful and actionable.

⚙️ How it's done:

SIEM or SOAR platforms can ingest threat intel feeds from tools like MISP, VirusTotal, or commercial CTI providers. Integration happens via APIs and can occur during ingestion or when querying alerts, improving triage and incident response.

🔒 Pro Tip:

Automated enrichment ensures known IOCs are tagged in real time — saving time and reducing manual analysis.

❓ Do you enrich logs automatically during ingestion, or do you use on-demand lookups? Which CTI platform do you rely on?

Enrichissement des logs avec le renseignement sur les menaces – Pourquoi et comment

🧠 Qu'est-ce que l'enrichissement de logs ?

C'est le fait d'ajouter des données contextuelles ou externes aux logs bruts. Associé à de l'intelligence sur les menaces, cela permet de mieux comprendre l'origine ou la dangerosité d'un indicateur (IP, domaine, empreinte...).

💡 Pourquoi c'est important dans un SOC :

Un log brut ne suffit pas. L'enrichissement permet de savoir si une IP est un serveur C2, si une empreinte correspond à un malware, ou si un domaine est utilisé pour du phishing. On passe ainsi d'un simple événement à une alerte qualifiée.

⚙️ Comment on le fait :

Des outils comme MISP, VirusTotal, ou OTX peuvent être intégrés à un SIEM/SOAR via API. L'enrichissement peut être fait à l'ingestion ou au moment de la recherche pour améliorer l'analyse et la réponse.

🔒 Astuce :

L'automatisation de l'enrichissement permet de gagner du temps et d'éviter les oubliés dans la détection d'IOC connus.

❓ Préférez-vous un enrichissement automatique ou manuel ? Quels outils CTI utilisez-vous ?



9- SIEM Use Case: Detecting Anomalous Logins

Detecting Anomalous Logins in a SIEM – Why and How

🧠 What is an anomalous login?

An anomalous login is a deviation from normal authentication behavior — like logins from unusual geolocations, during non-working hours, or from unauthorized devices.

💡 Why it matters in a SOC:

Credential misuse is a common attack vector. Detecting abnormal login behavior helps identify compromised accounts, insider threats, or lateral movement — often before any damage is done.

⚙️ How it's done:

SIEMs like Splunk, ELK Stack, or Microsoft Sentinel use correlation rules and detection logic (e.g., failed logins + success from different countries) and link with identity providers like AD, Okta, or Azure AD. Threat feeds may enrich source IPs.

🔒 Pro Tip:

Build baseline profiles per user/role (e.g., login times, locations). Alert on deviations rather than raw events for better precision.

❓ Do you use geolocation, time-based thresholds, or behavior models in your detection rules? What's your favorite use case?

Détection de connexions anormales avec un SIEM – Pourquoi et comment

🧠 Qu'est-ce qu'une connexion anormale ?

Il s'agit d'un comportement d'authentification inhabituel : heure inhabituelle, localisation inconnue, appareil non autorisé...

💡 Pourquoi c'est important dans un SOC :

Les identifiants compromis sont très utilisés par les attaquants. Identifier des connexions inhabituelles permet de détecter un compte compromis, une menace interne ou un mouvement latéral.

⚙️ Comment on le fait :

Les SIEM (Splunk, ELK, Sentinel...) utilisent des règles de corrélation (ex : échecs + succès depuis un autre pays) et croisent les données des fournisseurs d'identité (AD, Azure AD, Okta). Les IPs peuvent être enrichies via des flux de menaces.

🔒 Astuce :

Créez un profil de connexion normal par utilisateur ou rôle (ex : horaires, emplacement).

Déclenchez des alertes sur les écarts, pas sur chaque événement.

❓ Utilisez-vous la géolocalisation, les plages horaires ou les modèles de comportement ? Quel cas d'usage vous semble le plus pertinent ?



10- SIEM Query Examples: KQL, Lucene, DSL

Using Query Languages in a SIEM – Why and How

🧠 What are KQL, Lucene, and DSL?

SIEMs use query languages to extract, correlate, and filter log data.

KQL (Kusto Query Language) is used in Microsoft Sentinel.

Lucene is the foundation of ELK (Elasticsearch).

DSL (Domain Specific Language) is used to structure queries in JSON format in Elasticsearch.

💡 Why it matters in a SOC:

SOC analysts rely on queries to hunt for threats, investigate alerts, and create dashboards.

Mastering these syntaxes improves detection accuracy and incident response speed.

⚙️ How it's done:

You use filters like where, match, or must clauses to find suspicious patterns (e.g., multiple failed logins, logins from rare countries, or file access by privileged users).

🔒 Pro Tip:

Save and reuse tested queries for common investigations. Turn them into detection rules or dashboards to enhance your SOC's automation and visibility.

❓ Which query language do you use most in your SIEM? Have you created custom detection rules based on queries?

Langue de requête dans un SIEM – Pourquoi et comment

🧠 Qu'est-ce que KQL, Lucene et DSL ?

Les SIEM utilisent des langages de requête pour filtrer, corréler et analyser les logs.

KQL (Kusto Query Language) est utilisé dans Microsoft Sentinel.

Lucene est la base d'Elasticsearch (ELK).

DSL permet de structurer les requêtes Elasticsearch en format JSON.

💡 Pourquoi c'est important dans un SOC :

Les analystes SOC utilisent ces langages pour investiguer, rechercher des menaces ou créer des tableaux de bord. Les maîtriser améliore la rapidité et la qualité de la détection.

⚙️ Comment on le fait :

On utilise des filtres comme where, match ou must pour détecter des comportements suspects (ex : connexions échouées multiples, accès depuis des pays inhabituels, activités anormales d'un utilisateur admin).

🔒 Astuce :

Conservez vos requêtes utiles, réutilisez-les, et transformez-les en règles de détection ou en dashboards pour automatiser vos analyses.

? Quel langage de requête utilisez-vous le plus ? Avez-vous déjà créé des règles de détection à partir de vos requêtes ?

SIEM Query Examples: KQL, Lucene, DSL

<pre>SecurityEvent where LogonType == 3</pre> <p>KQL Microsoft Sentinel</p>	<pre>event.type: authentication AND status: failure</pre> <p>LUCENE Elasticsearch</p>	<pre>{ "query": "match": { "country": 'F' } }</pre> <p>DSL Elasticsearch </p>
--	--	---