세션 1

SIEM · ELK 개요

구성요소 · SIEM 개념 · 아키텍처와 데이터 흐름

01
SIEM 개념
보안 로그를 한곳에 모아 상관분석·탐지·대응

02
ELK 구성
수집·저장·검색·시각화 4대 구성요소

03
아키텍처
로그 발생부터 알럿까지의 데이터 흐름

04
검색 엔진
인덱스·역색인과 Discover 첫 검색

S1

SIEM · ELK 개요

세션 개요와 학습 목표

> SIEM의 전체 그림을 잡고 이후 세션에서 쓸 공통 용어를 확립한다.

이 세션에서 얻는 것

- **SIEM 정의**: 흩어진 보안 로그를 한곳에 모아 상관분석·탐지·대응하는 체계가 무엇인지 설명할 수 있다.
- **구성요소 구분**: 수집·저장·검색·시각화 네 단계가 각각 어떤 역할을 맡는지 나눠 이해한다.
- **데이터 흐름**: 로그가 발생해 인덱스에 적재되고 검색·시각화·알럿으로 이어지는 경로를 그릴 수 있다.

실습 환경 전제

- **공용 분석 서버**: ELK/OpenSearch는 고사양이라 개인 PC에 설치하지 않고 브라우저로 공용 서버에 접속해 실습한다.
- **실시간 로그**: 서버가 2초마다 웹·보안 로그를 흘려보내므로 접속 즉시 실제 데이터로 분석을 시작한다.

S1

SIEM · ELK 개요

SIEM이란 무엇인가

> SIEM은 보안 정보와 이벤트를 통합 관리해 위협을 가시화하는 체계다.

정의와 목적

- **SIEM 의미**: Security Information & Event Management의 약자로, 로그·이벤트를 통합해 탐지·대응·감사까지 지원한다.
- **통합 가시성**: 방화벽·웹서버·인증 로그가 서버마다 흩어져 있으면 공격의 전체 그림을 볼 수 없어 한곳으로 모은다.
- **상관분석**: 개별로는 정상인 이벤트도 시간·IP·계정으로 엮으면 브루트포스 같은 공격 패턴이 드러난다.

ELK와의 관계

- **스택으로 구현**: ELK/OpenSearch 스택으로 수집·저장·검색·시각화를 갖추면 그 자체가 SIEM 플랫폼이 된다.
- **탐지 규칙**: 저장된 로그에 저장 검색·알럿 규칙을 얹으면 자동 탐지·통보 체계로 확장된다.

S1

SIEM · ELK 개요

로그 관리와 보안 관제의 필요성

> 로그는 사고가 나기 전과 후를 잇는 유일한 증거이자 탐지 재료다.

```
[서버마다 흩어진 로그]
web-01:  /var/log/nginx/access.log
web-02:  /var/log/nginx/access.log
auth:    /var/log/auth.log
waf:     /var/log/modsec_audit.log

문제: 어느 서버가 언제 공격받았나?
  → 서버마다 grep 은 불가능

[SIEM 으로 통합]
모든 로그 → 인덱스 logs-demo-*
  → 한 화면에서 검색·집계·상관분석
  → client_ip 기준 전 서버 활동 추적
```

왜 필요한가

- **사고 조사**: 침해 발생 시 언제·어디서·무엇이 일어났는지 로그만이 시간순으로 증명한다.
- **실시간 탐지**: 흩어진 로그를 모아야 임계치·이상 징후를 실시간으로 잡아낼 수 있다.
- **규정 준수**: 접근·인증 로그 보존과 감사는 다수 보안 규정의 필수 요구사항이다.

S1

SIEM · ELK 개요

4대 구성요소 — 수집·저장·검색·시각화

SIEM 파이프라인은 네 단계가 순서대로 물려 돌아간다.

단계	역할	ELK	OpenSearch(이 랩)
수집	로그를 모아 파싱·정규화해 전송	Logstash · Beats	Data Prepper · Fluent Bit
저장	문서를 인덱스에 저장·보존	Elasticsearch	OpenSearch
검색	역색인으로 초고속 검색·집계	Elasticsearch	OpenSearch
시각화	Discover·Visualize·Dashboard·Alerting	Kibana	OpenSearch Dashboards

ELK와 OpenSearch의 관계

두 스택은 뿌리가 같아 화면과 API가 거의 동일하다.

항목	비교
기원	OpenSearch는 Elasticsearch 7.10 포크
엔진 포트	양쪽 모두 9200
대시보드 포트	양쪽 모두 5601
라이선스	OpenSearch는 Apache 2.0
화면	Discover·Visualize 거의 동일

실습 관점

- **학습 이전성**: 이 랩은 OpenSearch지만 개념·화면이 같아 Elastic 정품 ELK에도 그대로 통한다.
- **용어 혼용**: 인덱스·문서·KQL 등 핵심 용어는 양쪽에서 동일하게 쓰인다.

S1

SIEM · ELK 개요

구성요소별 역할 심화

> 각 구성요소가 무엇을 입력받아 무엇을 내보내는지 정확히 구분한다.

앞단 — 수집과 저장

- **수집기**: 원본 로그 라인을 필드로 파싱하고 정규화해 저장 엔진으로 안정적으로 전송한다.
- **저장 엔진**: 문서를 인덱스에 넣고 역색인을 만들어 대량 데이터에서도 밀리초 단위 검색을 가능하게 한다.

뒷단 — 검색과 시각화

- **검색·집계**: KQL·DSL로 필터링하고 Terms·Date Histogram 같은 집계로 통계를 산출한다.
- **시각화·관제**: Discover로 원본을 보고 Visualize·Dashboard로 관제 화면을 만들며 Alerting으로 통보한다.

S1

SIEM · ELK 개요

SIEM 데이터 흐름 파이프라인

로그 한 건이 발생해 알럿이 되기까지 여섯 단계를 거친다.

흐름의 핵심

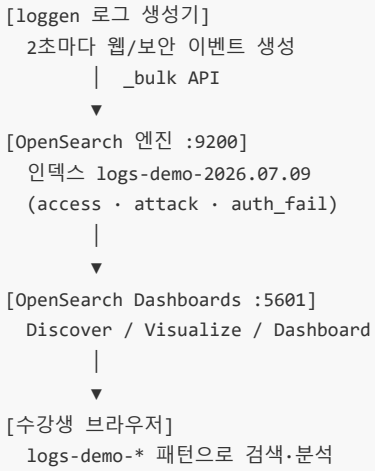
- **단방향 축적**: 로그는 계속 쌓이며, 앞 단계 산출물이 다음 단계의 입력이 되는 파이프라인이다.
- **인덱스가 중심**: 모든 검색·시각화·알럿은 결국 인덱스에 적재된 문서를 대상으로 동작한다.
- **탐지로 귀결**: 시각화는 관찰, 알럿은 자동화로, 최종 목표는 위협의 조기 탐지다.

S1

SIEM · ELK 개요

이 랩의 아키텍처

> 로그 생성기가 실시간으로 인덱스를 채우고 우리는 그 데이터를 분석한다.



구성 이해

- **자동 주입**: loggen이 access·attack·auth_fail 이벤트를 섞어 계속 적재하므로 항상 신선한 데이터가 있다.
- **일자별 인덱스**: 인덱스명이 날짜로 끝나 logs-demo-*로 모으면 여러 날을 한 번에 검색한다.
- **기본 패턴**: 인덱스 패턴 logs-demo-*는 기본값으로 지정돼 있어 바로 Discover에서 시작한다.

배포 형태 — 단일 노드와 클러스터

> 학습은 단일 노드로 충분하지만 운영은 다중 노드로 확장한다.

배포 유형

- **단일 노드**: 이 랩처럼 하나의 엔진에 discovery.type=single-node로 띄우며 학습·PoC에 적합하다.
- **클러스터**: 운영은 여러 노드로 샤드를 분산해 대용량·고가용성을 확보하고 노드 장애에 대비한다.
- **역할 분리**: 마스터·데이터·코디네이팅 노드로 역할을 나눠 안정성과 성능을 함께 확보한다.

리소스 주의

- **메모리 요구**: 엔진은 수 GB 힙을 쓰므로 개인 PC 대신 공용 서버에서 실습하도록 구성했다.
- **힙 설정**: 자바 힙은 물리 메모리의 50% 이하로 잡아 OOM 종료를 예방한다.

S1

SIEM · ELK 개요

검색 엔진의 기본 — 문서와 인덱스

> 관계형 DB와 달리 검색 엔진은 JSON 문서를 인덱스에 담는다.

핵심 개념

- **문서**: 로그 한 건이 하나의 JSON 문서이며 @timestamp·event·client_ip 같은 필드로 구성된다.
- **인덱스**: 유사한 문서를 모아 담는 논리적 컨테이너로, 이 랩은 logs-demo-YYYY.MM.DD를 쓴다.
- **매핑**: 각 필드의 타입(date·keyword·ip·number·text)을 정의해 검색·집계 동작을 결정한다.

RDB 대비

- **유연한 스키마**: 문서마다 필드 구성이 달라도 되어 이질적 로그를 한 인덱스에 담기 쉽다.
- **검색 최적화**: 조인 대신 역색인으로 전문 검색과 집계를 대규모로 빠르게 수행한다.

S1

SIEM · ELK 개요

역색인이 빠른 이유

> 역색인은 값에서 문서로 거꾸로 가리켜 검색을 즉시 끝낸다.

```
[원본 문서]
doc1: event=attack  ip=203.0.113.7
doc2: event=access  ip=10.0.0.9
doc3: event=attack  ip=10.0.0.9

[역색인 – 값 → 문서 목록]
event:
  "attack" -> [doc1, doc3]
  "access" -> [doc2]
client_ip:
  "203.0.113.7" -> [doc1]
  "10.0.0.9"    -> [doc2, doc3]

질의 event:attack
  → 곧바로 [doc1, doc3] 반환
  (전체 스캔 없이 목록 조회)
```

동작 원리

- **사전 색인**: 적재 시점에 각 필드 값을 문서 목록에 미리 매핑해 두어 질의가 조회로 끝난다.
- **전체 스캔 회피**: 모든 문서를 훑지 않고 해당 값의 문서 목록만 읽어 대량에서도 빠르다.
- **집계 가속**: 값별 문서 목록 구조 덕분에 Terms 집계 같은 통계도 신속하게 계산된다.

Discover로 첫 검색 시작하기

> Discover는 인덱스의 원본 문서를 시간순으로 탐색하는 출발점이다.

접속 절차

- **대시보드 로그인**: 공용 서버에 접속해 제공된 계정으로 로그인한 뒤 좌측 메뉴에서 Discover를 연다.
- **인덱스 패턴**: 좌상단이 logs-demo-*인지 확인하며, 이 패턴은 기본값으로 이미 지정돼 있다.
- **시간 범위**: 우상단을 Last 1 hour로 두고 Refresh하면 수천 건의 최근 로그가 나타난다.

화면 읽기

- **필드 목록**: 좌측 Available fields에서 필드를 클릭하면 상위 값 분포가 뜨고 컬럼으로 추가할 수 있다.
- **UTC 주의**: 로그는 UTC로 저장되므로 데이터가 안 보이면 시간 범위를 넓혀 Refresh한다.

S1

SIEM · ELK 개요

세션 정리와 다음 예고

> SIEM의 정의·구성·흐름을 하나의 그림으로 묶어 정리한다.

기억할 것

- **SIEM은 통합**: 흩어진 보안 로그를 한곳에 모아 상관분석·탐지·대응하는 체계이며 ELK 스택으로 구현한다.
- **4대 구성**: 수집·저장·검색·시각화가 파이프라인으로 물려 로그를 가치 있는 탐지 정보로 바꾼다.
- **역색인이 핵심**: 값에서 문서로 거꾸로 가리키는 역색인 덕분에 대량 로그도 즉시 검색·집계된다.

다음 예고 — 세션 2

- **로그 스키마**: 실습에서 쓸 필드 레퍼런스를 익히고 text와 keyword 타입의 차이를 이해한다.
- **적재 실습**: _bulk API로 문서를 직접 넣어 인덱싱과 정규화 과정을 손으로 확인한다.