빅데이터 분석기술을 활용한 이상징후탐지시스템(FDS) 구축

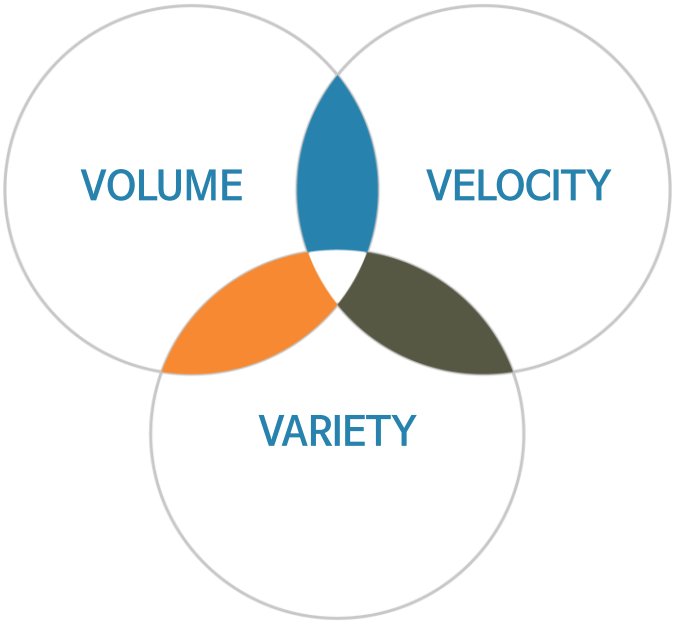
2015. 04. 16

주식회사 앤서

빅데이터?!

1. 사전적 의미의 빅데이터

» 빅데이터는 단순히 큰 데이터가 아닌, 크고, 빠르고 복잡한 형태를 지닌 데이터를 의미



VOLUME

빠른 속도로 증가하는 기업의 데이터 규모

-미국 기업 상당수의 데이터 규모 : 최소 100TB 이상
-데이터 규모 매년 140% 증가

VELOCITY

끊임없이 쏟아지는 새로운 데이터

-자동차 타이어 공기압, 연료 잔량 센서 : 100개+
-용광로 하나에서 1초에 쏟아지는 데이터 크기 : 1GB
-뉴욕 증권 거래소 하루 거래 데이터 크기 : 1TB

VARIETY

갈수록 다양해지는 데이터의 형태

-분석 기술의 발달로 정형/비정형 데이터 통합 분석
-서비스, 기술 발달에 따른 데이터 소스/형태의 다양화
-분석하지 못하고 버려졌던 데이터까지 분석의 영역으로

이제는 끊임없이 쏟아지는 다양한 대규모 데이터를 서로 엮어 실시간으로 분석할 수 있는 플랫폼이 필요합니다

2. 사업화를 위한 빅데이터의 구분

실시간 분석 vs 과거 데이터 분석

실시간으로 발생하는 대용량의 데이터 처리

인메모리 기술

FDS

보안 관제

내부정보 유출방지

장기간 보존된 대용량의 데이터 처리

분산처리 기술

CRM

성향분석

보고서

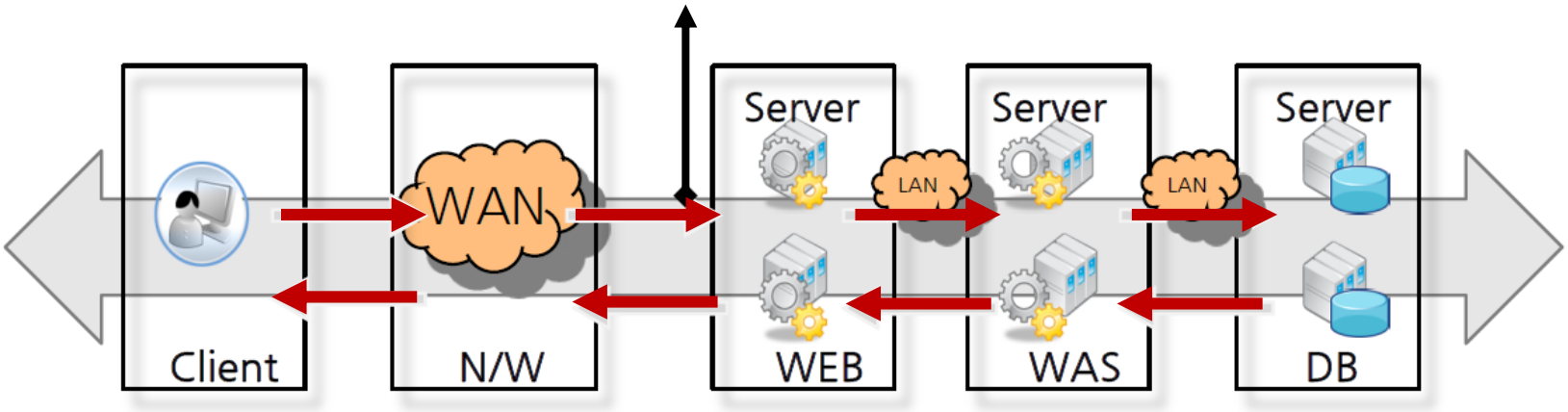
뭘 볼 것인가?

3. 목적에 맞는 제품 선정

예시1 : 서비스 품질관리

도입목적 : 핵심업무 10대 업무를 대상으로 업무 흐름에 따라 단위기능별 기능오류, 처리량, 응답속도를 측정하여 서비스 품질관제 시스템을 구축

구간별 실시간 Input / Output 응답속도 측정



3. 목적에 맞는 제품 선정

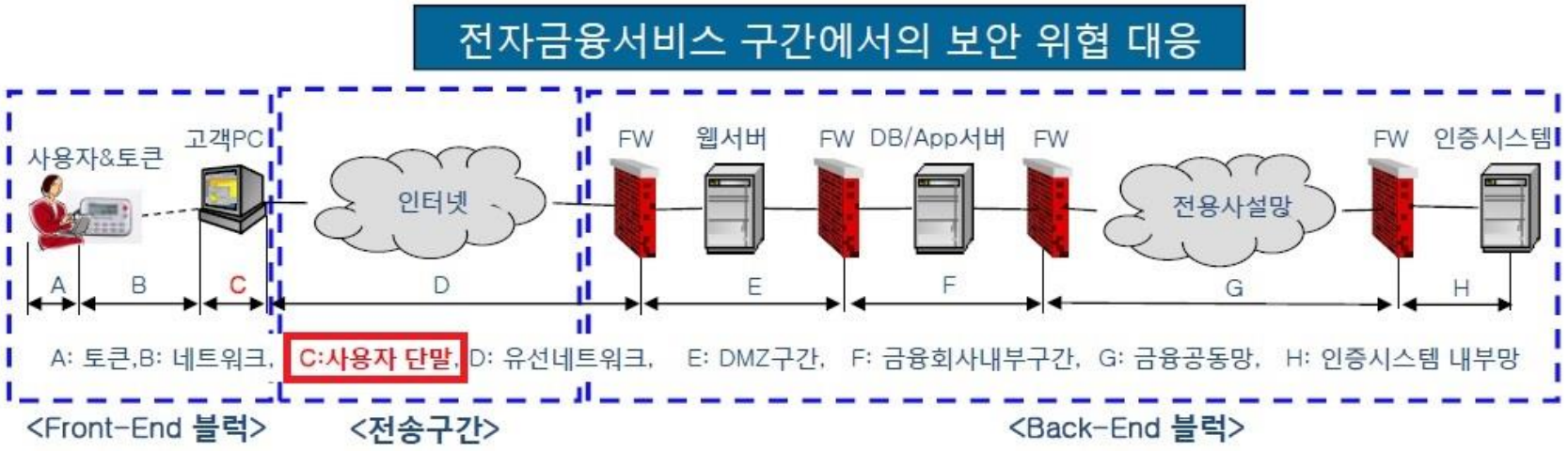
예시2 : 데이터 분석을 통한 버스노선 설정

도입목적 : 휴대폰 소지자의 등록주소와 심야에 활동지점 데이터를 공간적으로 분석하여 심야에 필요한 노선을 선정

3. 목적에 맞는 제품 선정

예시3 : 이상금융거래탐지시스템 구축

도입목적 : 해킹 및 고객정보 유출 사고를 통하여 보이스 피싱, 스미싱 등 고객에게 직접적으로 영향을 미칠 수 있는 2차 금융사고로 이어지지 않도록 빅데이터 기반의 통합 보안 모니터링 시스템을 구축하고 시나리오(룰, 패턴)를 통한 사전 예방 체계를 구축



FDS??
(빅데이터 관점 보안)

4. 빅데이터 관점의 보안

기존의 보안 분석 및 모니터링 시스템은 탐지/분석/대응 전과정에 걸쳐 변화된 보안 위협 환경에 부적절하여, 실시간 탐지, 최단 시간 내 분석으로 보안사고 사전 예방 및 전사적인 보안사고 대응체계를 강화할 필요가 있습니다.

모니터링 환경의 변화 필요성

위협의 고도화로 탐지/분석/대응에 장시간 소요

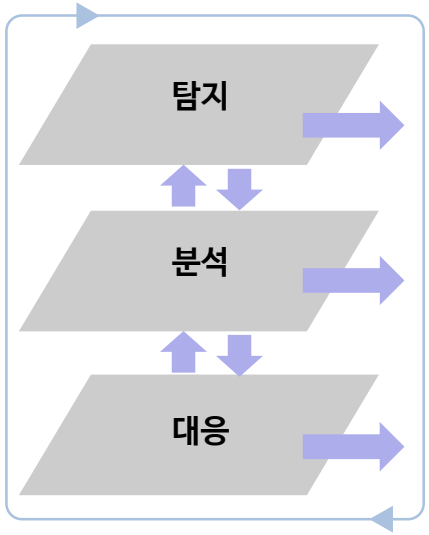
- 진화하는 공격으로 탐지 곤란
- 탐지/분석 위한 방대한 데이터 부족

BIG DATA 기반

- 시나리오 기반의 실시간 위협 탐지
- 실시간 분석 위한 방대한 데이터 인프라

기존 관제/모니터링의 한계

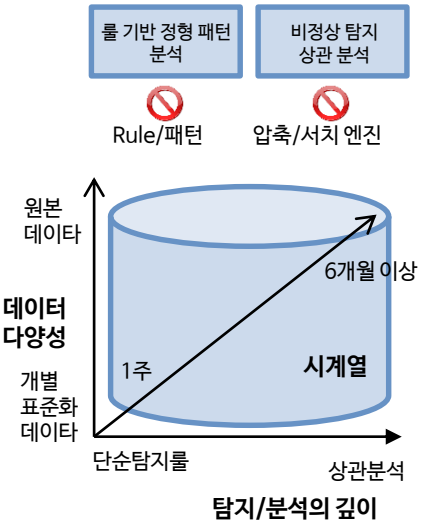
보안사고가 없는 것인가? 탐지가 안되는 것인가?



- ▶ 악성코드 1억 4천만 개
  → 탐지는 4천만 개
- ▶ **정상 사용자의** 메일, 웹링크를 통한 **우회 공격**
- ▶ 복수 시스템의 로그, 네트워크 데이터 필요
- ▶ 정상/비정상 구분 위한 **장기간 Baseline 데이터 필요**
- ▶ 임직원 대상 **구체적 변화관리 포인트 도출 어려움**

통합보안 관제/모니터링

보안사고에 대한 구체적인 탐지가능성 극대화

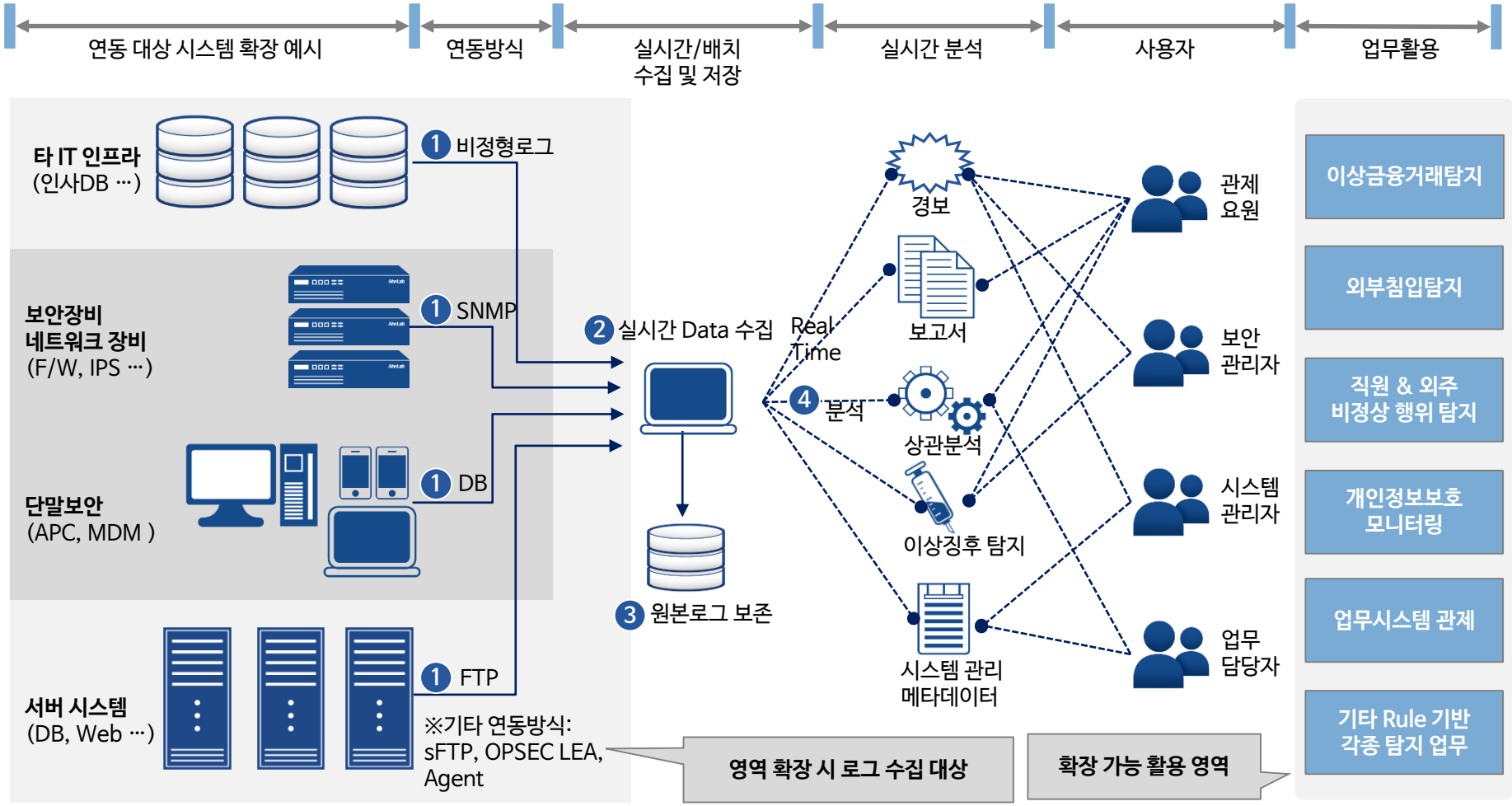


- **보안위협 탐지시간**
  → 실시간 탐지, 최단시간내 분석

  유형 1 – 정상사용자 우회 공격
  : 탐지불가 → 수시간 내 분석
  유형 2 – 악의적 내부자 유출
  : 탐지불가 → 실시간 징후 탐지
  유형 3 – 고도화된 Network 해킹
  : 탐지불가 → 실시간 복합 룰 탐지
- 탐지된 이벤트 기반 보안정책 변경, 직원 교육/훈련 등 **변화관리 활용**
- 탐지된 이벤트 기반으로 **기술적 예방책 마련**

4. 빅데이터 관점의 보안

분석 대상의 확대



Logpresso

5. Logpresso – 개요

대용량 비정형 데이터를 수집, 검색, 분석, 시각화하여 신속한 의사결정을 돕는 플랫폼

5. Logpresso - 주요기능1 (쿼리)

실시간 및 배치 처리에 필요한 4가지 유형 쿼리 지원

애드혹 쿼리

CLI나 GUI를 통하여 사용자가 직접 쿼리를 입력하거나, 클라이언트 SDK를 사용하여 프로그래핑 방식으로 임의의 쿼리를 직접 실행할 수 있습니다.

장시간 실행해야 하는 쿼리는 세션 종료와 관계없이 동작하도록 백그라운드 실행으로 전환하거나, 다시 포어그라운드로 전환하여 쿼리 결과를 확인할 수 있습니다.

실시간 쿼리

윈도우로 지정된 시간 동안 실시간으로 로거에서 수집되는 데이터, 실시간으로 테이블에 입력되는 데이터, 실시간으로 스트림 쿼리에서 출력되는 데이터를 대상으로 쿼리할 수 있습니다.

디스크에 데이터가 저장되어 있지 않은 상태에서도, 실시간 데이터를 원본으로 하여 쿼리를 수행할 수 있습니다.

스트림 쿼리

스트리밍 데이터에 대하여 쿼리가 무한히 실행됩니다. 스트리밍 가능한 커멘드로 구성된 경우 쿼리가 시스템 종료 시까지 무한히 실행되며, 통계나 정렬 등이 포함된 경우 쿼리 리프레시 간격에 맞추어 주기적으로 스트림 쿼리가 재시작됩니다.

스트림 쿼리는 쿼리가 재시작되더라도 빈틈없이 실행되므로 쿼리가 재시작되는 사이의 입력 데이터 손실 없이 정합성을 보장합니다.

스케줄 쿼리

예약된 일정에 맞추어 쿼리가 실행됩니다. 매시, 매일, 매주, 매월, 매년 혹은 특정 시각에 맞추어 배치 작업을 실행할 수 있습니다.

지정된 시각에 타 시스템의 데이터를 임포트하거나, 배치 쿼리 결과를 익스포트 하는 용도로 사용할 수 있습니다.

5. Logpresso - 주요기능1 (쿼리)

스트림 쿼리

스트리밍 모드: 시스템 종료 시까지 중단 없이 쿼리 실행

시나리오 예시: 실시간으로 수집되는 데이터를 정규화 후 다른 시스로그 서버로 중계

데이터 수신 → REX 정규표현식 필드 추출 → EVAL 라인 포멧팅 → SENDSYSLOG전송

리프레시 모드: 설정된 주기로 쿼리를 완료시키고 재생성

시나리오 예시: 실시간으로 수집되는 데이터에서 주기적으로 통계 추출 후 통계 테이블에 적재

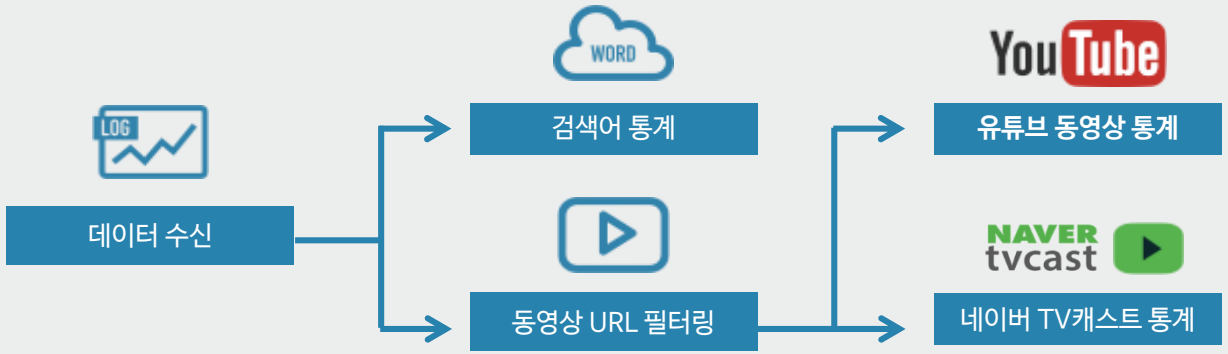
데이터 수신 → STATS 통계 추출 → IMPORT 테이블 적재

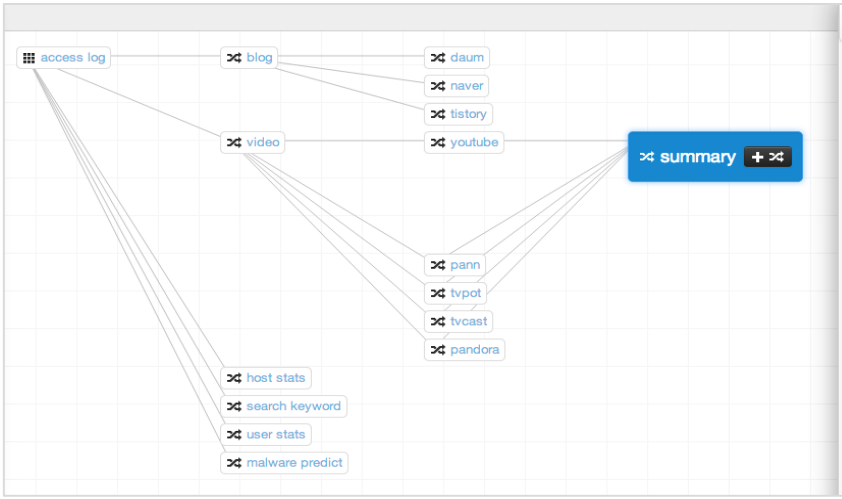
5. Logpresso - 주요기능1 (쿼리)

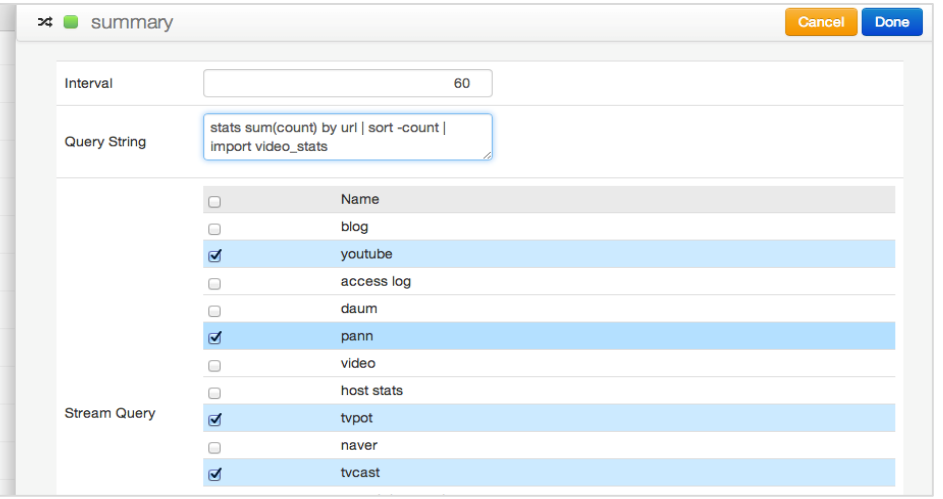
스트림 쿼리

스트림 쿼리 그래프

다수의 스트림 쿼리를 연결하여 부분적인 쿼리 결과를 공유하면서 고속 처리가 가능합니다.



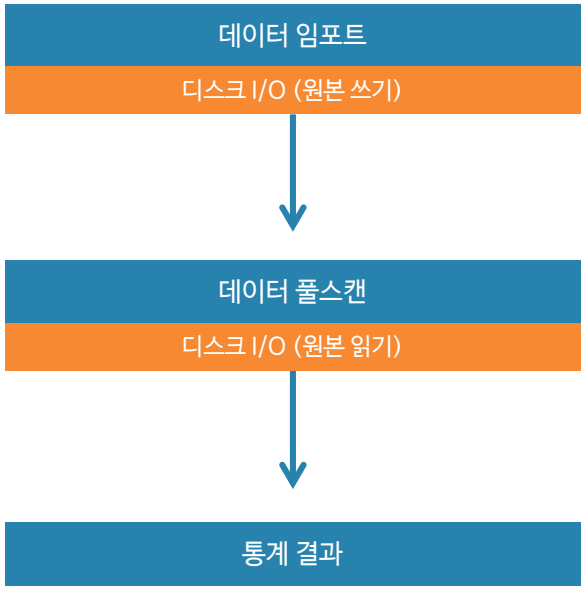




5. Logpresso - 주요기능2(실시간 써머리 로깅)

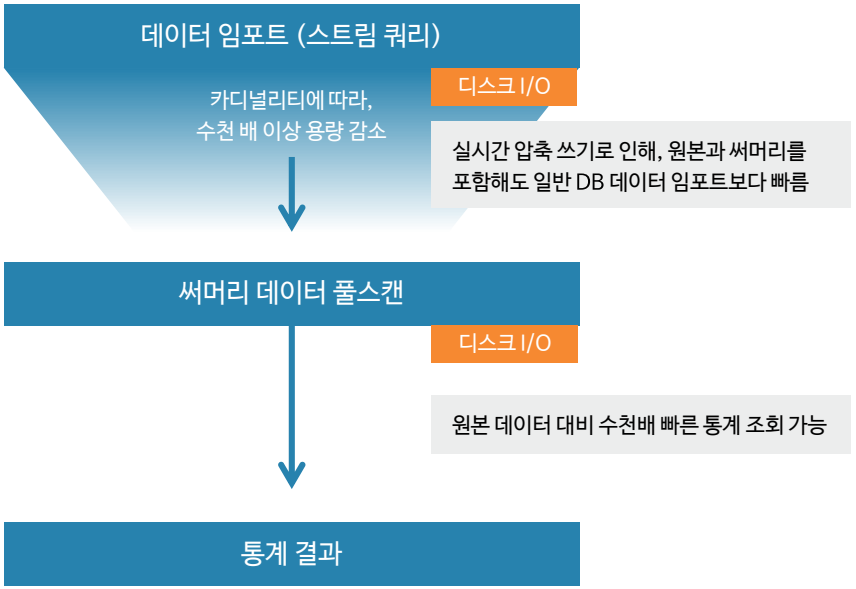
데이터 적재 단계에서 인메모리 통계 요약으로 추후 대용량 데이터 분석을 고속화

전처리 없는 통계 분석 방식



대시보드 리포팅이 정형화된 경우에도,
원본 데이터 전체를 쓰고 읽는 부하 발생

써머리 로깅 방식



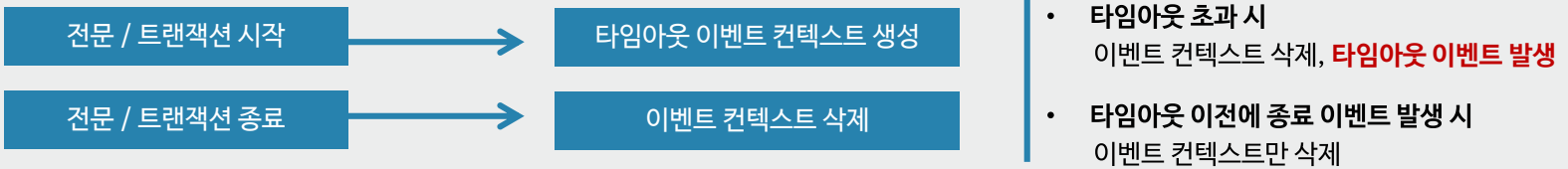
빈번하게 수행되는 쿼리를 대상으로,
입력 시점에 인메모리 통계 및 기록,
수천배 빠른 어낼리틱 쿼리 달성

5. Logpresso - 주요기능3(실시간 이벤트 처리)

이벤트 컨텍스트를 사용하여 실시간 이벤트 연관 분석을 수행할 수 있습니다.

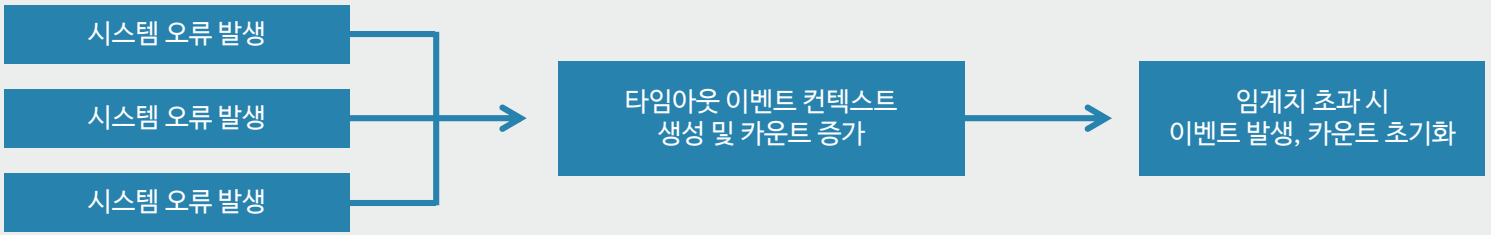
디스크 IO 없이 실시간으로 메모리에서 처리하므로 고속 이벤트 처리가 가능합니다.

1. 시작과 종료 확인

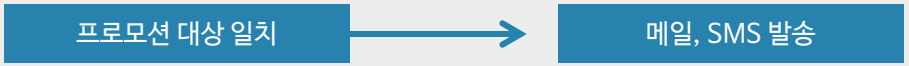


- **타임아웃 초과 시**
  이벤트 컨텍스트 삭제, **타임아웃 이벤트 발생**
- **타임아웃 이전에 종료 이벤트 발생 시**
  이벤트 컨텍스트만 삭제

2. 임계치 초과

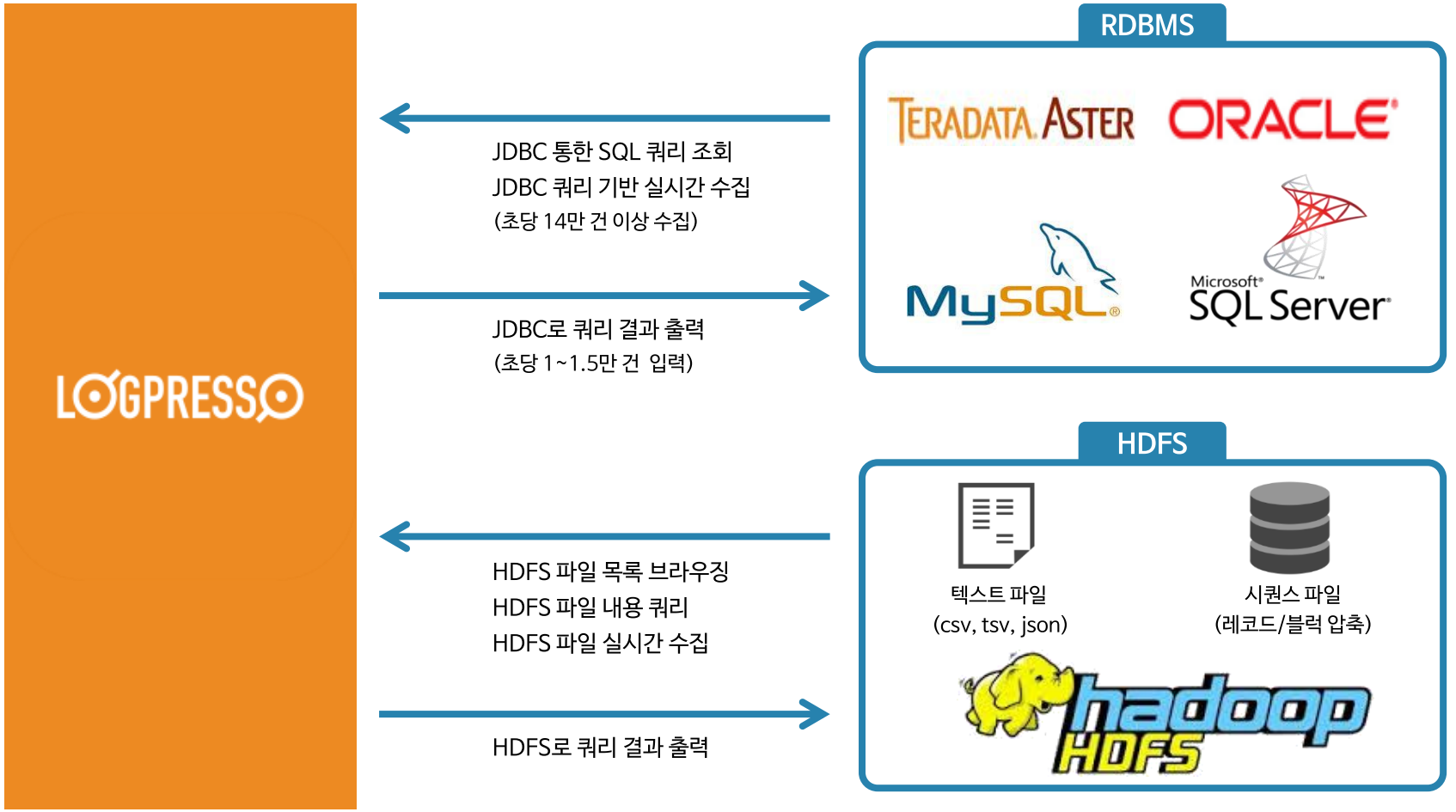


3. 조건 만족 시 자동 대응



5. Logpresso - 주요기능4(외부시스템 연동)

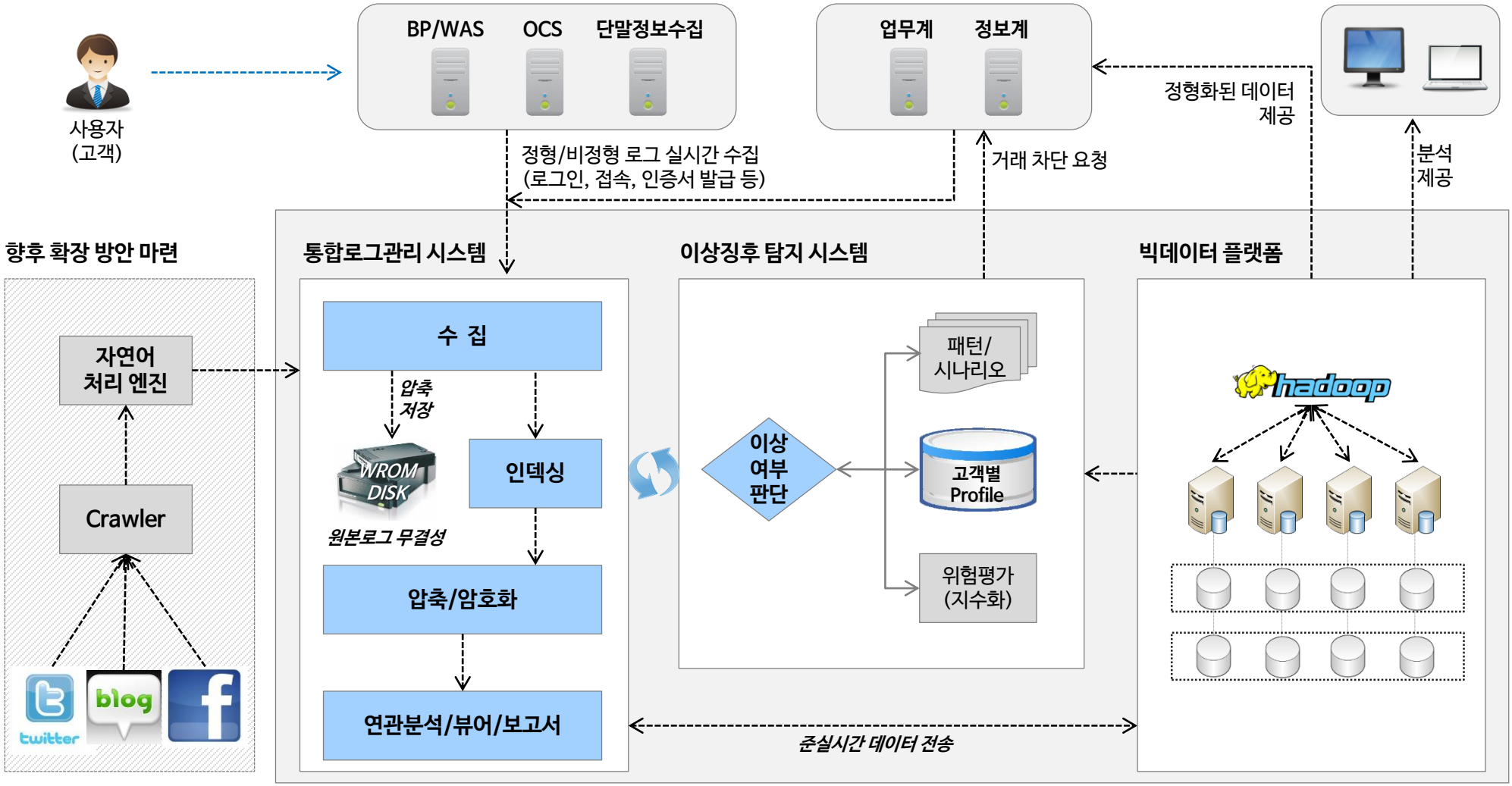
이벤트 탐지 결과는 RDBMS, Hadoop과 연동하여 신속하게 데이터 분석결과를 공유합니다.



구축 사례

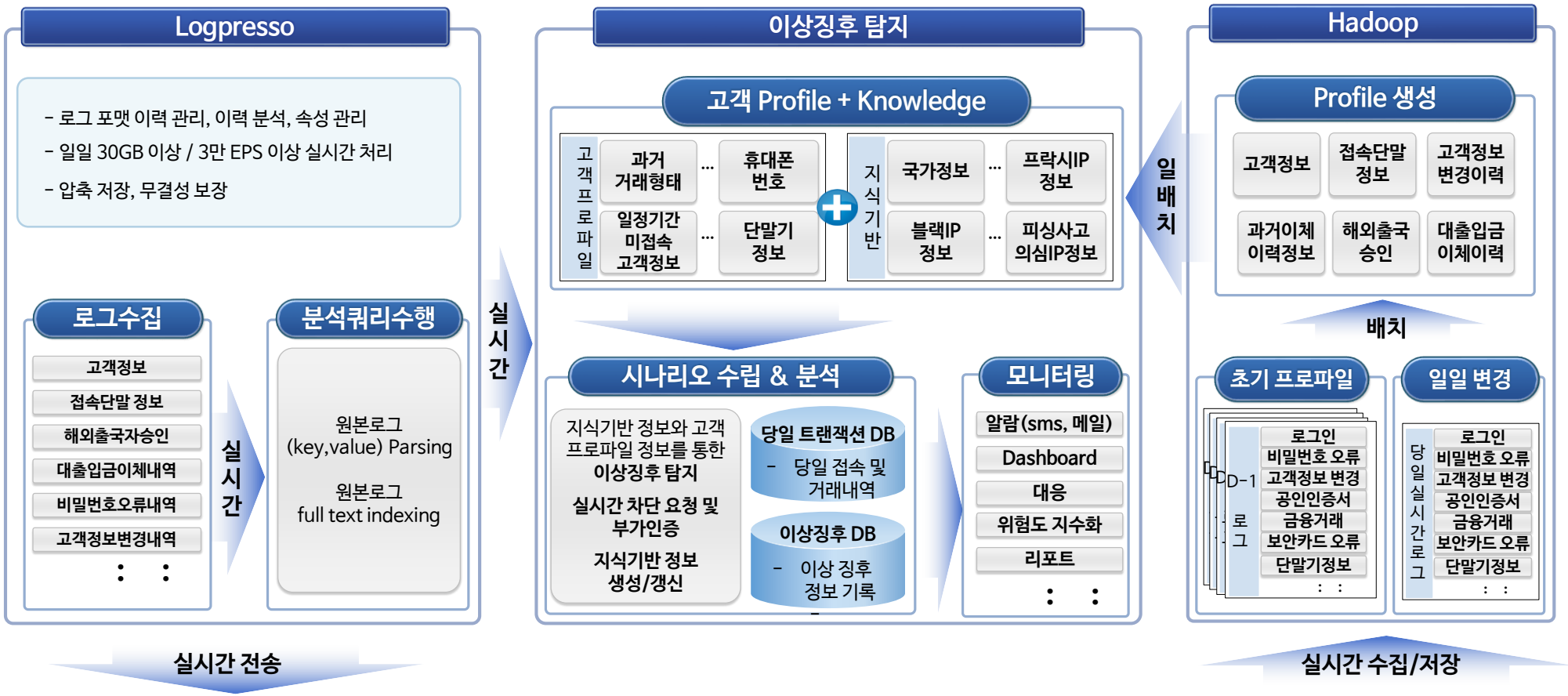
6. 구축사례 - 이상금융거래탐지시스템(1/3)

OOOO증권



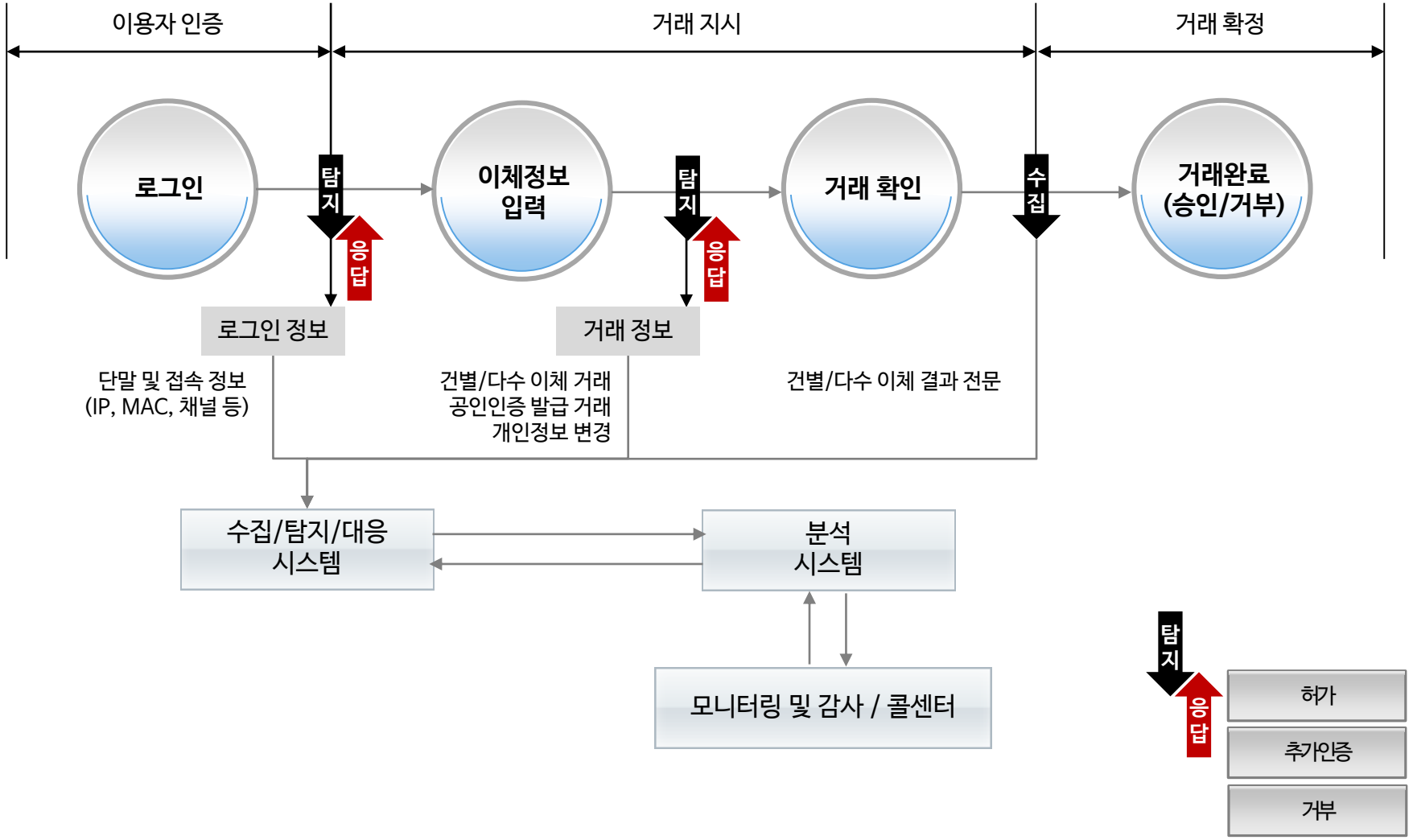
6. 구축사례 - 이상금융거래탐지시스템(2/3)

OOOO증권



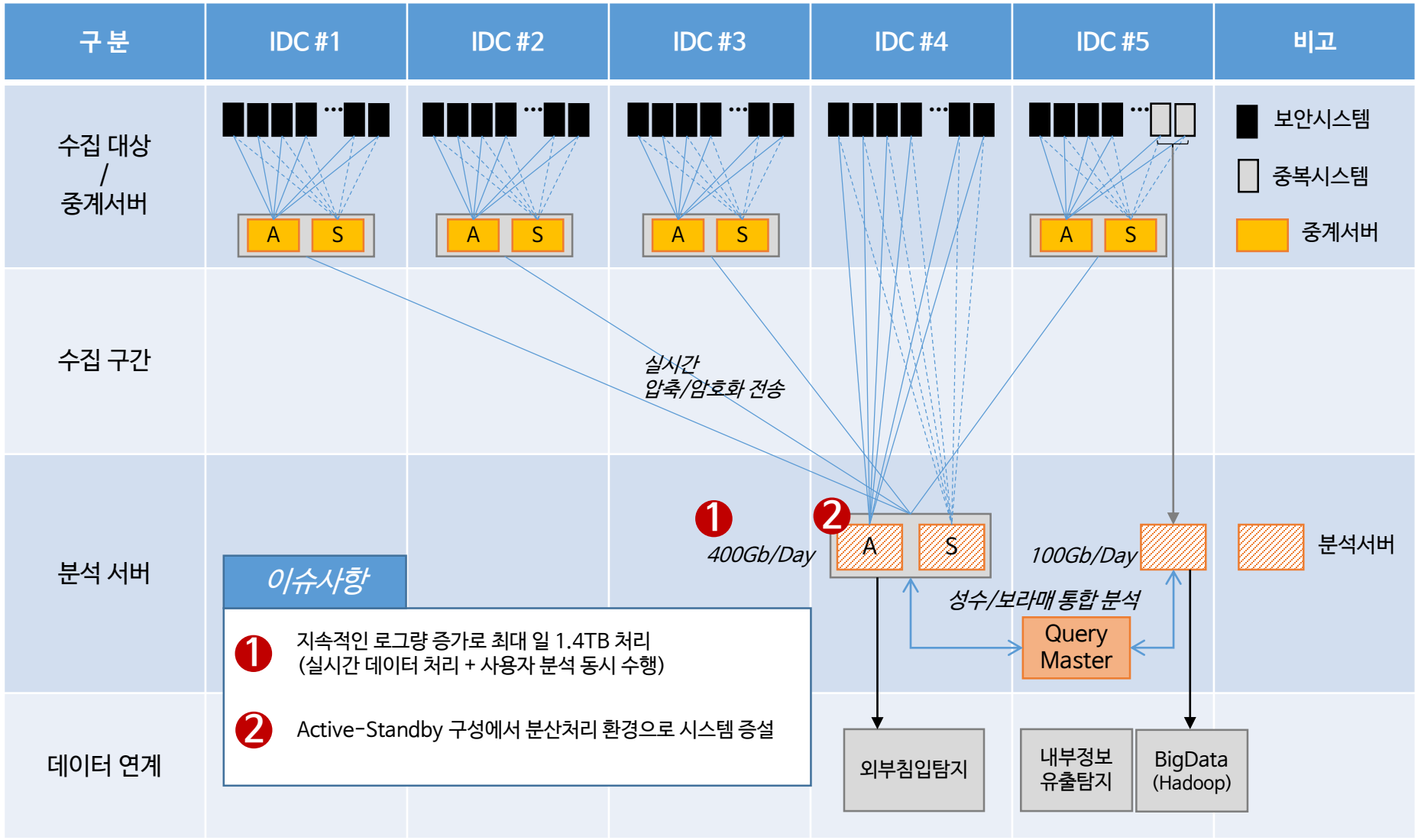
6. 구축사례 - 이상금융거래탐지시스템(3/3)

OO은행



6. 구축사례 - 보안로그 통합 관리(1/3)

SOOOO社



6. 구축사례 - 보안로그 통합 관리(2/3)

SOOOO社

2. 탐지시나리오 - No.2

- 취업관련사이트에 수시로 접근하는 내부임직원(App. F/W)이 일일 20개 이상의 문서파일을 일주일간 지속적으로 외부로 발송(App.F/W)

① 취업관련사이트 접속 and 문서파일 외부 전송 사용자 리스트

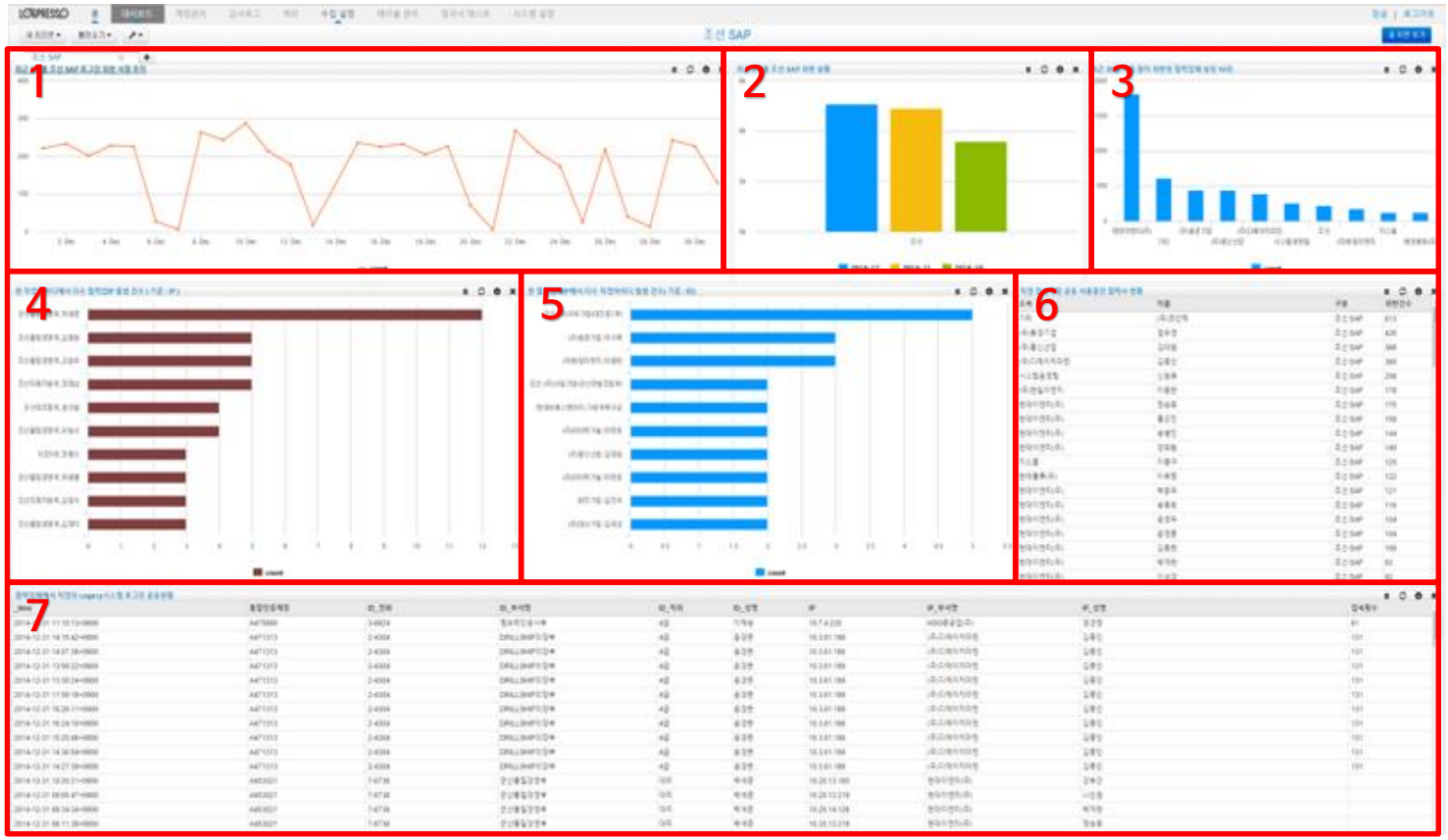
② 최근 24시간 취업관련사이트 접속자 Top 10

③ 최근 24시간 문서파일 외부전송 사용자 Top 10

6. 구축사례 - 보안로그 통합 관리(3/3)

OO중공업

SAP 분석 시나리오

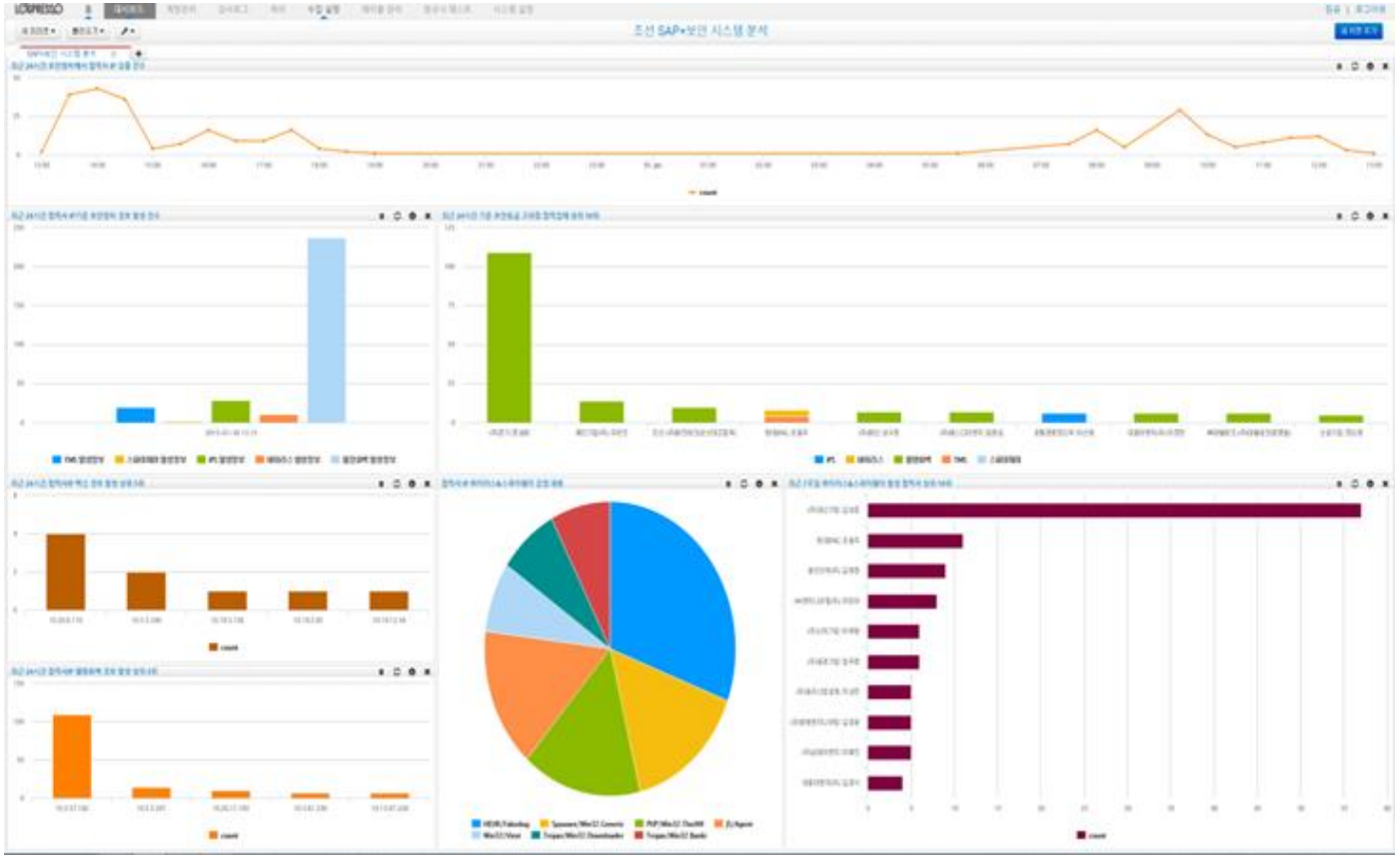


① 최근 1개월 조선SAP 로그인 위반
② 최근 3개월 조선 SAP 위반
③ 최근 1개월 위반 협력업체 TOP10
④
⑤
⑥
⑦ 협력업체 직영의 Legacy 시스템 로그인 공유현황

6. 구축사례 - 보안로그 통합 관리(3/3)

OO중공업

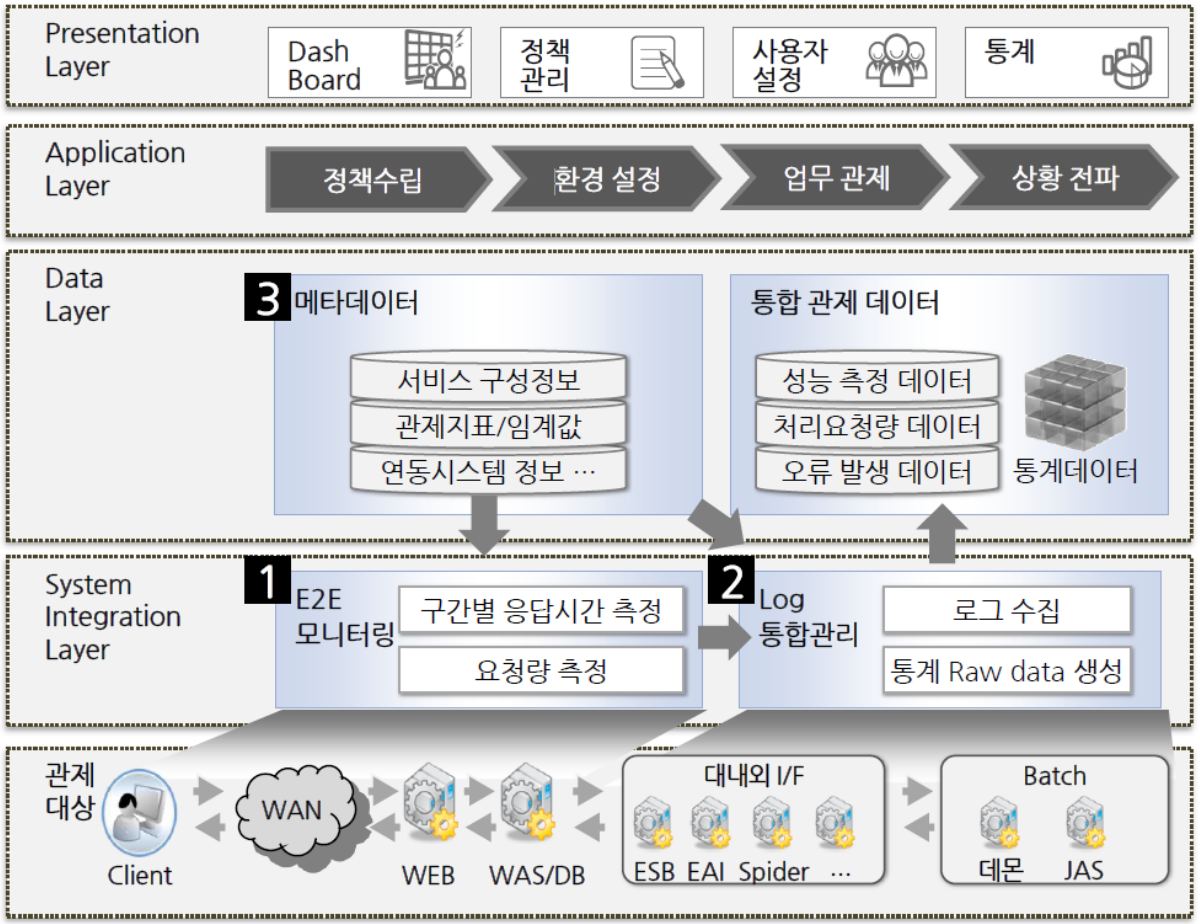
SAP + 보안시스템 분석 시나리오



① 최근 24시간 보안장비에서 검출된 협력사 추이
② 최근 24시간 협력사 IP 기준 보안장비 경보 발생 건수
③ 최근 24시간 고위험 협력업체 상위 TOP10
④
⑤
⑥
⑦ 최근 7일 악성코드 발생 협력사 상위 TOP10

6. 구축사례 - 서비스 품질 관리(1/2)

LOOOOO社



1 E2E 모니터링

- 서비스 및 2,300여개의 단위기능 별 고객 체감 응답시간과 서버 응답시간 측정

2 Log 통합관리

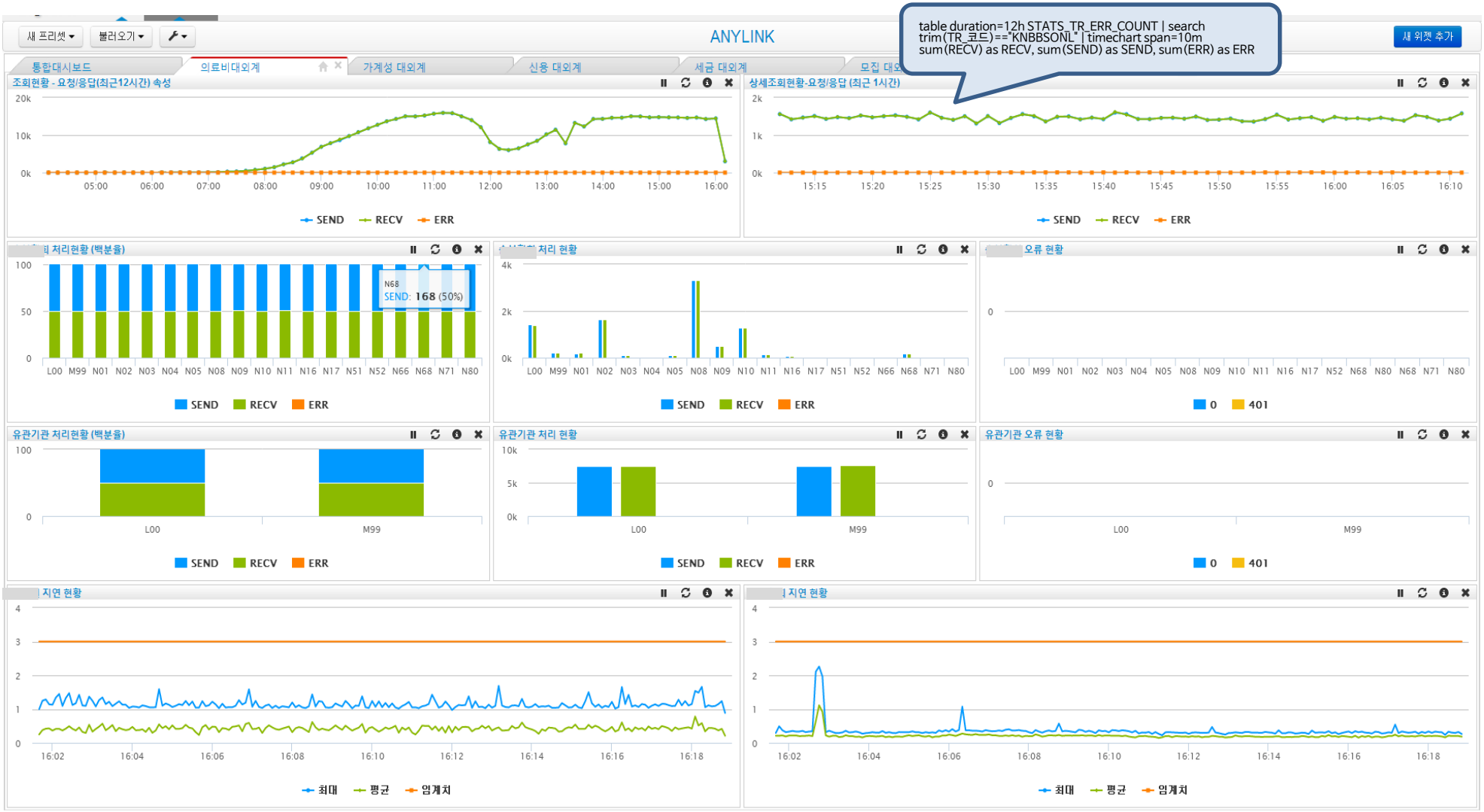
- 대내외 I/F, Batch 관리 솔루션에서 발생하는 대용량 로그데이터에 대한 수집정책 관리 및 수집/분석

3 메타데이터

- 서비스 구성요소 및 연관 정보 관리를 통한 서비스 Visibility 제공

6. 구축사례 - 서비스 품질 관리(2/2)

OOOO협회



감사합니다