2016. 12

Big Data
Stream Analytics
-
LOGPRESSO

목차

1. 로그프레소 개요

수집부터 시각화까지 의사결정에 필요한 모든 단계를 지원하는 실시간 빅데이터 웨어하우스 입니다.

수집

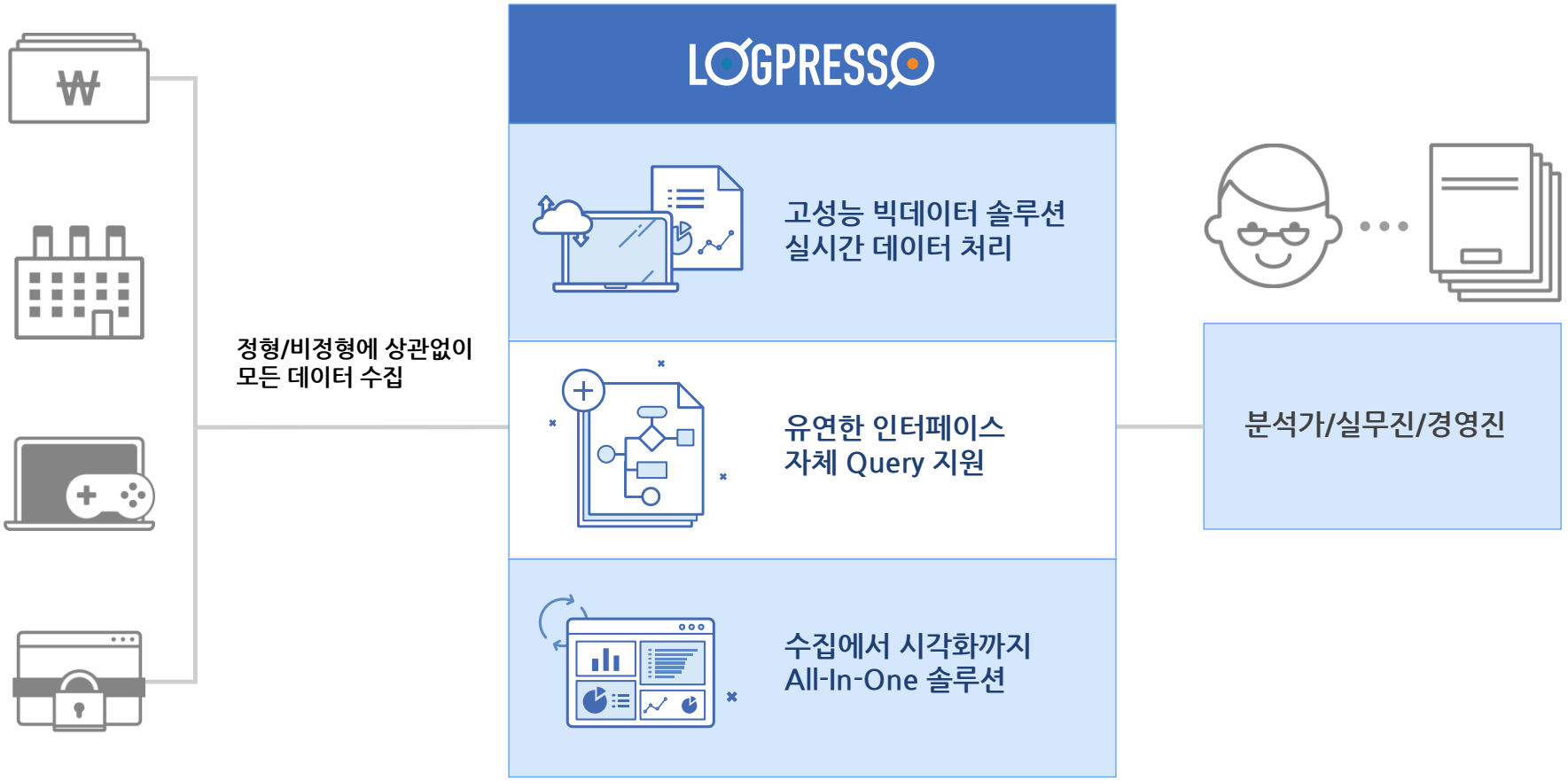
저장

분석

시각화

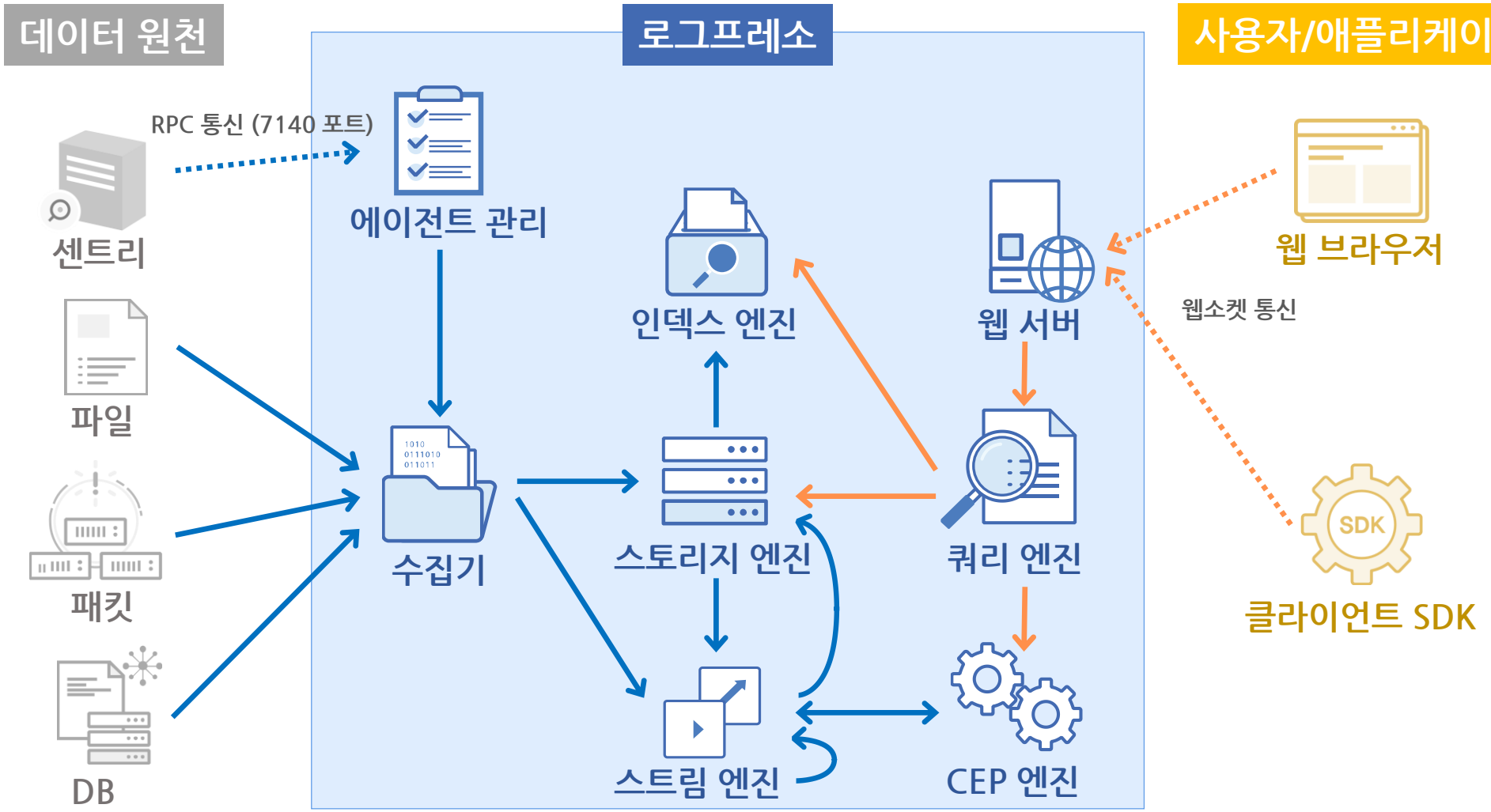
1-1. 로그프레소 개요

빅데이터를 수집·저장·분석·시각화하여 신속한 의사결정을 돕는 실시간 데이터 웨어하우스입니다.



1-2. 로그프레소 아키텍쳐

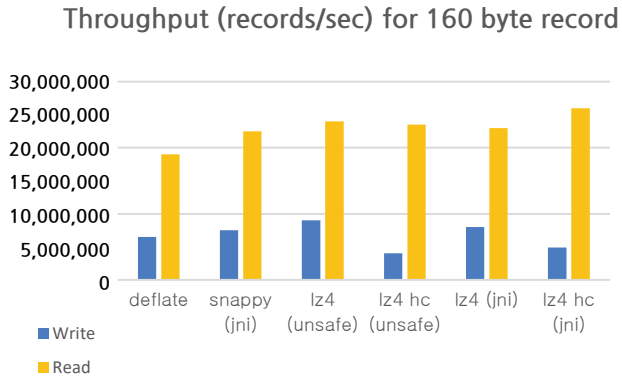
실시간 스트림과 배치 처리의 단순하고 효율적인 통합을 제공합니다.



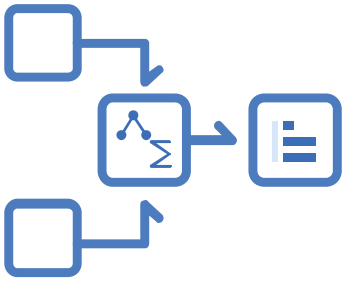
단일 프로세스에서 하나의 쿼리 문법으로 스트림 처리와 배치 처리를 모두 효율적으로 수행합니다.

1-3. 로그프레소 성능

로그프레소는 범용 하드웨어에서 뛰어난 성능을 보장합니다.



회	검색 대상	결과	소요시간
1	010****2419	204건	4.735초
2	010****1188	92건	0.742초
3	010****5561	119건	1.053초
4	010****2573	25건	0.356초



실시간 데이터 압축 저장

- 데이터 수정을 지원하지 않는 설계로 100만건/초 이상의 쓰기 성능을 제공합니다.
- 원본의 10% 수준으로 임시 파일 없이 실시간 압축 쓰기를 수행합니다.
- deflate, snappy, lz4, lz4 hc 4가지 압축 알고리즘을 제공합니다.
- 스키마가 없는 데이터 저장소로 언제든지 필드를 추가하거나 삭제할 수 있습니다.

실시간 풀 텍스트 인덱싱 + 고속 검색

- 단일 테이블에 30만건/초 이상 입력하는 인덱싱 성능
- 10억 건을 1초에 검색할 수 있는 검색 성능
  - 실제 SMS 백오피스 운영 서버 계측: 누적 600GB, 매일 30GB 적재
  - E5-2670 2.6GHz, RAM 96GB, SAS 7.2K RPM 하드웨어

실시간 스트림 쿼리

- 시스템 정지 시까지 무기한 동작하는 쿼리 (Continuous Query)
- 데이터 입력 후 쿼리 수행 완료까지 밀리초 단위의 레이턴시
  - FDS 응용에서 50개 이상의 패턴 및 프로파일 룰 대조 후 평균 0.1초 이내 탐지 응답

2. 실시간 빅데이터 웨어하우스

2-1. 수집

수집 저장 분석 시각화

2-1. 수집 : (1) 다양한 데이터 수집 방식 지원

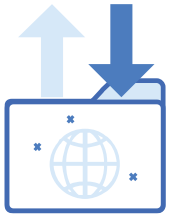
로그프레소를 설치하는 즉시 실시간 데이터 수집 및 데이터 통합을 시작할 수 있습니다.

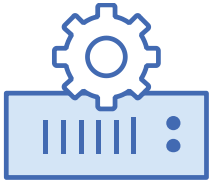

File


Syslog

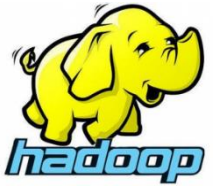

Datebase


twitter

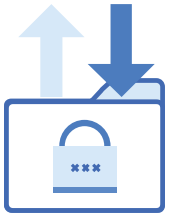

FTP


SNMP

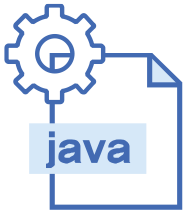

Performance



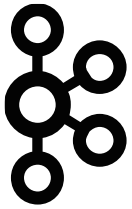
hadoop


SFTP


Netflow

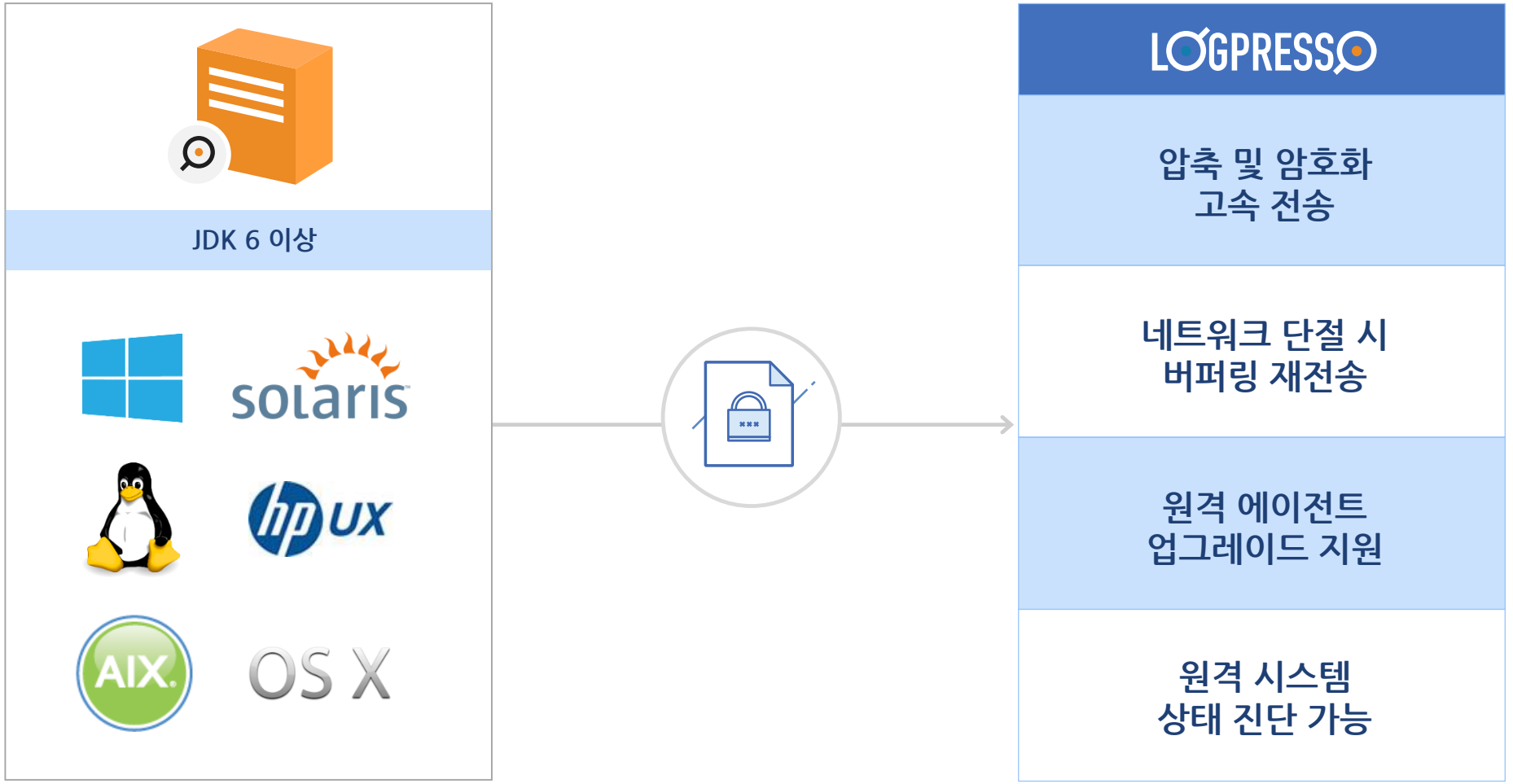


JMX


Kafka

2-1. 수집 : (2) 에이전트 지원

운영체제 별 특화된 데이터 수집을 지원하고, 압축/암호화하여 고속 데이터 전송을 수행합니다.



2. 실시간 빅데이터 웨어하우스

2-2. 저장

수집 저장 분석 시각화

2-2. 저장 : (1) 실시간 압축 저장

원본 데이터 포맷을 그대로 보존하면서 100만건/초 이상 실시간 압축 저장을 수행합니다. (웹 로그 기준)

원본 데이터에 대한 바이너리 인코딩 및 실시간 압축 쓰기 수행 (I/O 비용 감소)

원본 비정형 데이터

```
Oct 7 20:32:06 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 72.52.98.94#5939: query: logpresso.net IN
AAAA -
Oct 7 20:32:30 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 216.218.207.130#48281: query: logpresso.org
IN AAAA -
Oct 7 20:32:35 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 216.218.228.98#36493: query: logpresso.com
IN AAAA -
Oct 7 20:34:33 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 144.76.95.231#54247: query: araqne.org IN A -
Oct 7 20:35:44 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 168.126.63.60#57825: query:
wpad.hq.eediom.net IN A -E
Oct 7 20:36:11 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 168.126.63.59#45355: query: PC-
XERAPH.hq.eediom.net IN A -E
Oct 7 20:37:48 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 218.85.157.4#16590: query: newmets.net IN
AAAA -E
Oct 7 20:38:16 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 216.218.139.146#20958: query: newmets.net
IN AAAA -
Oct 7 20:40:33 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 218.30.103.29#43729: query: c828.com IN A -E
Oct 7 20:40:33 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 74.125.191.80#57392: query: ns1.8con.net IN
A -
Oct 7 20:40:33 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 74.125.191.84#51790: query: c828.com IN A -E
Oct 7 20:40:34 6ac2553f-fc02-44c8-baab-cc62ad8396d2
named[6187]: client 74.125.191.81#32794: query: c828.com IN A
```

데이터 타입 인코딩

코드	타입	설명
0	null	존재 안 함
1	bool	참/거짓
2	short	2byte 정수
3	int	4byte 정수
4	long	8byte 정수
5	string	문자열
6	date	타임스탬프
7	ipv4	IPv4 주소
8	ipv6	IPv6 주소
9	map	키/값 쌍
10	array	개체 목록
11	blob	바이너리
12	short (zigzag)	2byte 정수
13	int(zigzag)	4byte 정수
14	long(zigzag)	8byte 정수
15	float	단정도 실수
16	double	배정도 실수

실시간 블럭 압축된 원본 데이터 유지 (deflate, snappy, lz4, lz4hc 알고리즘 선택)

블럭헤더 압축된 원본 데이터

바이너리 코덱을 통하여 타입(T)-길이(L)-값(V) 형식으로 바이너리 인코딩된 다수의 레코드를 블럭 단위 압축 (ASN.1 DER과 유사함)

블럭헤더 압축된 원본 데이터

검색/애드혹 쿼리 모드: 낮은 압축률 대신 고속 검색 위주 설정
대용량 배치 분석 모드: 검색 성능이 떨어지는 대신 고효율 압축

블럭헤더 압축된 원본 데이터

원본 데이터 암호화 및 블럭 단위 무결성 검증 지원

블럭헤더 압축된 원본 데이터

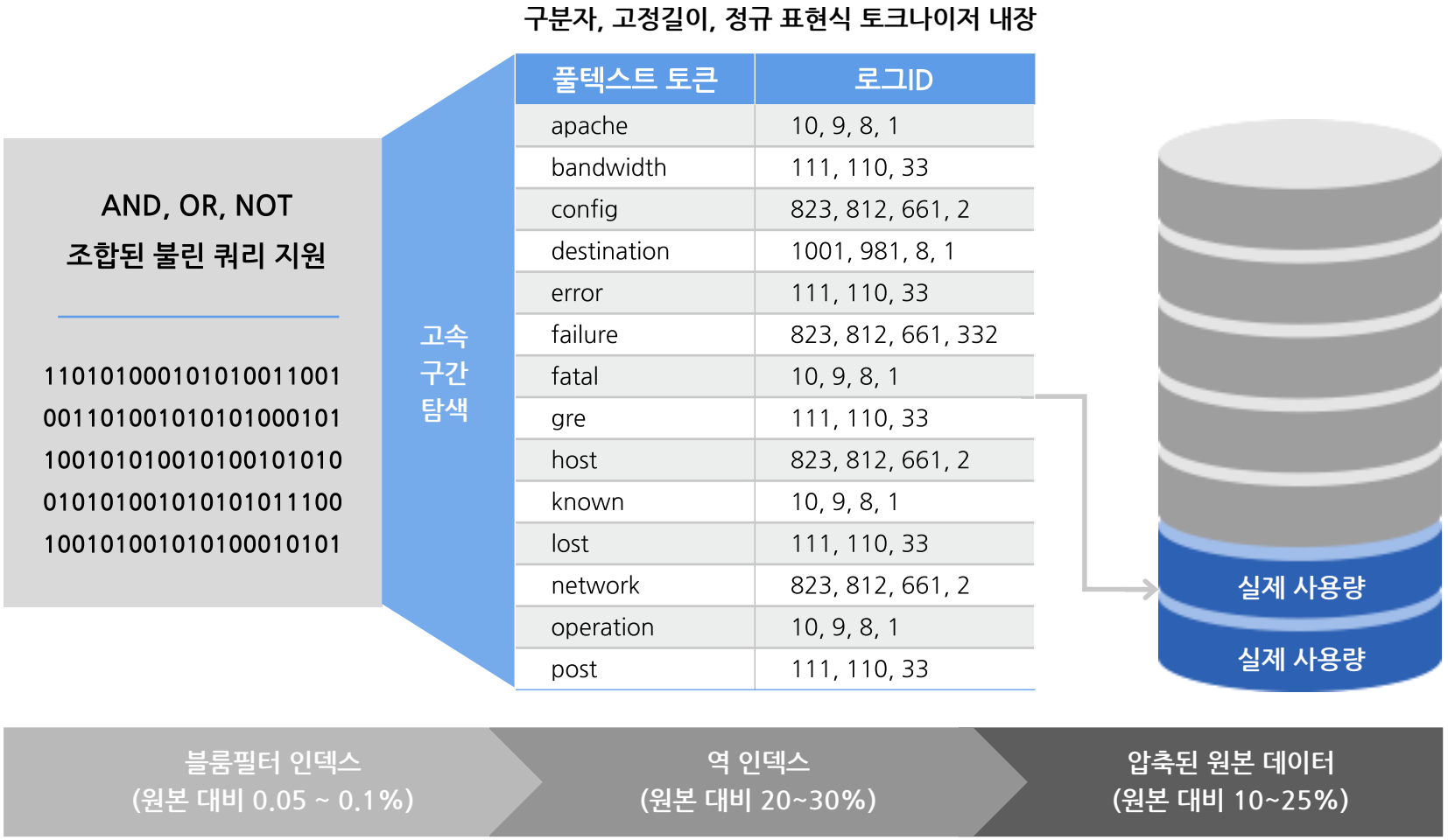
원본 데이터 암호화 및 블럭 단위 무결성 검증 지원

블럭헤더 압축된 원본 데이터

원본 데이터 암호화 및 블럭 단위 무결성 검증 지원

2-2. 저장 : (2) 실시간 풀텍스트 인덱싱

멀티코어를 극대화하는 병렬처리로 20만건/초 이상 인덱싱과 고속 풀텍스트 검색을 실현합니다.



2-2. 저장 : (3) 실시간 필드 인덱싱

원본 데이터를 그대로 유지하면서 필드와 타입을 구분하여 인덱스하고 검색할 수 있습니다.

원본 데이터

```
10.0.81.15 - - [01/May/1998:22:00:02 +0000] "GET /images/bg_generic_longjto.jpg HTTP/1.0" 200 21127
10.0.81.15 - - [01/May/1998:22:00:02 +0000] "GET /english/venues/images/venue_header.gif HTTP/1.0" 200 711
```

쿼리를 이용한 인덱스 토큰 추출

```
rex field=line "(?<ip>₩S+ )₩S+ ₩S+ ₩[(?<t>.*?)₩] "(?<method>₩S+) (?<path>₩S+) ₩S+" (?<status>₩S+) (?<bytes>₩S+)"
| eval ip = ip(ip)
| eval bytes = long(bytes)
| fields ip, method, path, status, bytes
```

필드 인덱스를 이용한 검색

```
fulltext ip == ipv4("10.0.135.39")
fulltext ip >= ipv4("10.0.135.0") and ip <= ipv4("10.0.135.255")
fulltext path == "sitemap"
fulltext status != "200" and method != "GET"
```

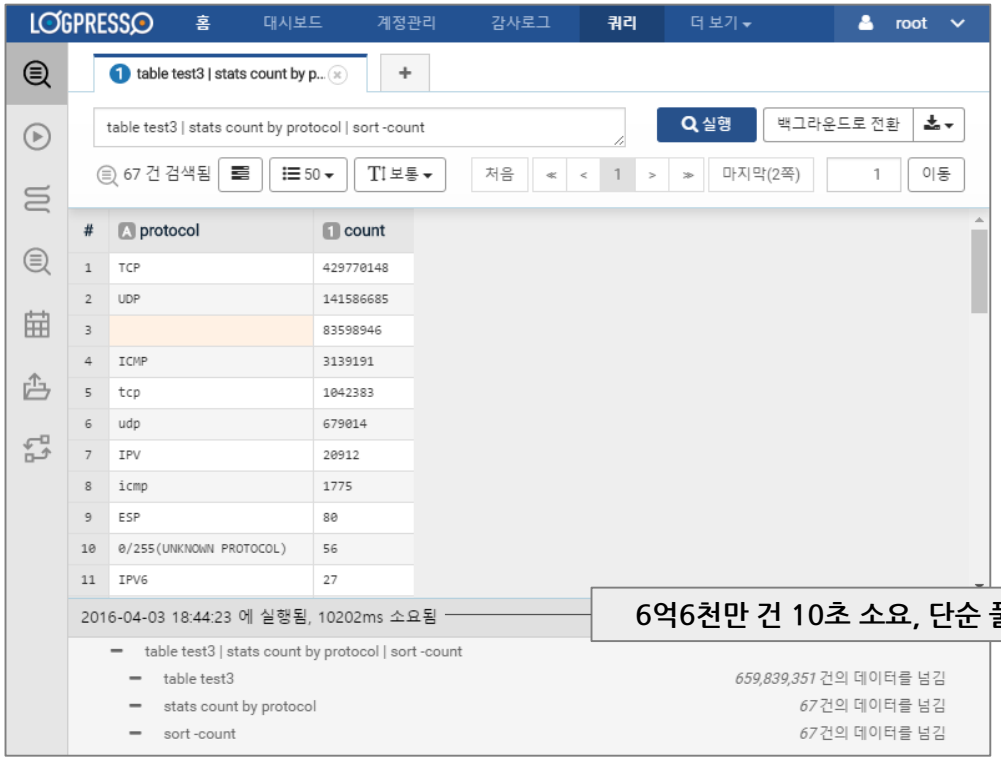
2-2. 저장 : (4) 컬럼 스토리지

테이블 설정만으로 데이터 스캔 성능을 극대화합니다.

- 쿼리에 반드시 필요한 컬럼 값만 추출하므로 컬럼 수가 많을수록 차이가 극대화됩니다.
- 추출 대상 데이터가 모두 인접하므로 CPU 캐시 활용을 극대화합니다.
- 벡터 타입을 자동 인식하여 디코딩을 가속화 합니다.

2-2. 저장 : (4) 컬럼 스토리지

기본 설정에 비해 최대 40배 이상의 성능을 나타냅니다.



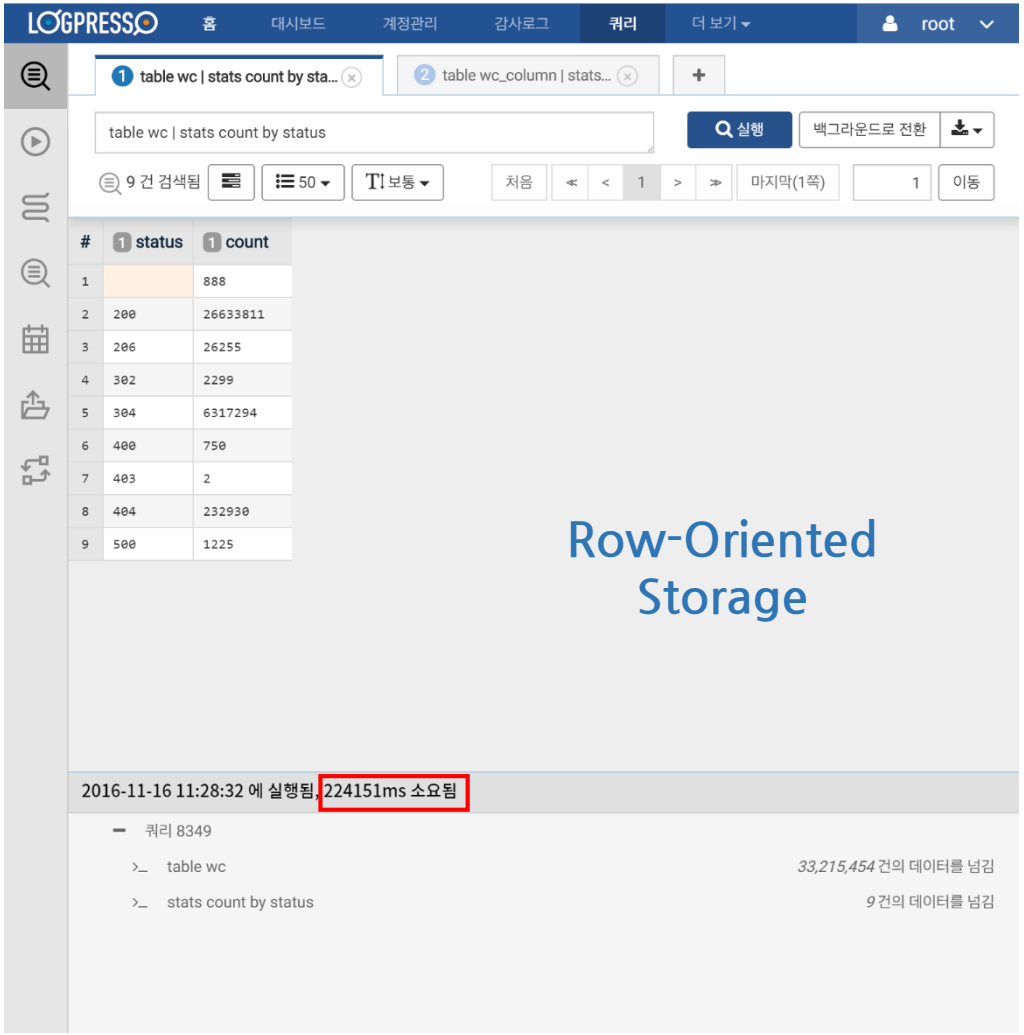
6억6천만 건 10초 소요, 단순 풀스캔 통계 6400만건/초

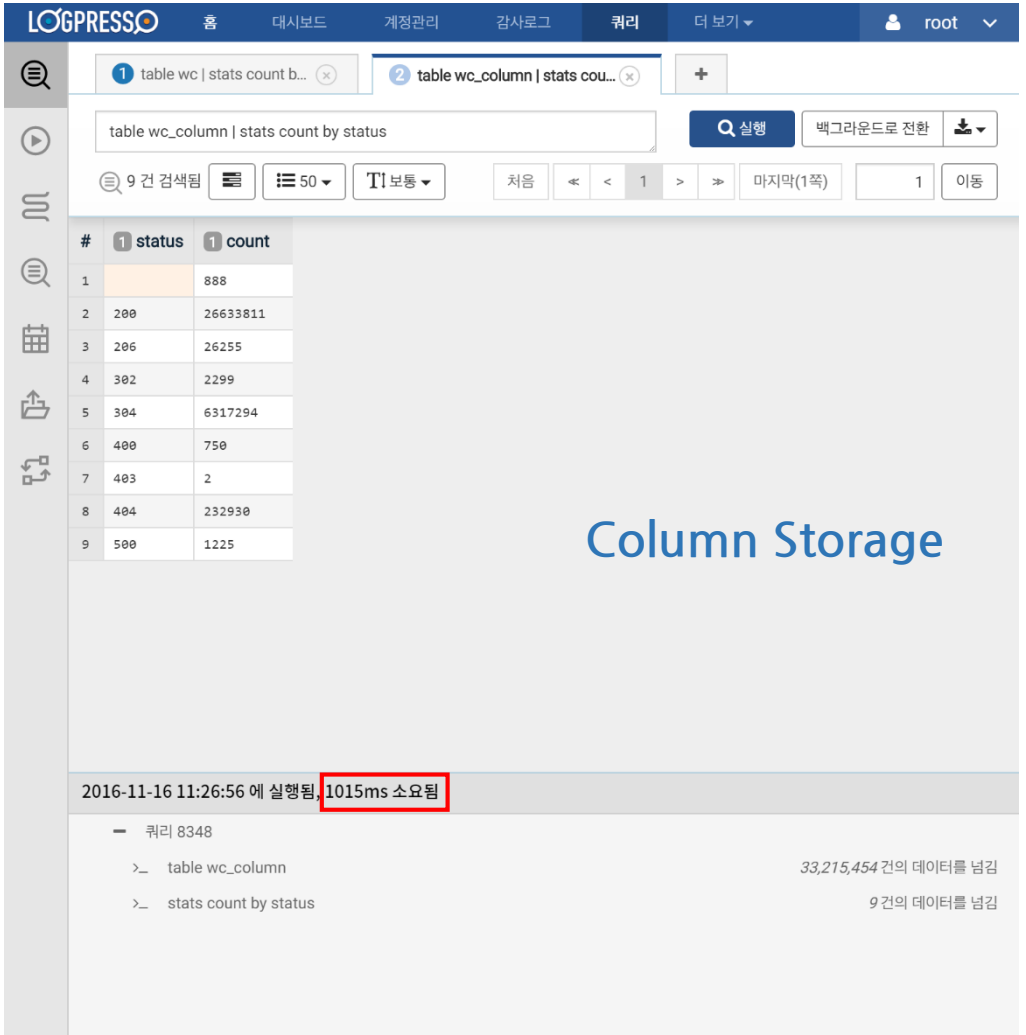
구 버전 대비
x20~40배 이상

구분	사양
운영체제	윈도우즈 10
CPU 및 메모리	인텔 코어 i7-4770 3.4GHz, RAM 32GB
OS 디스크	SanDisk 2.5" 128GB SATA III Internal SSD (SD6SB1M-128G-1022i)
데이터 디스크	Barracuda Desktop 6-Gb/s 3TB Hard Drive (ST3000DM001)

2-2. 저장 : (4) 컬럼 스토리지

기존 row 방식의 테이블과 columnar 방식의 테이블 성능 비교 (224초 vs 1초 : 224X 차이)





2-2. 저장 : (5) 실시간 데이터 암호화

수집된 원본 데이터의 전송부터 파일 저장까지 평문을 노출하지 않고 실시간 암호화를 수행합니다.

2-3. 분석

수집 저장 분석 시각화

2-3. 분석 : (1) 쿼리 문법 및 기능

전용 쿼리 문법으로 대용량, 비정형 데이터를 손쉽게 가공하고, 원격 데이터 소스를 통합하여 분석합니다.

Extract (추출) → Transform (변환) → Load (적재)

명령어	기능
table	로그프레소 테이블 스캔
fulltext	로그프레소 인덱스 검색
textfile	로컬 텍스트 파일 스캔
csvfile	로컬 CSV 파일 스캔
zipfile	로컬 ZIP 파일 스캔
ftp cat	원격 FTP 파일 스캔
sftp cat	원격 SFTP 파일 스캔
hdfs cat	원격 HDFS 파일 스캔
dbquery	JDBC 기반 SQL 쿼리
wget	HTTP 데이터 쿼리
stream	스트림의 출력 실시간 수신
logger	데이터 수집 시 실시간 수신

<다양한 데이터 원본 공급자>

- **100종 이상의 커맨드, 함수 및 그룹 함수**
- **사용자 정의 쿼리 문법 확장 지원**
- **사용자 정의 함수 (UDF) 확장 지원**

명령어	기능
parse	파서를 이용한 필드 추출
rex	정규표현식 기반 필드 추출
search	조건 표현식 필터링
sort	정렬 및 Top N
stats	그룹 함수 계산
timechart	시간대별 그룹 함수 계산
eval	표현식 평가 후 필드 할당
lookup	룩업 테이블 매핑 (geoip 등)
join	일반 데이터와 서브쿼리의 조인
streamjoin	스트림 데이터와 서브 쿼리의 조인
limit	쿼리 결과 페이징
fields	조회 필드 프로젝션
boxplot	박스플롯 값 계산
explode	배열을 다수의 튜플로 분해
evtctxadd	CEP 컨텍스트 생성
evtctxdel	CEP 컨텍스트 삭제

명령어	기능
import	로그프레소 테이블 적재
outputcsv	CSV 파일 출력
outputjson	JSON 파일 출력
outputtxt	텍스트 파일 출력
sftp put	원격 SFTP 파일 출력
hdfs put	원격 HDFS 파일 출력
dboutput	JDBC 기반 DBMS 적재
sendsyslog	Syslog 패킷 전송

데이터 마트 적재

hadoop ORACLE Microsoft SQL Server

BI 도구 시각화

tableau QlikView Spotfire

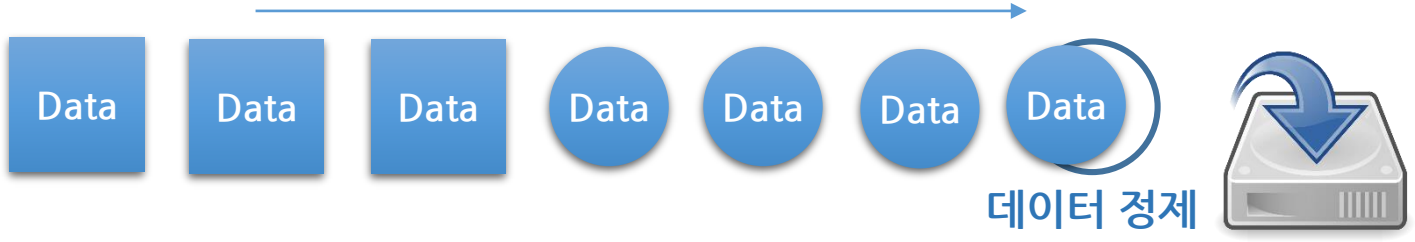
2-3. 분석 : (2) 비정형 데이터 파싱 - 60종 이상의 파서 내장

어떤 유형의 비정형 데이터라도 파싱하여 필드를 추출할 수 있습니다.

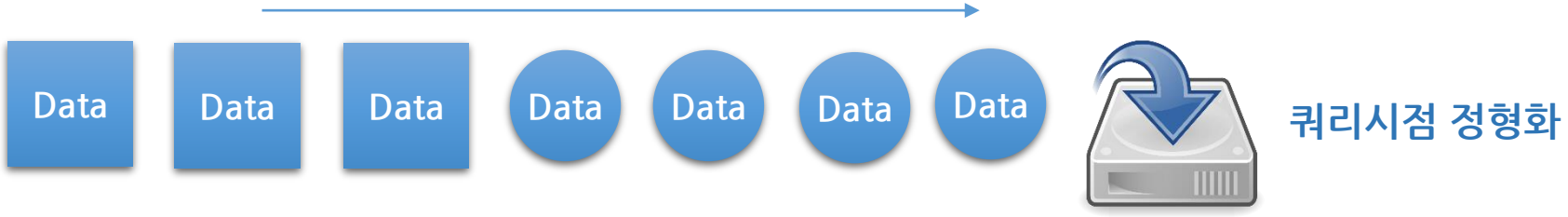
대분류	이름	설명
다용도 파서	구분자	지정된 구분자로 자르고 토큰 순서대로 설정된 필드 이름을 부여합니다.
	정규표현식	정규표현식을 사용하여 필드를 추출합니다.
	WELF	WELF(Web Trends Enhanced Log Format) 형식을 파싱하여 키/값 쌍을 추출합니다.
	필드 매핑	필드 값을 다른 필드에 할당합니다.
	태깅	특정 필드와 상수 쌍을 태깅합니다.
	체인	다수의 파서를 순차적으로 호출하여 복잡한 파싱을 수행합니다.
	쿼리	로그프레소 쿼리 문법과 함수를 이용하여 파싱을 수행합니다.
	그루비	내장된 그루비 스크립팅 엔진을 이용하여 사용자 스크립트로 파싱을 수행합니다.
특정 포맷 전용 파서 (이하 생략)	아파치 웹 로그	주어진 로그포맷 설정에 맞춰 아파치 웹 로그를 파싱합니다.
	넷스크린, SRX	주니퍼 넷스크린 방화벽, IPS 로그, SRX서비스 게이트웨이 로그를 파싱합니다.
	티핑포인트	HP 티핑포인트 IPS 로그를 파싱합니다.
	포티게이트	포티게이트 UTM 로그를 파싱합니다.
	팔로알토	팔로알토 네트웍스 PA 시리즈 UTM로그를 파싱합니다.
	디펜스프로	라드웨어 디펜스프로 Anti DDoS 로그를 파싱합니다.
	스나이퍼	윈스테크넷 스나이터 IPS 로그를 파싱합니다.
	MF2, NXG	시큐아이 MF2, NXG 방화벽 로그를 파싱합니다.
	트러스가드	안랩 트러스가드 UTM 로그를 파싱합니다.
	와플	와플 웹방화벽 로그를 파싱합니다.
	위가디아	퓨처시스템 위가디아 XTM, FM, IPS, SSLVPN 제품군 로그를 파싱합니다.

2-3. 분석 : (2) 비정형 데이터 파싱 - 60종 이상의 파서 내장

로그프레소는 선 수집, 쿼리시 파싱으로 데이터 유실 방지



데이터를 정형화 한 후 저장하는 경우 **데이터의 형상이 변경되면 이를 적용할 때까지 수집불가**



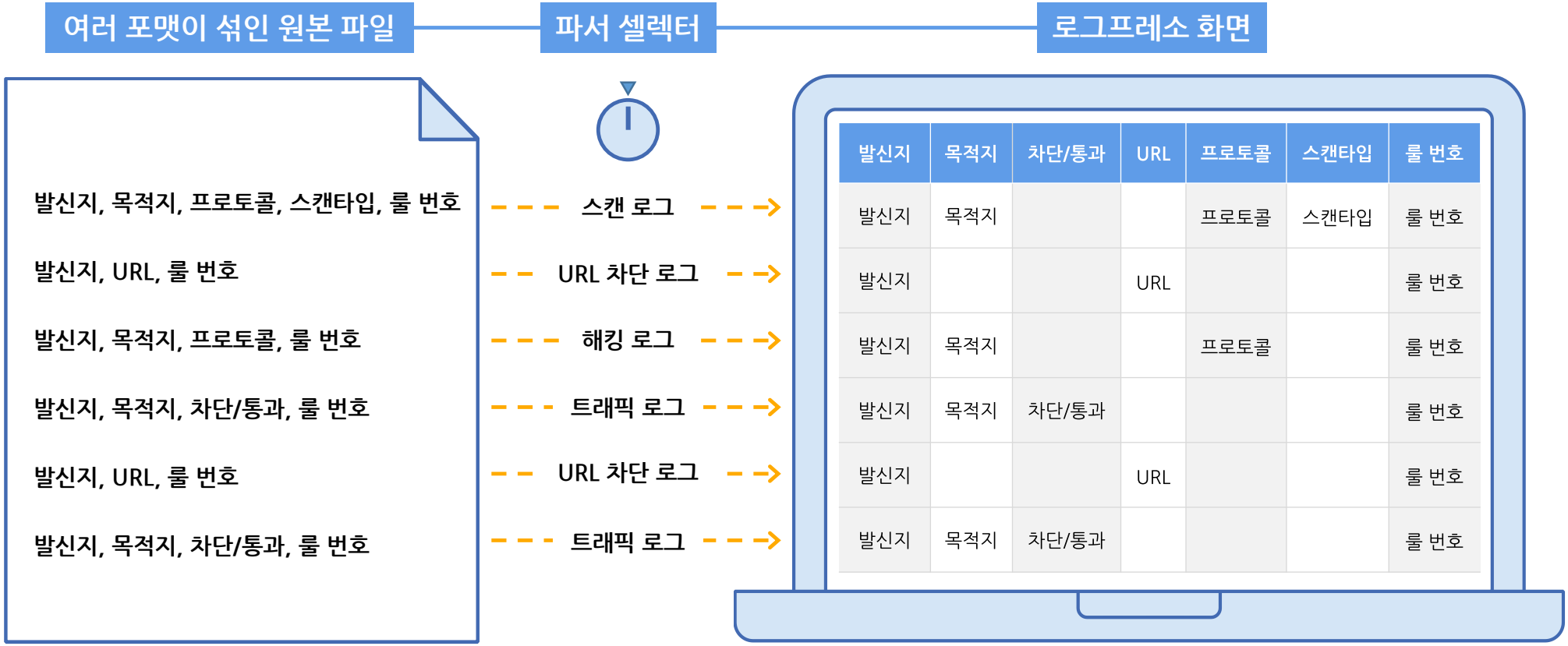
원본 형태로 수집하고 쿼리 시점에 정형화하는 경우는 유실없이 데이터 수집가능
사후 파서를 보완하면 완벽하게 정형화

2-3. 분석 : (3) 비정형 데이터 파싱 - 복합된 비정형 데이터 필드 추출

UTM 로그처럼 다수의 포맷이 섞여있는 경우에도 단순 정규식 추출 이상의 파싱을 수행할 수 있습니다.

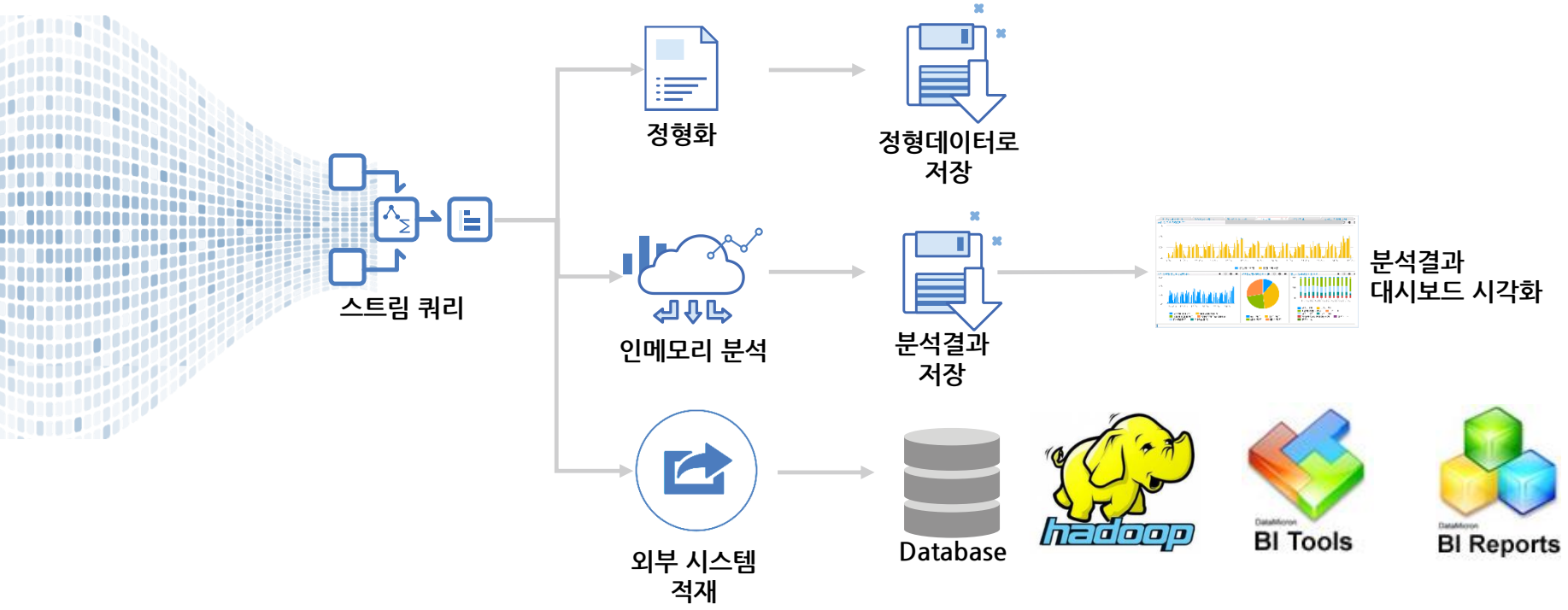
고성능(+)유연성을 만족시키는 파서 구성 지원

- 쿼리를 이용한 파싱
- 그루비 스크립트를 이용한 파싱
- 자바스크립트를 이용한 파싱



2-3. 분석 : (4) 스트림 분석 - 실시간 분석 및 DB 적재

데이터 수집 단계에서 정규화 및 인메모리 분석 및 외부 연동의 실시간 적용

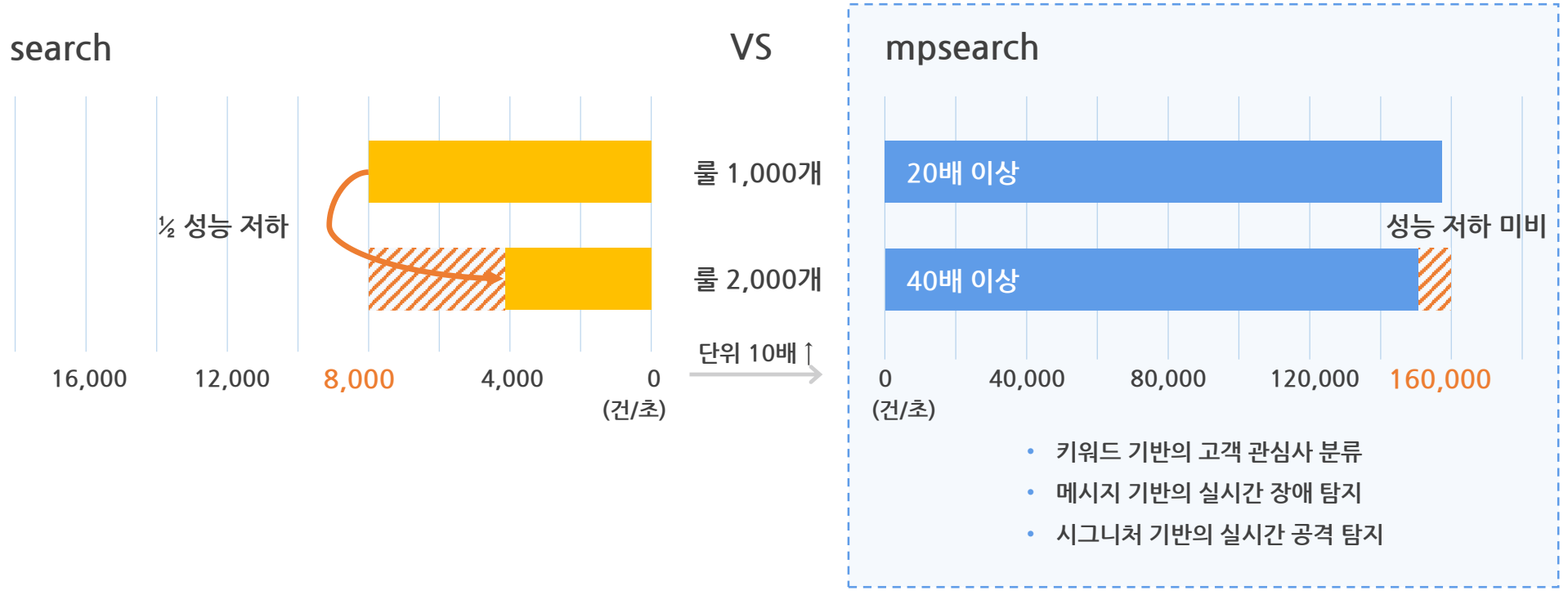


실시간 스트림 분석

- 데이터 수집과 동시에 파서를 적용, 데이터를 정규화, 필터링 및 분기 로직을 통과하여 저장
- DBMS, Hadoop 등의 시스템에 데이터를 수집과 동시에 변형을 가하여 연동 저장 가능(실시간 ETL)
- 원본 데이터 전체를 저장하지 않고도 스트림 쿼리를 통한 통계 분석 및 외부 데이터 적재 가능

2-3. 분석 : (4) 스트림 분석 - 실시간 멀티패턴 검색

입력 스트림에 수천 개의 패턴을 적용해도 성능 저하 없이 실시간으로 탐지합니다.

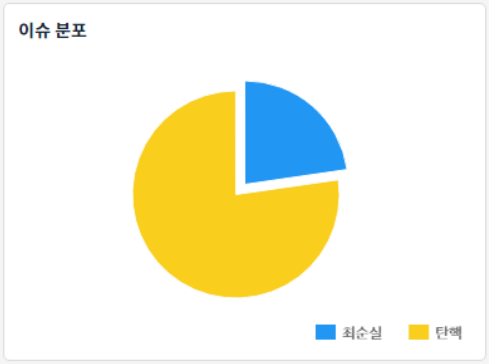


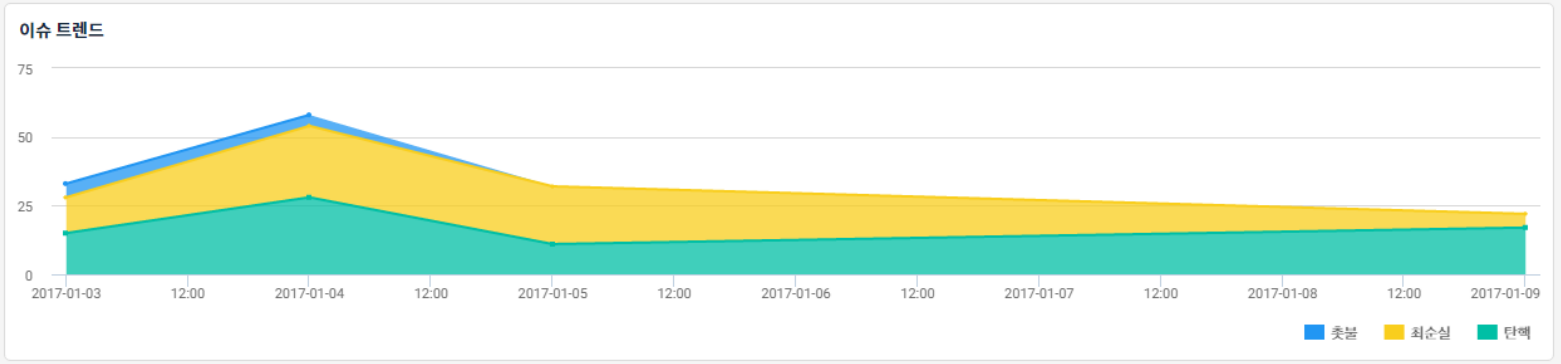
항목	모델 및 사양
CPU	1.7 GHz Intel Core i5
메모리	4 GB 1600 MHz DDR3
운영체제	OS X Yosemite Version 10.10.4

2-3. 분석 : (4) 스트림 분석 - 실시간 멀티패턴 검색

입력 스트림에 수천 개의 패턴을 적용해도 성능 저하 없이 실시간으로 탐지합니다.

RSS수집기와 멀티패턴 쿼리로 만든 뉴스 데이터 추출 대시보드의 예



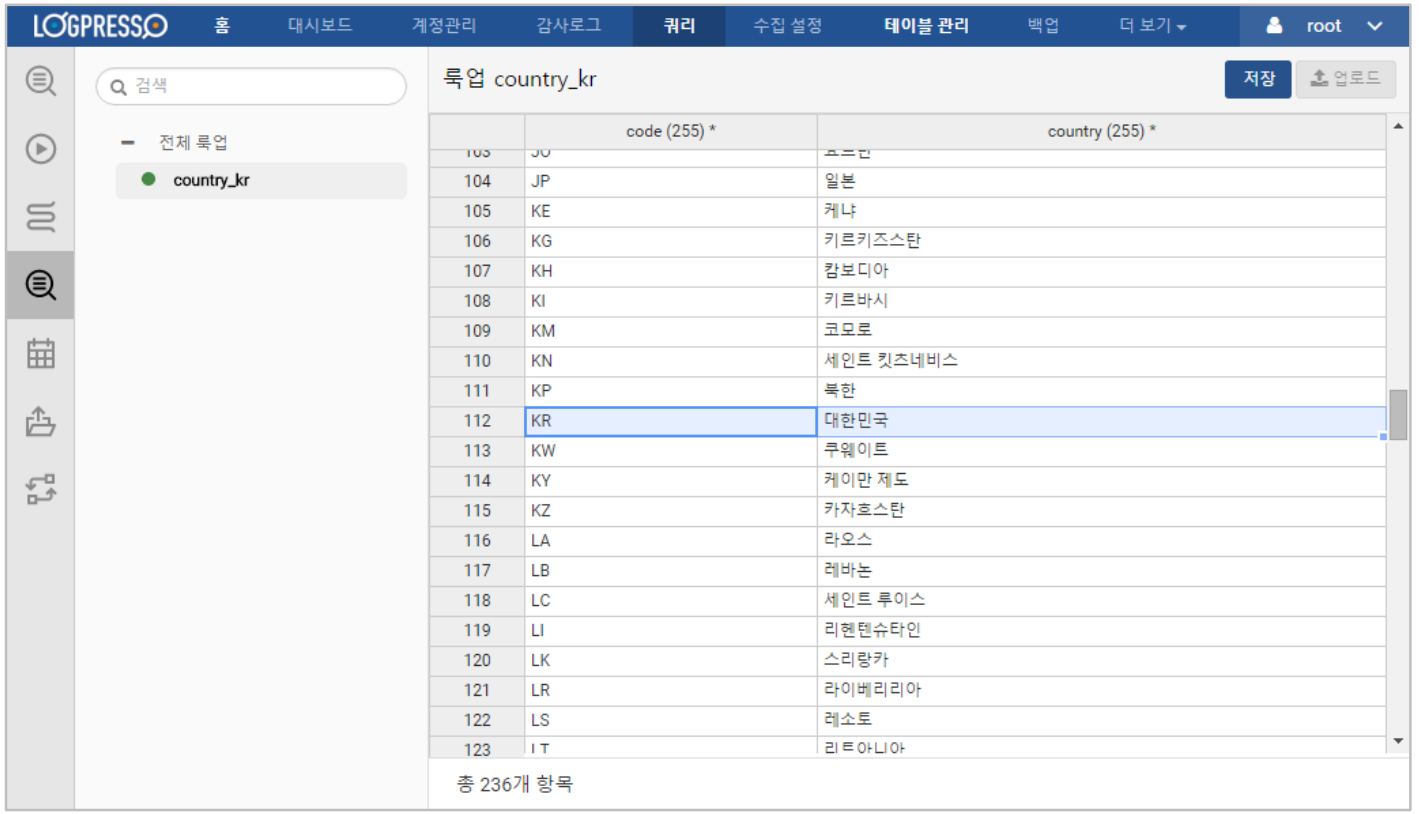


JTBC 뉴스

rule	구분	title	link	content
탄핵	속보	[사회현장] 마지막 청문회 '맹탕' 우려 등	http://news.jtbc.join	[앵커]오늘(9일) 뉴스 현장은 국정조사 7차 청문회 현장과 함께 하겠습니다. 여상원 변호사, 노영희 변호사, 그리고 강찬호 중앙일보 논설위원과 함께 합니다. 세 분 어서오십시오. 오전 청문회는 끝났고, 잠시 후
탄핵	속보	[속보] 특검 "김상률·김종덕·정관주·신동철 오늘 영장 가능성"	http://news.jtbc.join	특검 "김상률·김종덕·정관주·신동철 오늘 영장 가능성"(JTBC 방송뉴스팀)
탄핵	속보	[영상구성] 국조특위 '7차 최순실 청문회'	http://news.jtbc.join	국조특위 '7차 최순실 청문회'남궁곤, 정동춘 증인 출석노승일, 참고인 신분으로 출석하지만, 핵심 증인 대부분 불출석![정제원 의원/바른정당 : 오늘 텅빈 증인석을 바라보니까 참 청문위원로서 자괴감과 무력감과
최순실	속보	[영상구성] 국조특위 '7차 최순실 청문회'	http://news.jtbc.join	국조특위 '7차 최순실 청문회'남궁곤, 정동춘 증인 출석노승일, 참고인 신분으로 출석하지만, 핵심 증인 대부분 불출석![정제원 의원/바른정당 : 오늘 텅빈 증인석을 바라보니까 참 청문위원로서 자괴감과 무력감과
탄핵	속보	7차 청문회, 텅 빈 증인석…조윤선 오후 출석에 주목	http://news.jtbc.join	[앵커]특검이 정신없이 비쁜 모양인데요, 이번에는 국회 취재기자 연결합니다. 박사라 기자! 오늘(9일) 최순실 국정농단 사건에 대한 사실상 마지막 청문회가 열렸는데, 증인들이 대부분 안나오지 않았습니까?[기자
탄핵	속보	특검, 삼성 뇌물죄 수사 정점…최순실 '강제수사' 검토	http://news.jtbc.join	[앵커]특검이 삼성그룹의 뇌물죄 수사와 관련해 오늘(9일) 오전 최지성 부회장과 장충기 사장을 소환했습니다. 현장에 나가있는 취재기자 연결해 자세한 소식 알아봅니다.최규진 기자, 특검팀이 처음으로 삼성그룹
최순실	속보	특검, 삼성 뇌물죄 수사 정점…최순실 '강제수사' 검토	http://news.jtbc.join	[앵커]특검이 삼성그룹의 뇌물죄 수사와 관련해 오늘(9일) 오전 최지성 부회장과 장충기 사장을 소환했습니다. 현장에 나가있는 취재기자 연결해 자세한 소식 알아봅니다.최규진 기자, 특검팀이 처음으로 삼성그룹
탄핵	속보	[최순실 국정농단 국정조사 청문회] 1월 9일 JTBC 뉴스특보	http://news.jtbc.join	[앵커]최순실 국정농단 사건에 대한 국회 국정조사 특위, 사실상 마지막 청문회입니다. 이제 곧 진행될 7차 청문회, 지금부터 특보로 전해드리겠습니다. 그런데 오늘(9일) 청문회 마지막 청문회인데 참 맥이 빠지게
최순실	속보	[최순실 국정농단 국정조사 청문회] 1월 9일 JTBC 뉴스특보	http://news.jtbc.join	[앵커]최순실 국정농단 사건에 대한 국회 국정조사 특위, 사실상 마지막 청문회입니다. 이제 곧 진행될 7차 청문회, 지금부터 특보로 전해드리겠습니다. 그런데 오늘(9일) 청문회 마지막 청문회인데 참 맥이 빠지게
탄핵	속보	특검, 뇌물죄 수사망 확대…이재용 부회장 소환 임박	http://news.jtbc.join	[앵커]중요한 수사 중 한 축이 삼성의 뇌물죄인데 오늘(9일) 그룹의 2인자가 검찰에 나오네요?Q. 최지성·장충기 소환…삼성 수뇌부 정조준[손정혜/변호사 : 특검, 삼성미래전략실 지원 경위 따질 것.]Q. 이재용 소
탄핵	속보	[속보] 삼성 장충기 특검 출석… 박 대통령-이재용 의혹 추궁	http://news.jtbc.join	삼성 장충기 특검 출석… 박 대통령-이재용 의혹 추궁(JTBC 방송뉴스팀)
탄핵	정치	[속보] 특검 "김상률·김종덕·정관주·신동철 오늘 영장 가능성"	http://news.jtbc.join	특검 "김상률·김종덕·정관주·신동철 오늘 영장 가능성"(JTBC 방송뉴스팀)
탄핵	정치	[영상구성] 국조특위 '7차 최순실 청문회'	http://news.jtbc.join	국조특위 '7차 최순실 청문회'남궁곤, 정동춘 증인 출석노승일, 참고인 신분으로 출석하지만, 핵심 증인 대부분 불출석![정제원 의원/바른정당 : 오늘 텅빈 증인석을 바라보니까 참 청문위원로서 자괴감과 무력감과
최순실	정치	[영상구성] 국조특위 '7차 최순실 청문회'	http://news.jtbc.join	국조특위 '7차 최순실 청문회'남궁곤, 정동춘 증인 출석노승일, 참고인 신분으로 출석하지만, 핵심 증인 대부분 불출석![정제원 의원/바른정당 : 오늘 텅빈 증인석을 바라보니까 참 청문위원로서 자괴감과 무력감과
탄핵	정치	7차 청문회, 텅 빈 증인석…조윤선 오후 출석에 주목	http://news.jtbc.join	[앵커]특검이 정신없이 비쁜 모양인데요, 이번에는 국회 취재기자 연결합니다. 박사라 기자! 오늘(9일) 최순실 국정농단 사건에 대한 사실상 마지막 청문회가 열렸는데, 증인들이 대부분 안나오지 않았습니까?[기자
탄핵	정치	[최순실 국정농단 국정조사 청문회] 1월 9일 JTBC 뉴스특보	http://news.jtbc.join	[앵커]최순실 국정농단 사건에 대한 국회 국정조사 특위, 사실상 마지막 청문회입니다. 이제 곧 진행될 7차 청문회, 지금부터 특보로 전해드리겠습니다. 그런데 오늘(9일) 청문회 마지막 청문회인데 참 맥이 빠지게
최순실	정치	[최순실 국정농단 국정조사 청문회] 1월 9일 JTBC 뉴스특보	http://news.jtbc.join	[앵커]최순실 국정농단 사건에 대한 국회 국정조사 특위, 사실상 마지막 청문회입니다. 이제 곧 진행될 7차 청문회, 지금부터 특보로 전해드리겠습니다. 그런데 오늘(9일) 청문회 마지막 청문회인데 참 맥이 빠지게

2-3. 분석 : (5) 데이터 매핑 - 룩업

인사 DB, 자산 DB, 블랙리스트, 위협 인텔리전스 등 다양한 메타데이터를 편집하고 분석에 적용합니다.



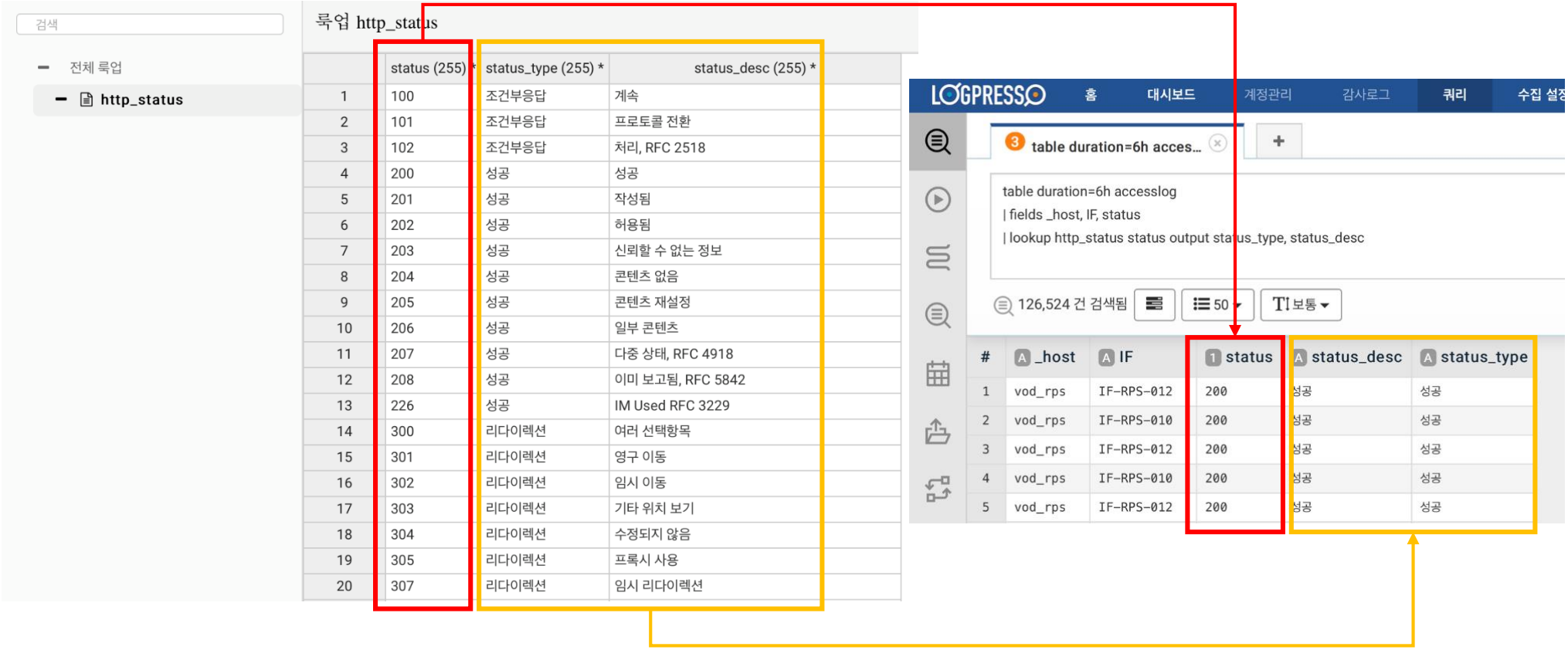
로컬 파일 룩업

외부 DB 룩업

2-3. 분석 : (5) 데이터 매핑 - 룩업

인사 DB, 자산 DB, 블랙리스트, 위협 인텔리전스 등 다양한 메타데이터를 편집하고 분석에 적용합니다.

룩업 화면과 룩업이 적용된 쿼리 결과 화면

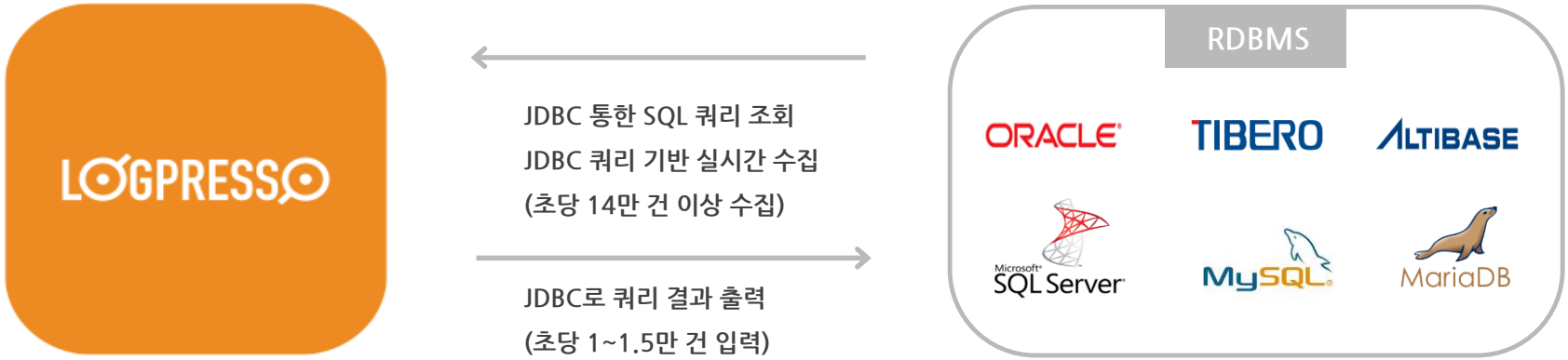


	status (255) *	status_type (255) *	status_desc (255) *
1	100	조건부응답	계속
2	101	조건부응답	프로토콜 전환
3	102	조건부응답	처리, RFC 2518
4	200	성공	성공
5	201	성공	작성됨
6	202	성공	허용됨
7	203	성공	신뢰할 수 없는 정보
8	204	성공	콘텐츠 없음
9	205	성공	콘텐츠 재설정
10	206	성공	일부 콘텐츠
11	207	성공	다중 상태, RFC 4918
12	208	성공	이미 보고됨, RFC 5842
13	226	성공	IM Used RFC 3229
14	300	리다이렉션	여러 선택항목
15	301	리다이렉션	영구 이동
16	302	리다이렉션	임시 이동
17	303	리다이렉션	기타 위치 보기
18	304	리다이렉션	수정되지 않음
19	305	리다이렉션	프록시 사용
20	307	리다이렉션	임시 리다이렉션

#	_host	IF	status	status_desc	status_type
1	vod_rps	IF-RPS-012	200	성공	성공
2	vod_rps	IF-RPS-010	200	성공	성공
3	vod_rps	IF-RPS-012	200	성공	성공
4	vod_rps	IF-RPS-010	200	성공	성공
5	vod_rps	IF-RPS-012	200	성공	성공

2-3. 분석 : (6) 데이터 통합 - RDBMS

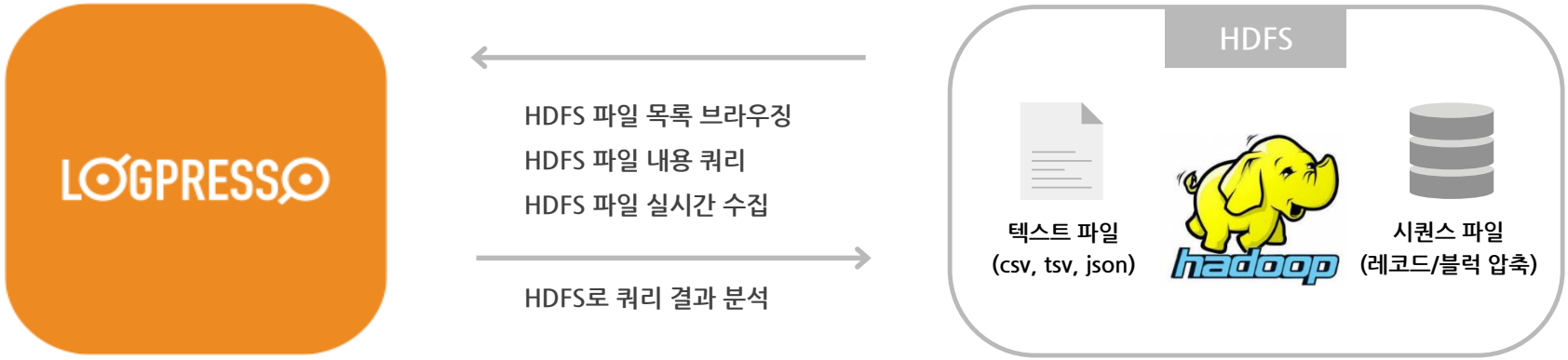
각종 상용 RDBMS와 연동하여, RDBMS에 저장한 데이터를 다양하게 분석할 수 있습니다.



시나리오	설명
로그프레소 → RDBMS 데이터 흐름	로그프레소가 ETL의 역할로 실시간 데이터 수집, 시각화, 데이터 분석을 수행하고 파싱, 정규화, 가공된 결과를 SQL 데이터베이스에 적재
RDBMS → 로그프레소 데이터 흐름	데이터베이스를 쿼리하여 로그프레소가 추가적인 분석 및 시각화 수행
로그프레소 → RDBMS 양방향 데이터 흐름	로그프레소와 SQL 데이터베이스가 각기 다른 데이터를 수집 및 처리하고, 양쪽 데이터를 **조인**하여 추가적인 정보 획득

2-3. 분석 : (6) 데이터 통합 - 하둡 (HDFS)

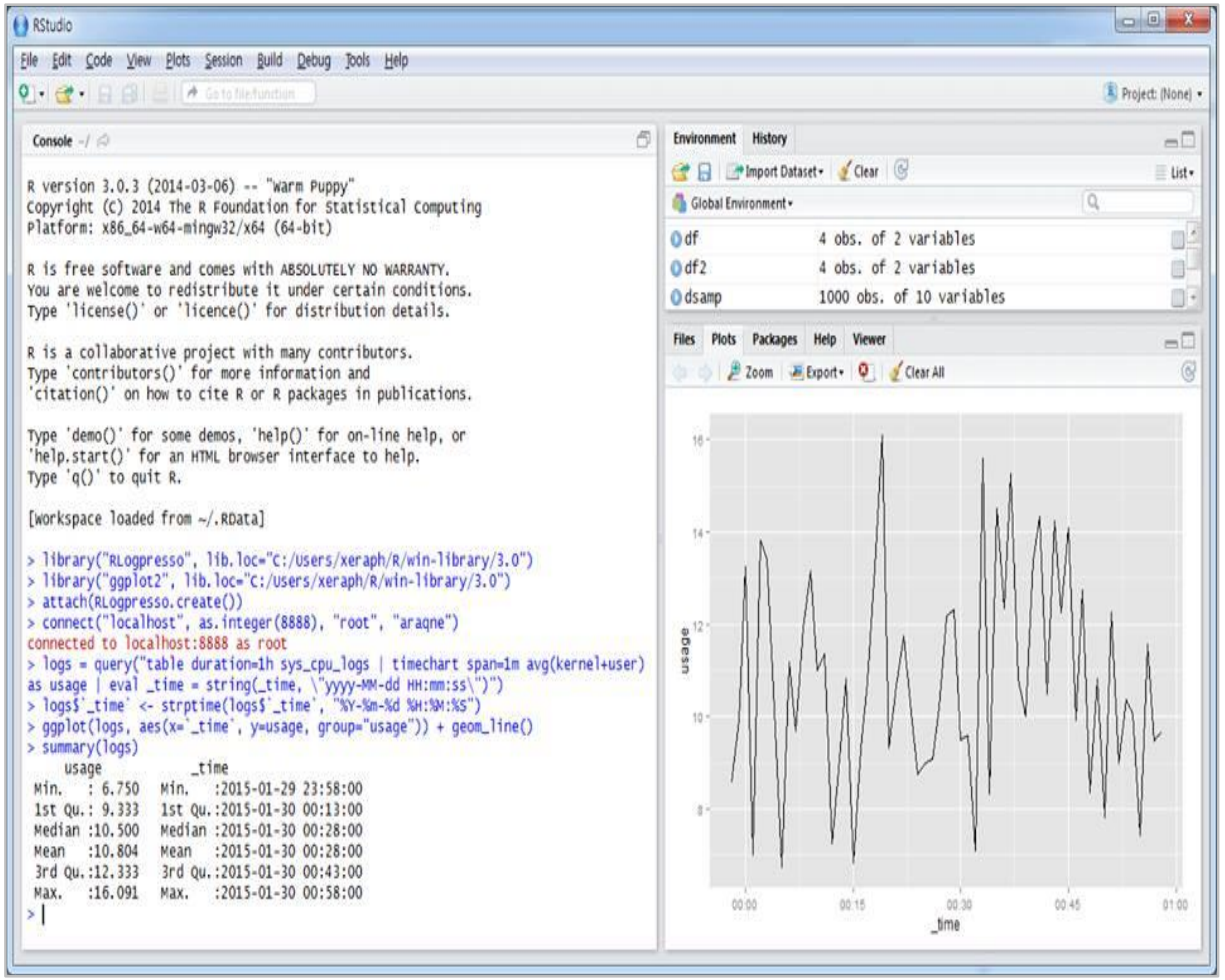
하둡(HDFS)과 완벽하게 연동해 사용할 수 있어, 하둡으로 수집한 데이터를 로그프레소를 이용해 빠른 분석이 가능합니다.



시나리오	설명
로그프레소 → 하둡 데이터 흐름	로그프레소가 ETL의 역할로 실시간 데이터 수집, 시각화, 데이터 분석을 수행하고 그 결과를 하둡으로 전달, 하둡이 DW로서 배치 처리 수행
하둡 → 로그프레소 데이터 흐름	하둡이 대용량 데이터 수집 및 처리를 수행하고, HDFS에 적재된 파일을 로그프레소가 추가적인 분석 및 시각화 수행
로그프레소 → 하둡 양방향 데이터 흐름	로그프레소와 하둡이 각기 다른 데이터를 수집 및 처리하고, 양쪽 데이터를 **조인**하여 추가적인 정보 획득

2-3. 분석 : (7) 고급 분석 - R 연동 지원

Rlogpresso 라이브러리를 이용하여 R에서 로그프레소 쿼리를 실행할 수 있습니다.



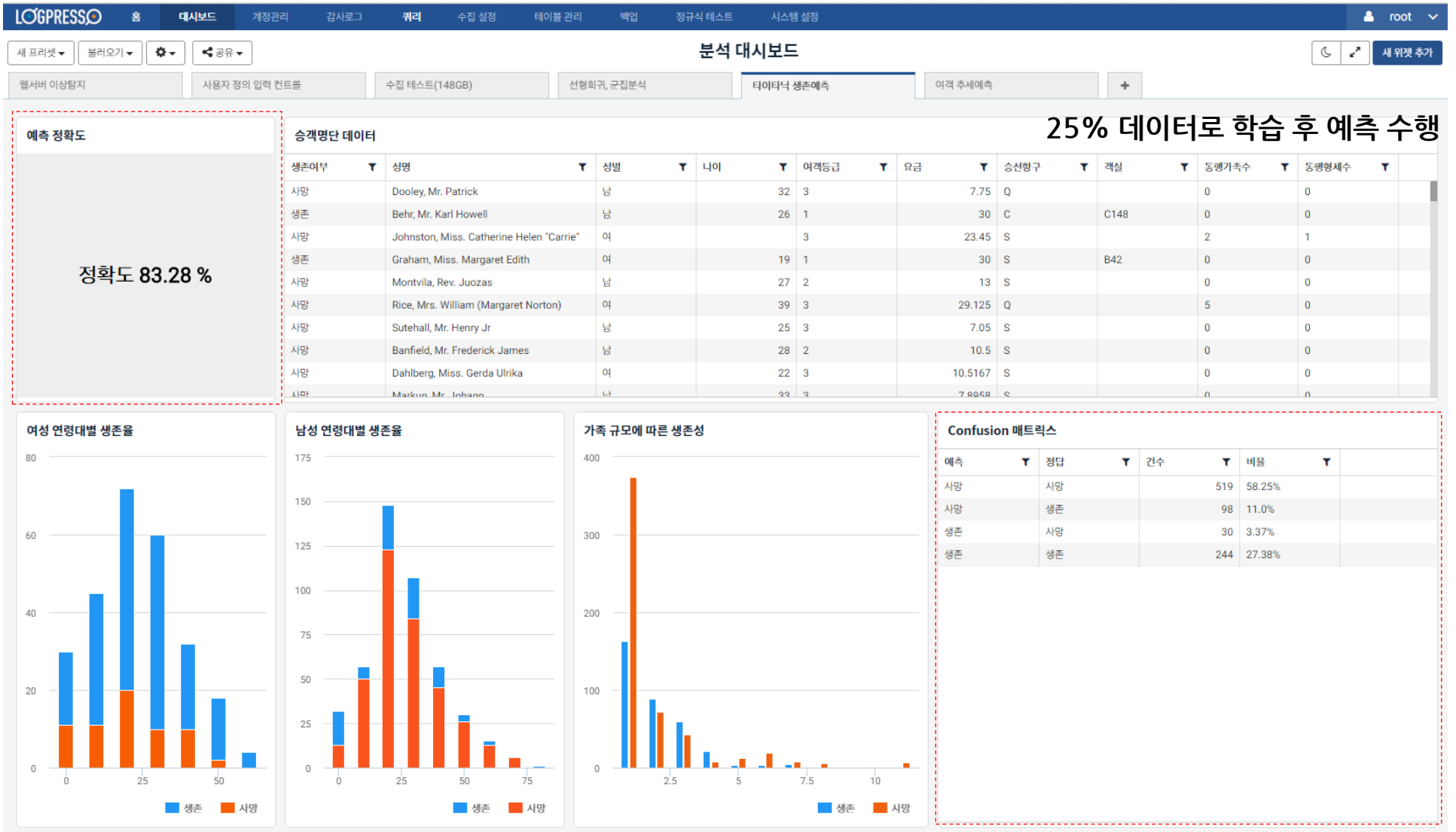
인메모리 적재
고급 통계 분석

LOGPRESSO

실시간 빅데이터
수집 및 1차 통계 가공

2-3. 분석 : (7) 머신러닝 분석

Supervised Machine Learning 기법을 통한 예측 분석

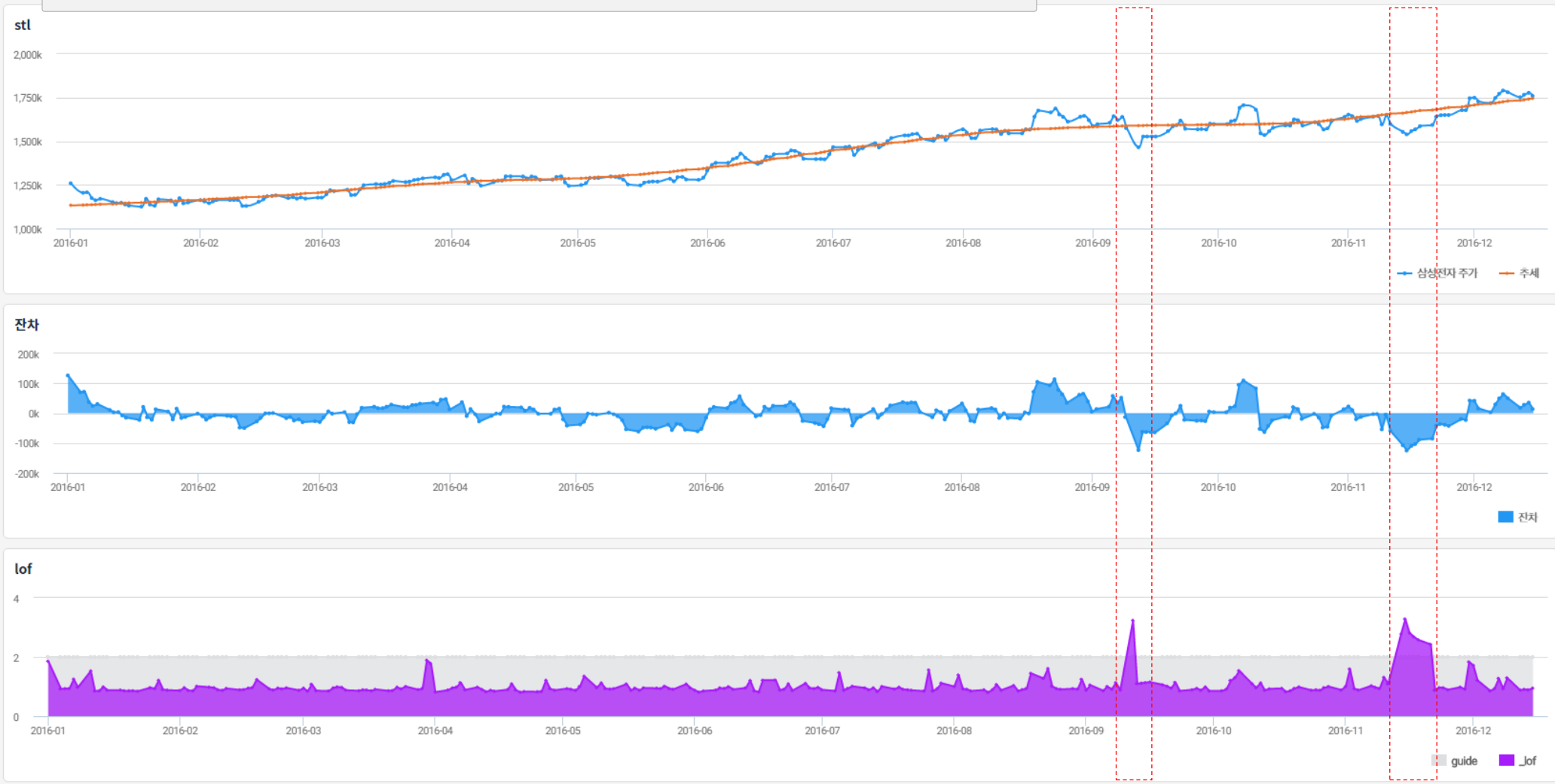


2-3. 분석 : (7) 머신러닝 분석

Unsupervised Machine Learning 기법을 통한 Anormaly Detection

```
table finance | fields Adj_Close, _time | stl Adj_Close | lof _error
```

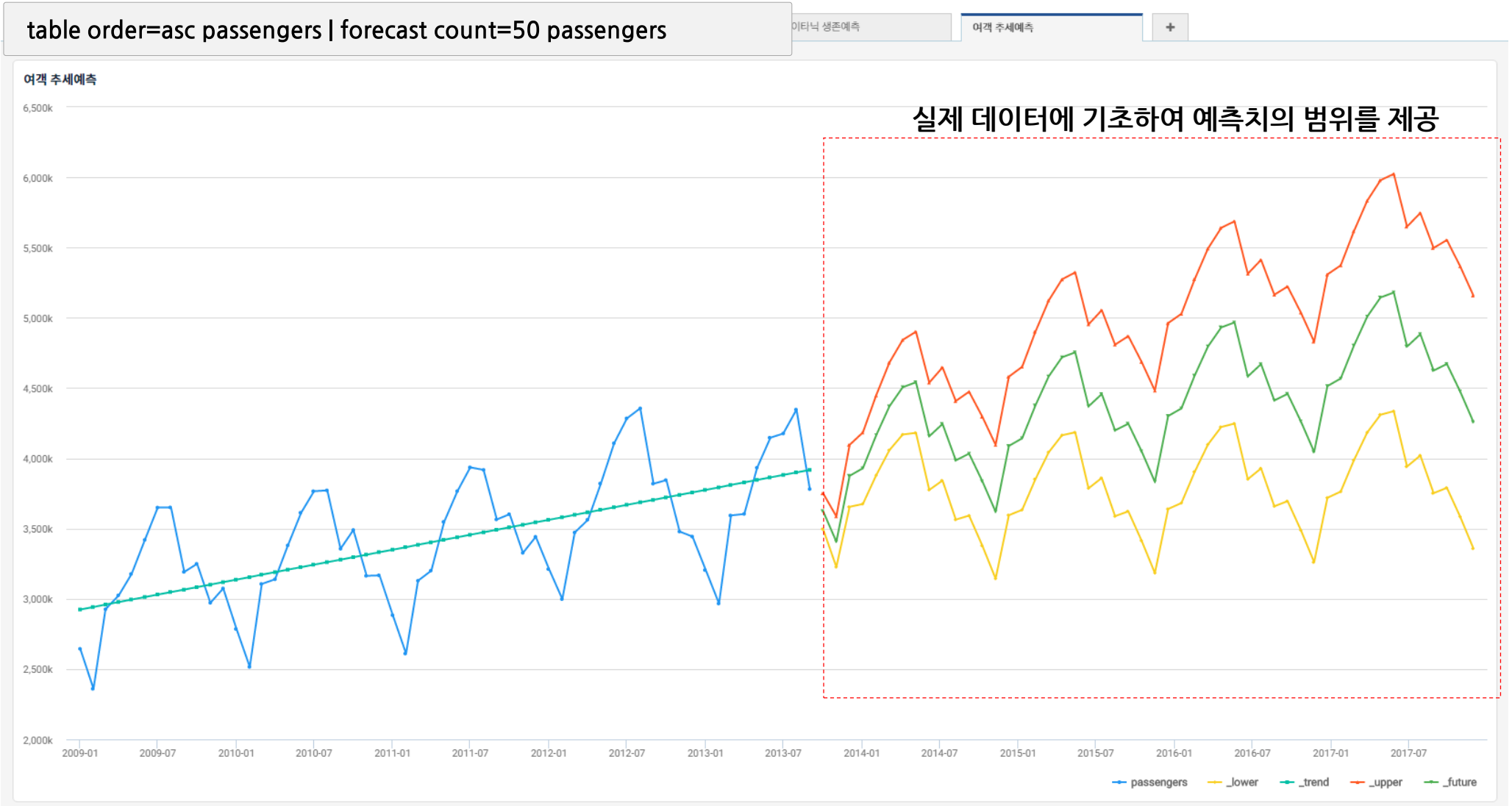
주가 하락의 이상치



2-3. 분석 : (7) 머신러닝 분석

Machine Learning Regression을 통한 추세 예측

```
table order=asc passengers | forecast count=50 passengers
```



2-3. 분석 : (8) 분산 쿼리 옵티마이저

데이터 로컬리티를 극대화하는 분산 쿼리 플랜을 작성합니다.

분산 어낼리틱 쿼리 플랜

클라이언트 쿼리 : 출발지 IP 통계

table n1:t1, n2:t2, n3:t3 | rex | stats count by src ip

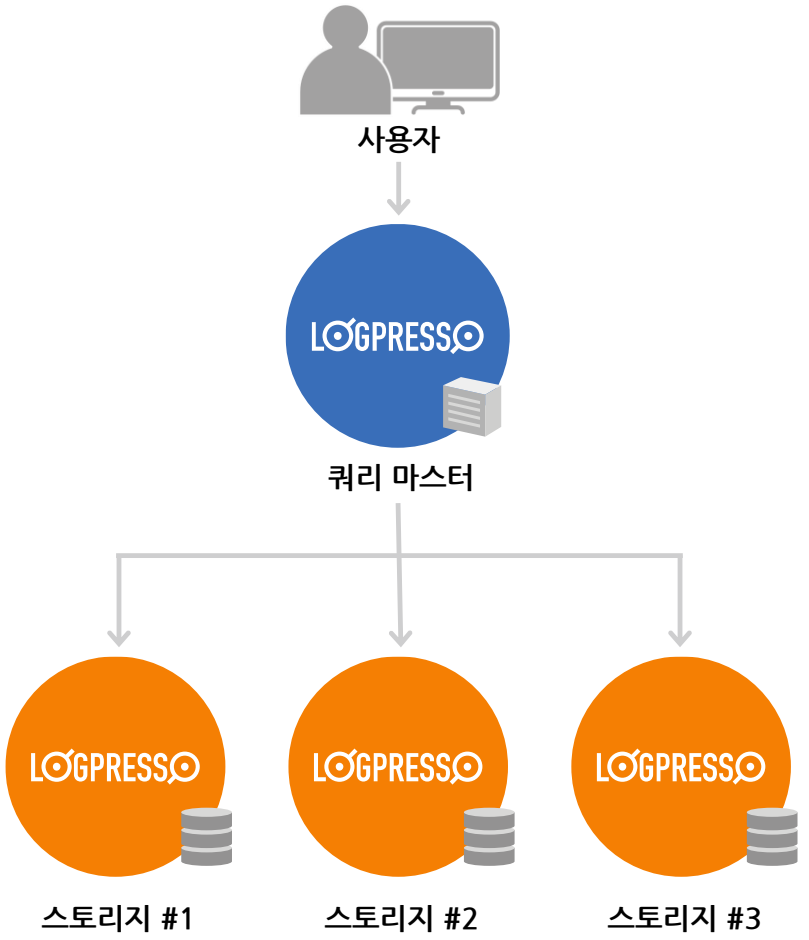
스토리지 노드 쿼리

- table n1:t1 | rex | stats count by src ip
- table n1:t1 | rex | stats count by src ip
- table n1:t1 | rex | stats count by src ip

마스터 노드 병합 쿼리

- merge | stats sum(count) as count by src_ip

각 노드별로 부분 통계 데이터만 전송하고 마스터에서
병합하여 네트워크를 통한 데이터 전송을 최소화 합니다.

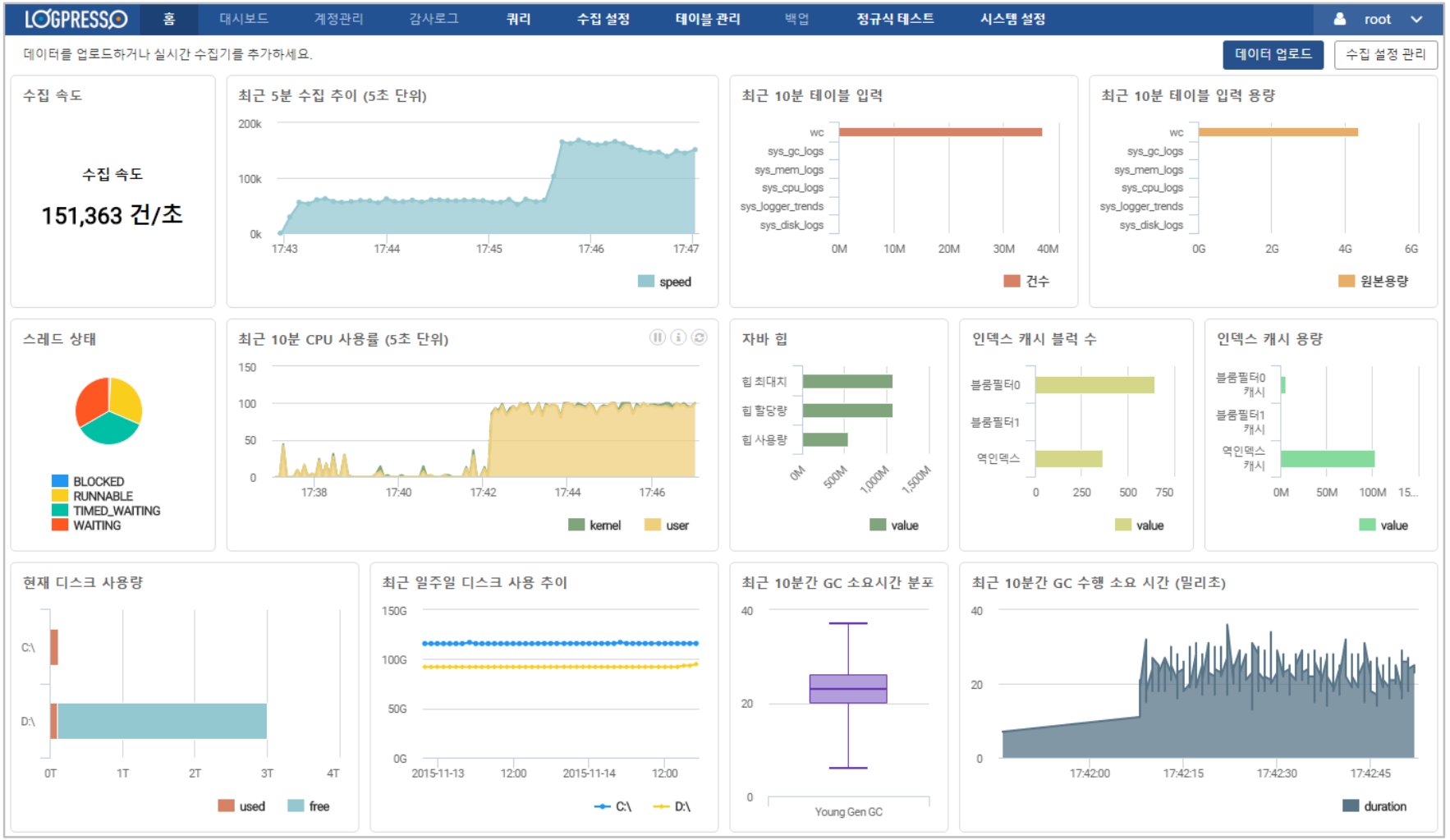


2-4. 시각화

수집 저장 분석 시각화

2-4. 시각화 : (1) 자체 모니터링

애플리케이션 데이터를 수집, 분석 후 통계와 추이 등을 쿼리로 위젯화 하여 도킹 배치

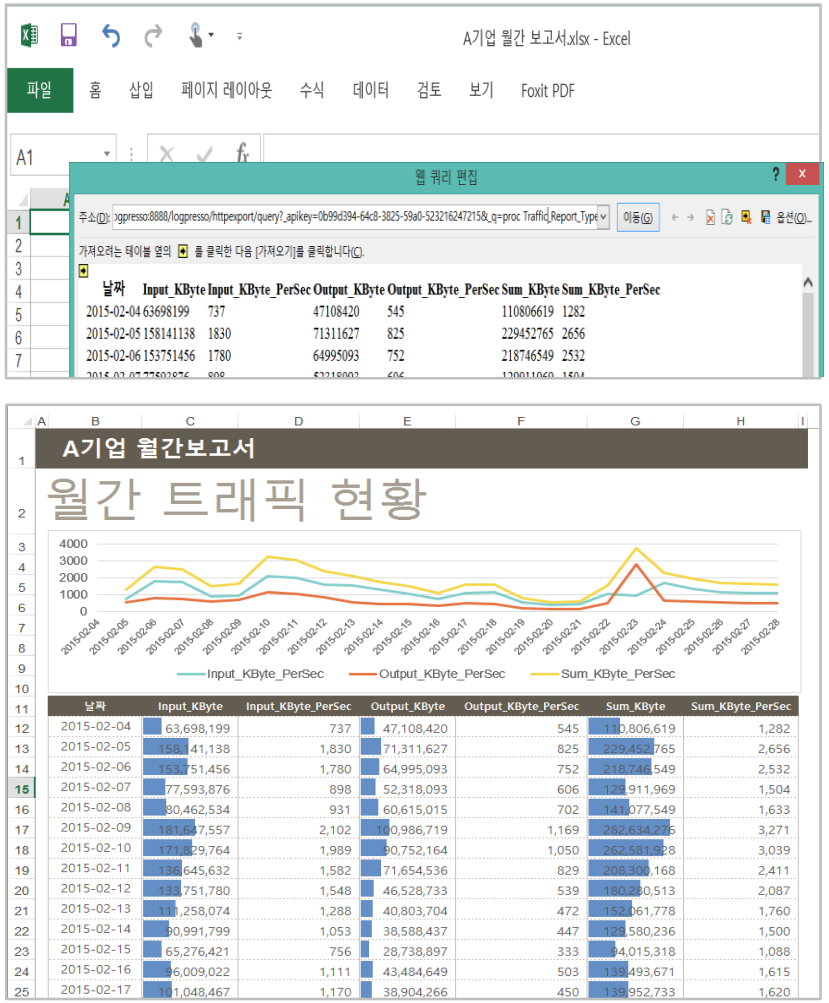


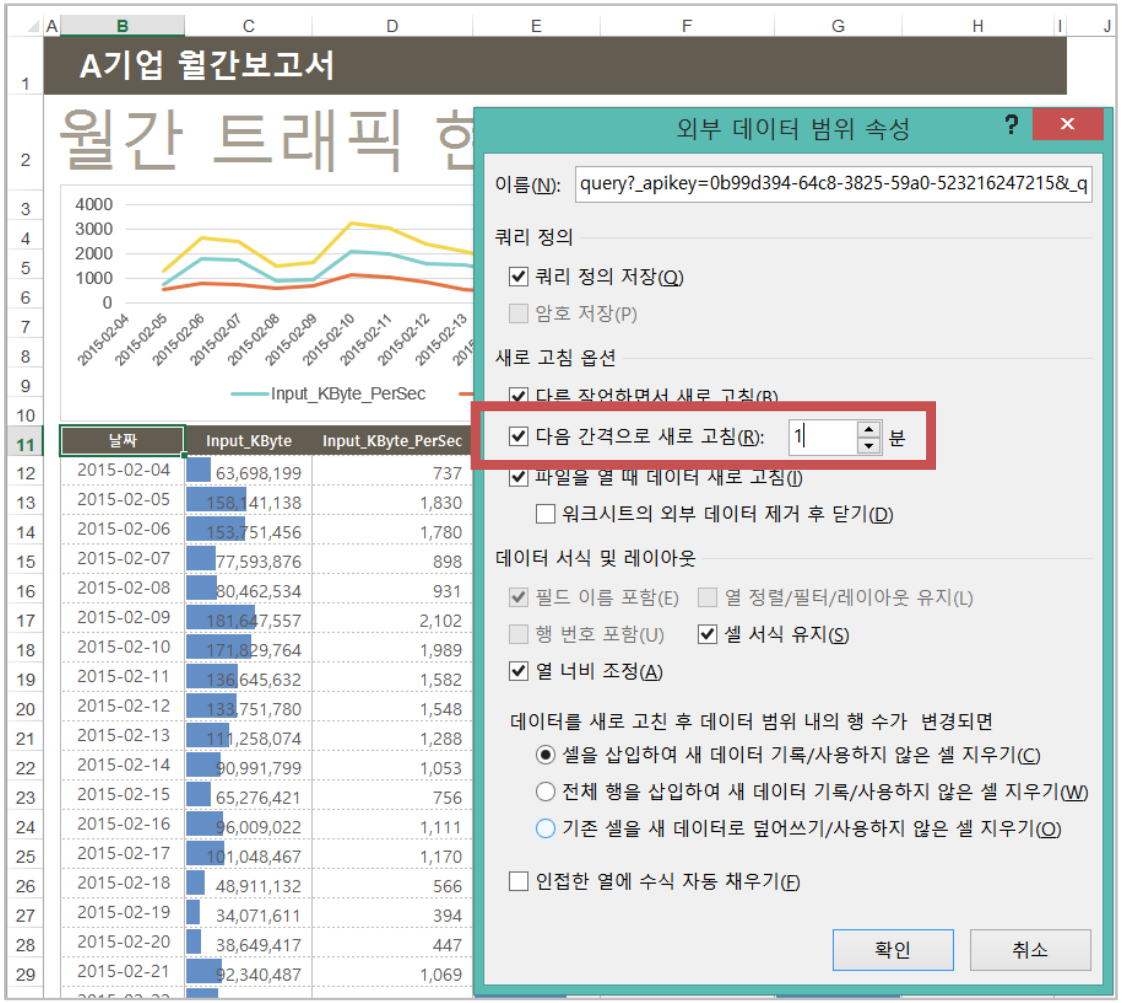
2-3. 시각화 : (2) 대시보드 시각화

대시보드 위젯에 쿼리를 설정하여 각종 통계 및 추이를 실시간으로 시각화합니다.

2-3. 시각화 : (3) 웹쿼리를 통한 엑셀 연동

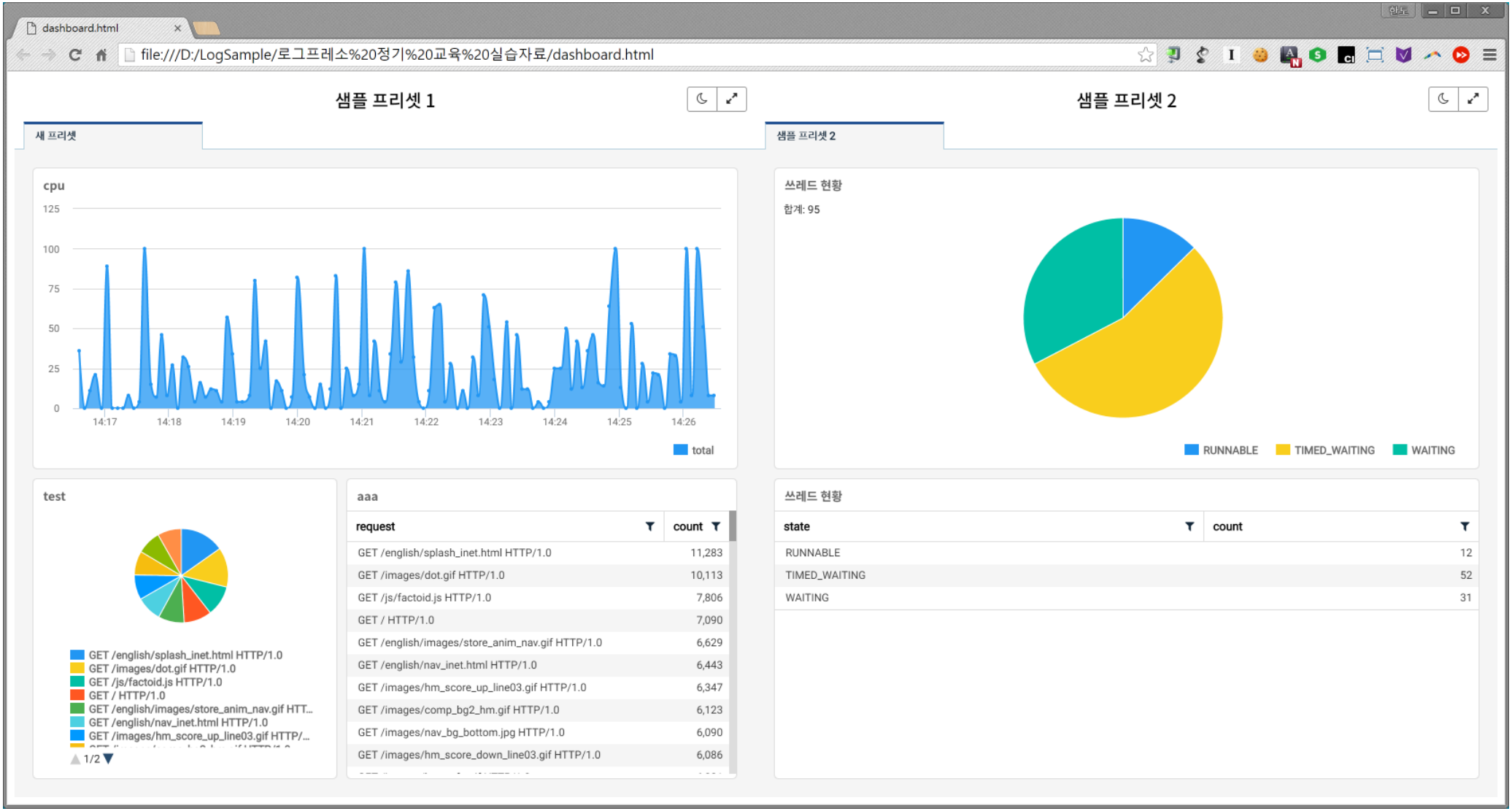
엑셀 웹쿼리를 통해 로그프레소 쿼리 결과를 시트에 삽입하고 자동 갱신되도록 설정할 수 있습니다.





2-3. 시각화 : (4) 대시보드 외부 공유

로그프레소의 대시보드를 외부 웹화면에 그대로 임베딩



로그프레소 기능 요약

벡터화된 쿼리 실행 + 스키마리스 컬럼스토리지 + 대시보드 드릴다운 구현

LOGPRESSO

수집

에이전트 기반 수집
Windows, Linux, Solaris, AIX, HP-UX, OSX

네트워크 기반 수집
SYSLOG, SNMP, JDBC, HTTP, FTP, SFTP, JMX, PCAP, HDFS

모바일 SDK 지원
Android, iOS

저장

비정형 컬럼스토리지
스키마 정의 불필요

실시간 압축
원본의 10%-25% 디스크

실시간 인덱싱
30만 EPS 이상

실시간 암호화
암호화 규제 준수

분석

고속 임의 통계 분석
x100배 빠른 분석 성능

고속 풀텍스트 검색
원본 10억건 검색 1초

스트림/CEP 쿼리
쿼리 기반 실시간 처리

원격 데이터 조인
원격 파일, DB 분석 통합

시각화

사용자정의 대시보드
별도의 BI 툴 불필요

실시간 시각화
수집-분석의 GAP 없음

드릴다운 지원
다차원 시각화 분석

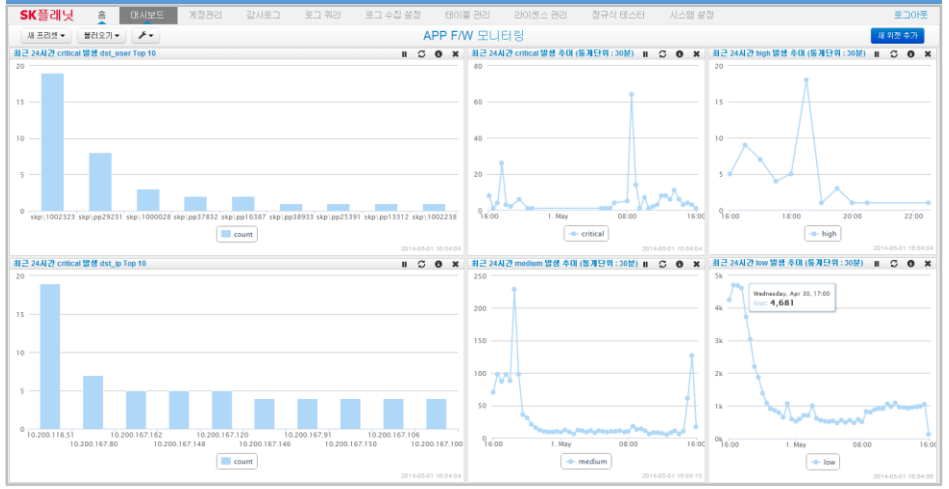
다중 사용자 지원
분석 결과의 공유

3. 적용분야 및 구축사례 : 적용 가능 분야

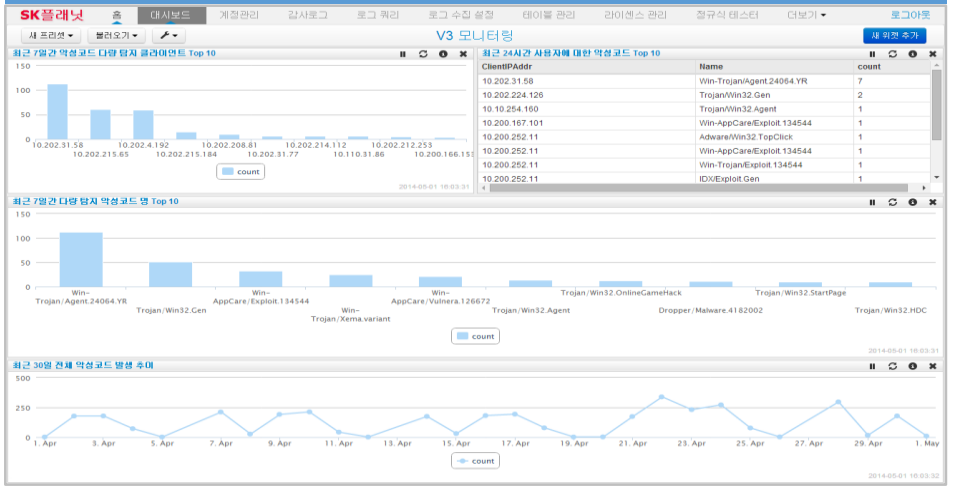
구분	활용 방안	설명
게임	아이템 구매 및 플레이 로그를 분석한, 레벨 최적화 설계 및 아이템 구매 유도 방안 도출	중도 탈락률 및 이탈률이 높은 레벨을 분석하고, 아이템 구매 경향을 분석하여, 레벨의 난이도 조정 및 유료 아이템 구매 유도 방안을 도출한다
	플레이 로그 분석을 통한 게임 봇 실시간 탐지	정상 사용자와 봇의 이용 패턴을 파악하고, 지속적으로 이용 패턴 변화를 분석하여, 봇을 이용한 게임 어뷰징을 방지한다
	결제 및 아이템 구입 로그 분석을 통한 아이템 부정 구입 및 해킹 탐지 등	게임 내 아이템 및 게임 머니 유통 현황과 구입 현황을 실시간 분석하여, 해킹으로 인한 아이템 부정 취득 및 게임 내 경제 인플레이션을 방지한다
모바일 서비스	서버 로그 및 사용자 단말기 로그 통합 분석을 통한 사용량 지표 및 서비스 품질 지표 실시간 추출	지역, 통신망, 스마트폰 플랫폼, 스마트폰 하드웨어 별로 서비스 품질 지표를 추출하여 모바일 애플리케이션 서비스 품질을 관리한다. ex) 음악, 비디오 스트리밍 서비스에서 특히 중요
물류	차량 미터기 및 GPS 데이터 실시간 수집 및 분석을 통한 경로 최적화 및 운행 방식 최적화	경로 및 운전 습관에 따른 유류사용량, 브레이크 라이닝/미션/타이어 손실량 분석. 이를 기반으로 경로 및 운전습관에 따른 인센티브 제공, 동기부여, 비용절감
통신	기지국별 응답시간 및 통신 감도 분석을 통한 실시간 품질 지표 생성 및 경보 시스템 구축	전체 기지국에 대한 실시간 품질 지표를 구축하여, 고객의 클레임 발생에 대비하여 선제적 품질 관리 체계 구축
	셋탑박스 사용자 데이터 수집을 통한, 채널 시청률 및 광고 집행 현황 전수 분석	기존 셋탑박스 네트워크 구성에 별도의 변경을 가하지 않고 데이터를 수집하여, IPTV 및 케이블TV 사용자의 채널 변경 패턴 및 사용량 추출 Ex.) 유료서비스 유입 행태, 광고 반응 행태, 시청 패턴 추출, 실시간 시청량 분석
제조	생산공정 로그 전수검사를 통한, 품질 지표 실시간 추출 및 검사장비 비중 축소를 통한 비용 절감	생산공정에서 발생하는 각종 로그를 전수 분석하여 품질지표를 구축하고, 이를 기반으로 검사장비 사용횟수 및 비중 축소 Ex.) 반도체 공정의 경우 전체 비용의 1/5 가량이 반도체 검사장비 비용
전력	스마트 전력 계량기 실시간 연동을 통한, 수요기반 전력 요금 과금 및 예측 시스템 구축	AMI(스마트 전력 계량기) 데이터를 실시간으로 전송받아, 디테일한 수요 기반 전력 과금 시스템 및 수요 예측 시스템 구축
금융	모든 거래 채널 로그 및 사용자 로그인 데이터를 기반으로, 실시간 이상 금융거래 방지 시스템 구축	이상 거래로 의심되는 각종 장비, 사용자 정보를 프로파일링하고 이상 거래 패턴을 내장해 다양한 로그인 방식 및 거래 방식으로 이루어지는 이상금융거래를 탐지함으로써 고객 및 금융사 손실 최소화

3. 적용분야 및 구축사례 : (1) 빅데이터 보안 - 통합보안

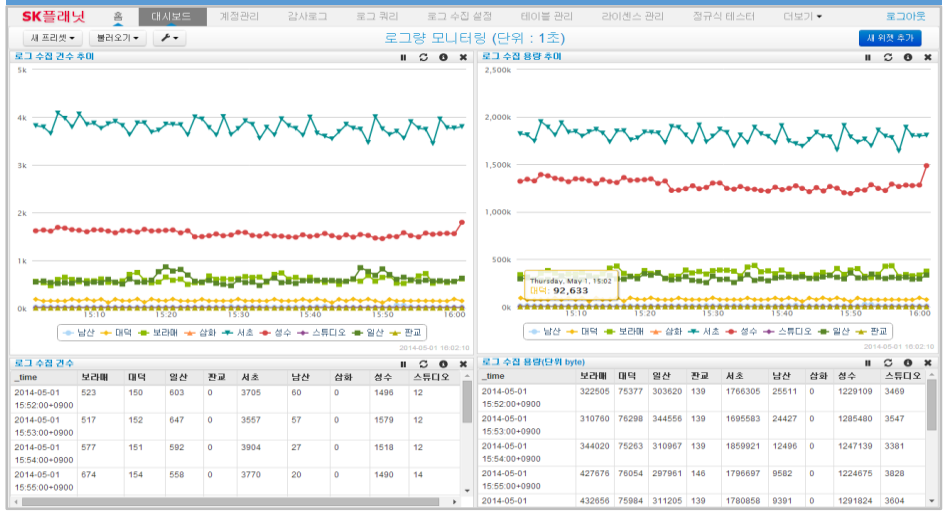
APP FW 모니터링



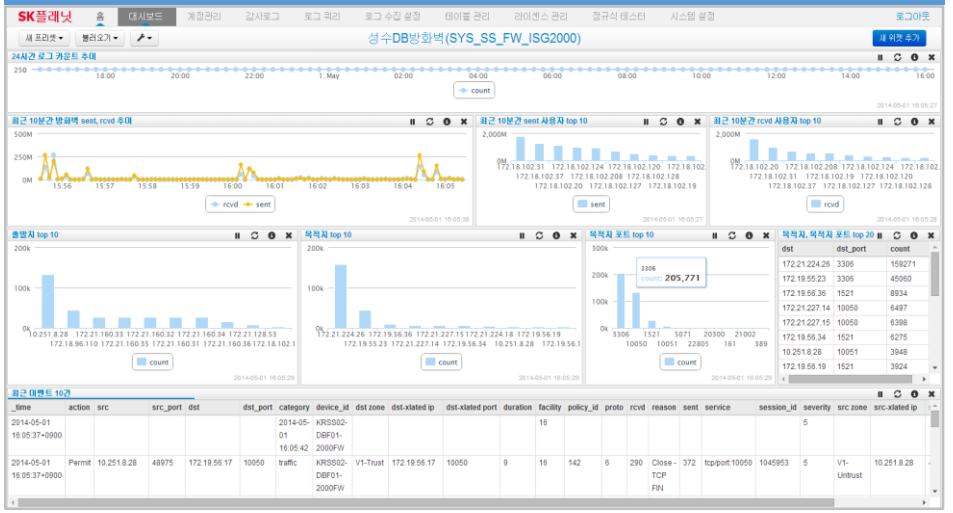
V3 모니터링



로그량 모니터링 - 1초 단위

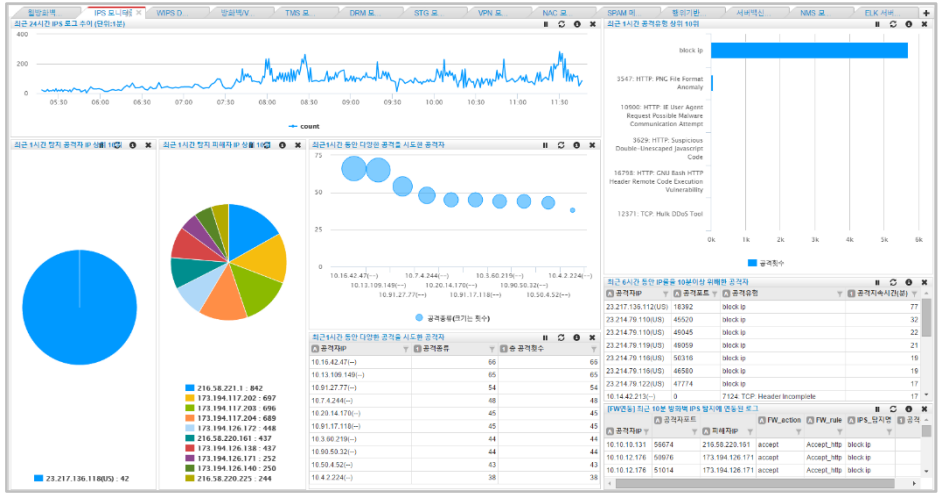


DB 방화벽 모니터링

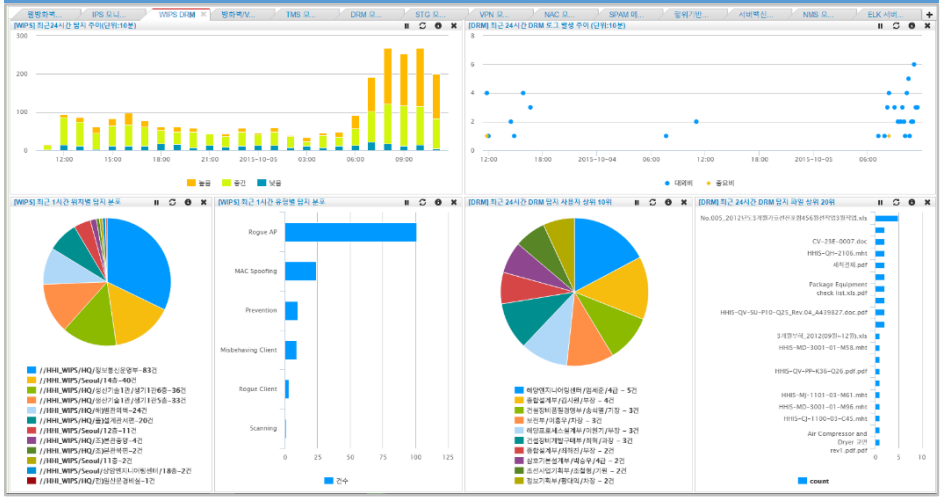


3. 적용분야 및 구축사례 : (1) 빅데이터 보안 - 통합보안

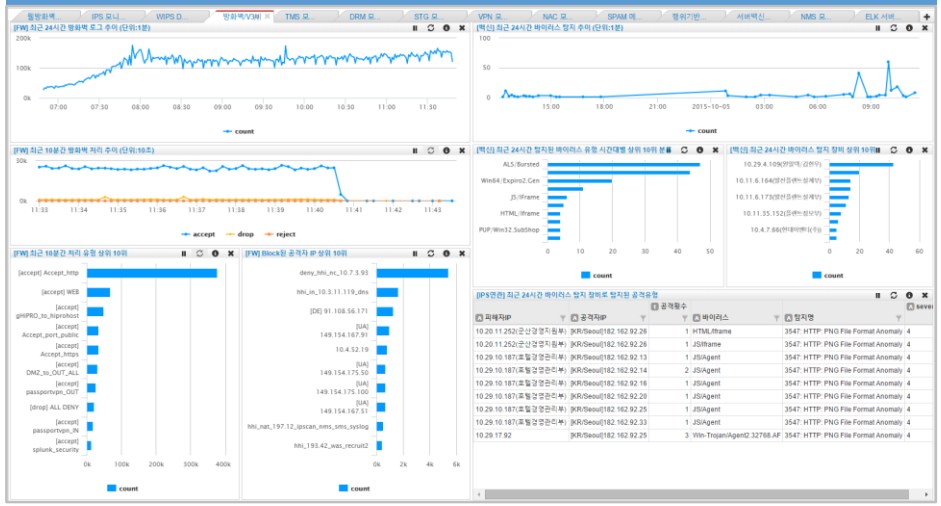
IPS 모니터링



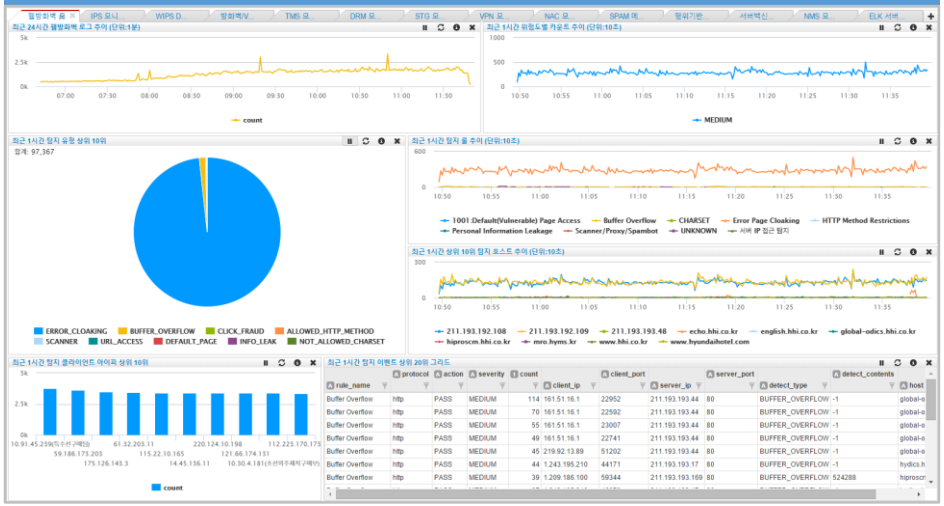
WIPS_DRM 모니터링



방화벽 모니터링 - 1분 단위



웹 방화벽 모니터링 - 1분 단위

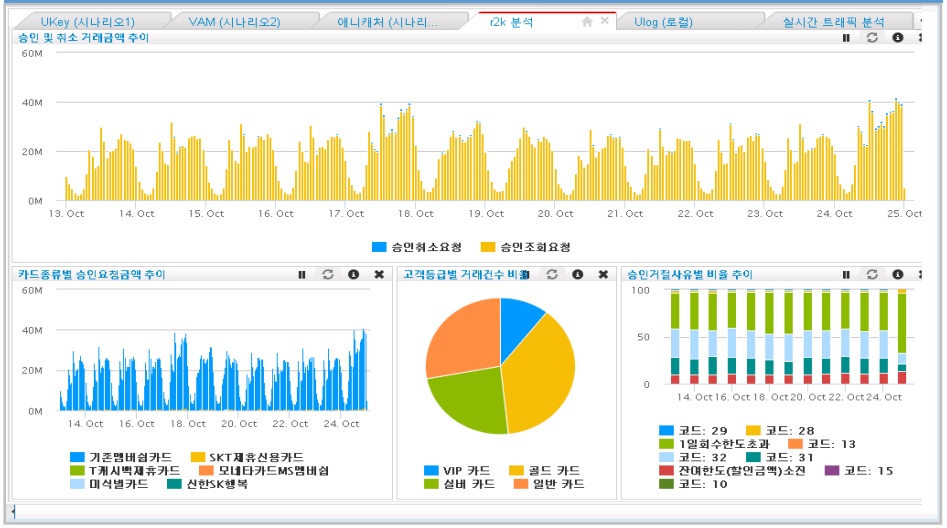


3. 적용분야 및 구축사례 : (2) 방송 - 실시간 시청률 분석 통계

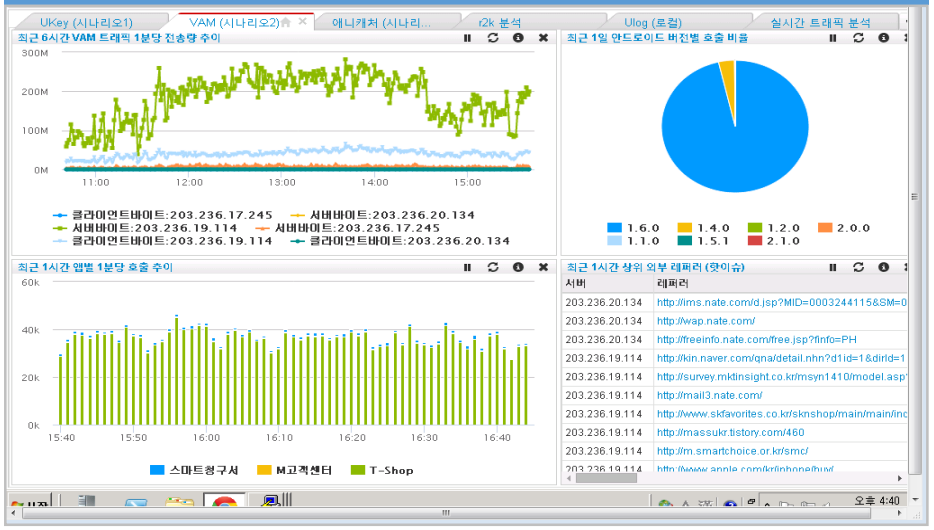
실시간 시청률 분석 - 종합 관제

채널별 상세현황

추이분석 - 일 단위



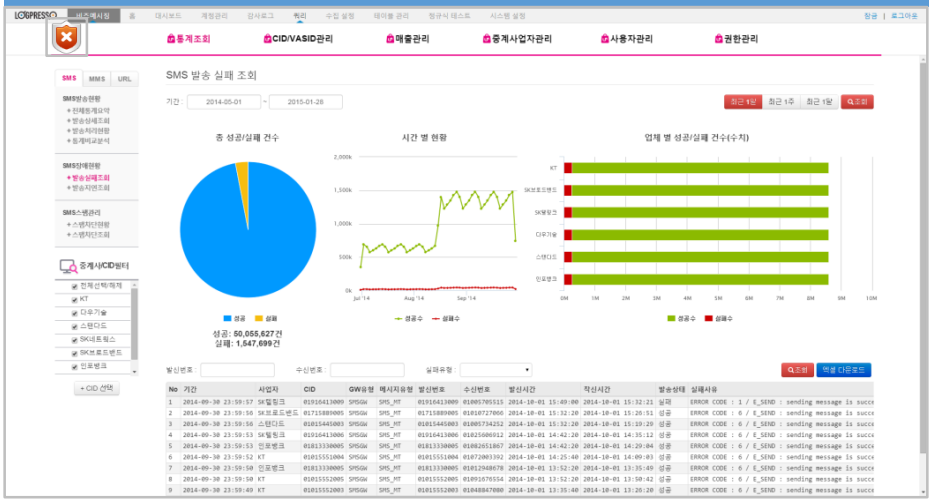
추이분석 - 분 단위



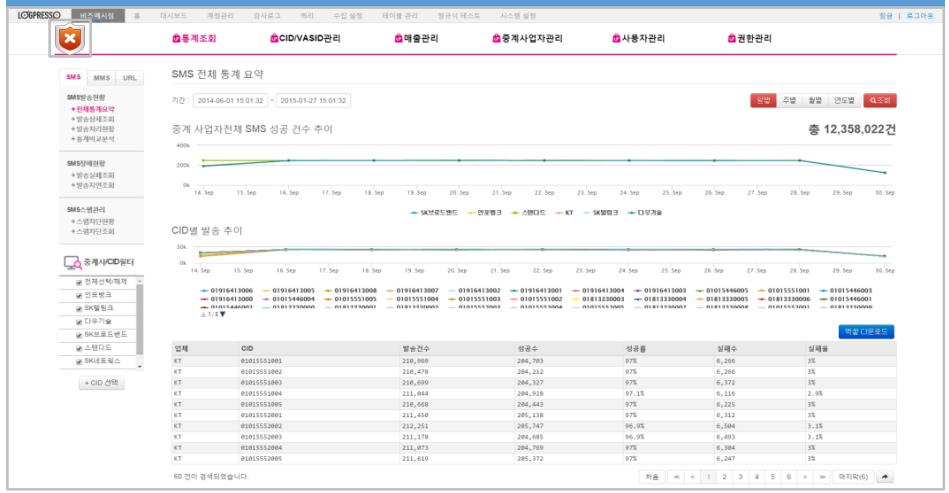
3. 적용분야 및 구축사례 : (3) 통신 - 빅데이터를 활용한 업무 관리

실시간 추이 - 스팸 차단

서비스 별 성공률 분석



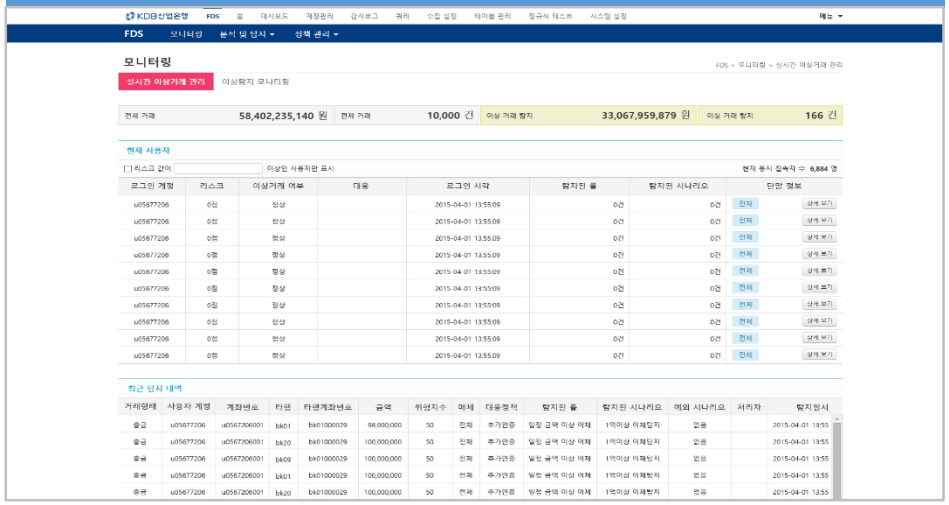
일간 통계



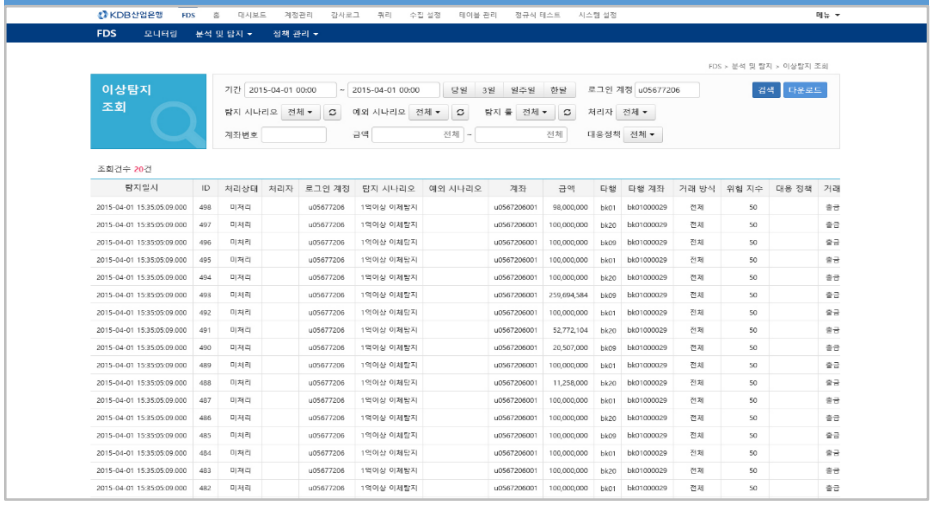
월간 통계

3. 적용분야 및 구축사례 : (4) 금융 - 이상금융거래탐지(FDS)

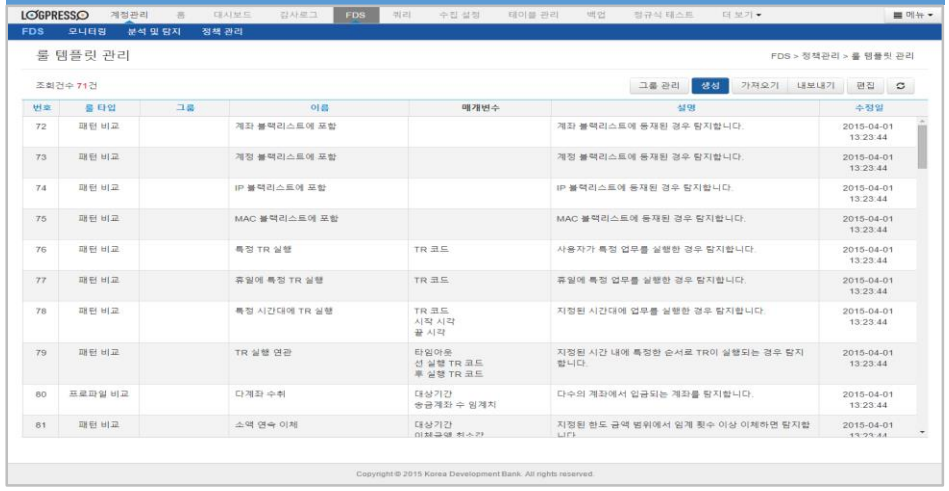
이상 거래 모니터링



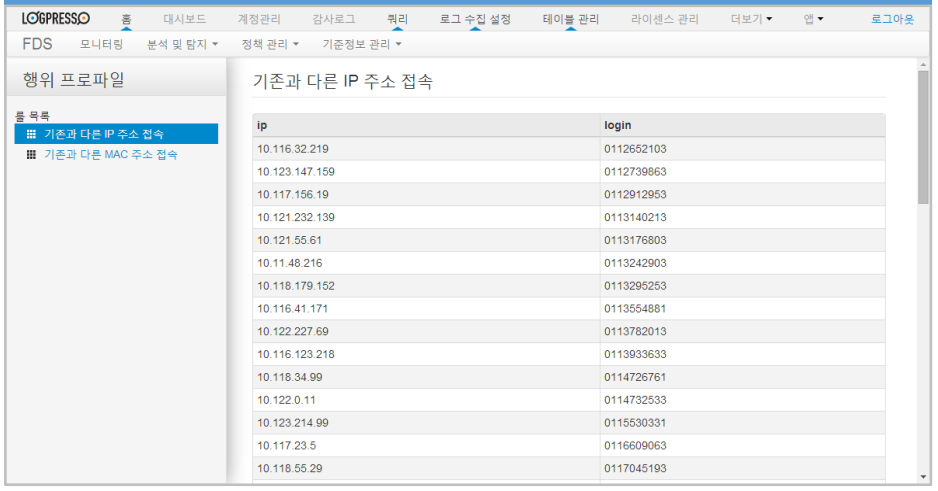
이상 탐지 조회



룰 템플릿 관리



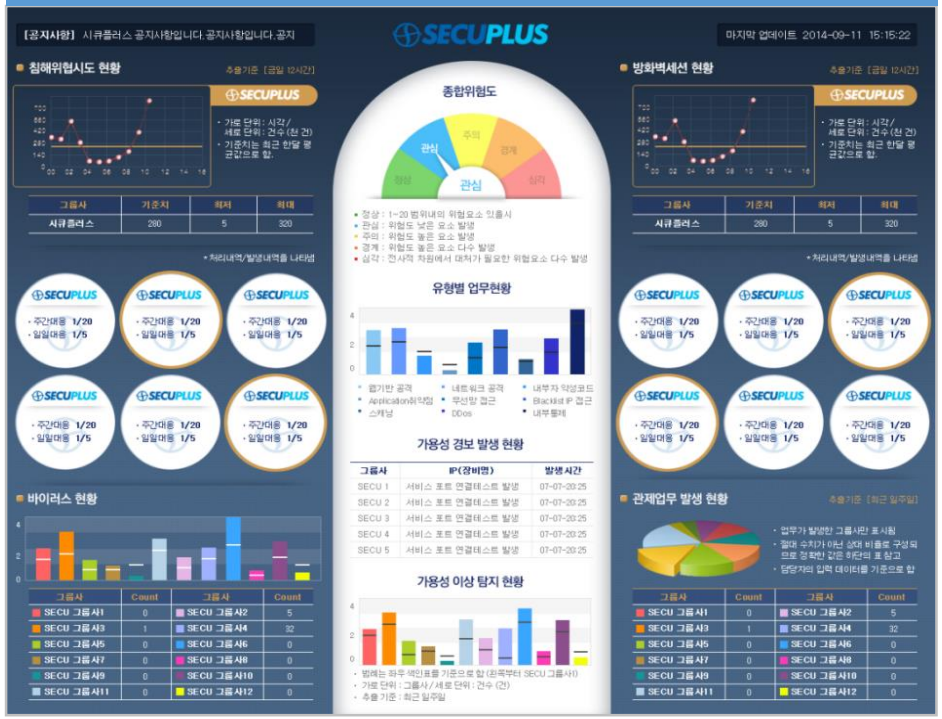
행위 프로파일 조회



3. 적용분야 및 구축사례 : (5) 정보보호 포탈

정보보호 포탈 대시보드 예시 (1/2)

지역별 보안 운영 현황 정보 조회 예시



국가별 보안 운영 현황 정보 조회 예시

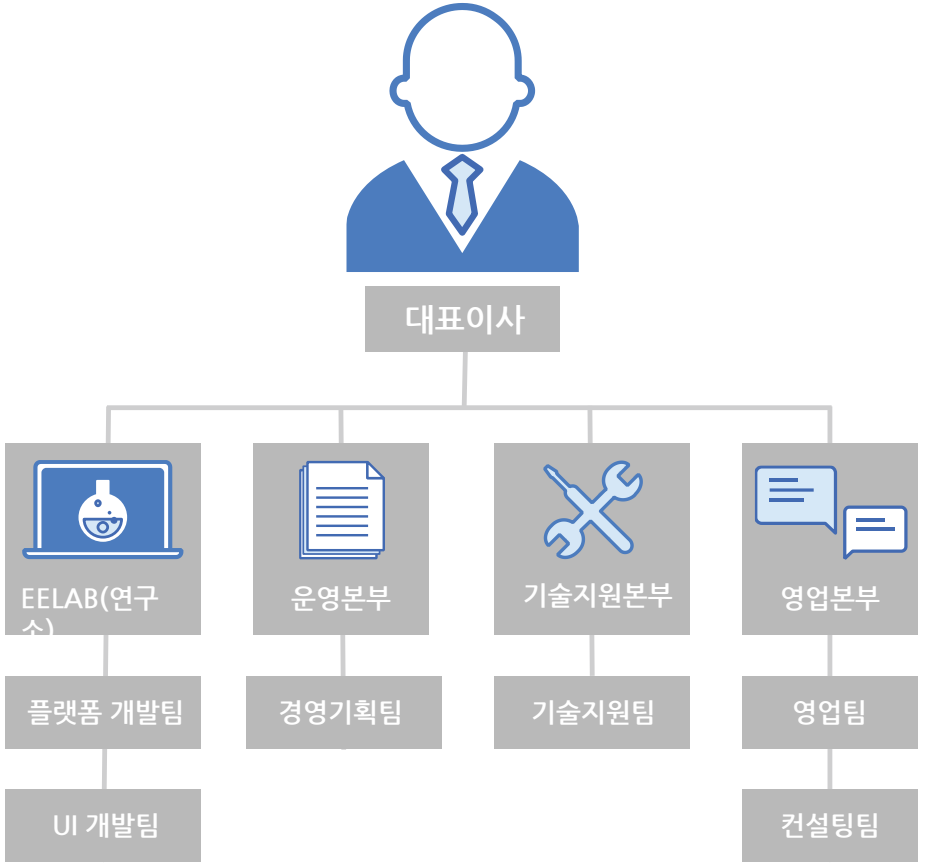
전체 보안 현황을 확인할 수있는 대시보드 형태의 웹 기반 UI 제공

3. 적용분야 및 구축사례 : (6) 구축실적

구분	고객사
공공	ncia 정부통합전산센터, 사이버경찰청 NATIONAL POLICE AGENCY, 병무청, 경상남도교육청, 도로교통공단, KOGAS, 농림축산식품부 Ministry of Agriculture, Food and Rural Affairs, 대한민국육군 Republic of Korea ARMY, 한국교육과정평가원 Korea Institute for Curriculum and Evaluation, KLID 한국지역정보개발원 Korea Local Information Research & Development Institute, 한국 KOREA FIRE SAFETY ASSOCIATION 소방안전협회, 국세청 NATIONAL TAX SERVICE, 관세청 KOREA CUSTOMS SERVICE, 캠코 한국자산관리공사, 서울특별시 SEOUL METROPOLITAN GOVERNMENT
금융	금융결제원 金融決濟院, 우리은행, NH농협생명, 저축은행중앙회, 수협, 신협, DGB금융그룹 대구은행, 키움증권 대한민국 주식시장 점유율 9년 연속 1위, KDB산업은행, MIRAE ASSET 미래에셋증권, 이비카드, KRX 한국거래소 KOREA EXCHANGE, KB캐피탈, IBK기업은행 금융그룹 IBK시스템, Standard Chartered SC스탠다드 캐피탈, 우리투자증권, Your Best Credit Partner SGI서울보증
방송 통신	LG U+, SK telecom, SK 플래닛, KBS
일반	삼성꿈장학재단 Samsung Dream Scholarship Foundation, 농협정보시스템, 중앙대학교병원, ChosunMedia 디지틀조선일보, Hansol 한솔넥스지, 현대중공업, NICE, HYOSUNG FMS, 세브란스 SEVERANCE 강남세브란스병원 GANGNAM SEVERANCE HOSPITAL, DSME 대우조선해양, KEPCO 전력연구원
협력사	SECUPLUS, ANSWER BIG DATA & BIG INSIGHT, DATAValue, WOOSAN CnC Woosan Communications & Computers (주)우산씨앤씨, IGLOOSECURITY, geni NETWORKS, KE

4. 회사 소개 : (1) 회사 소개

회사명	㈜ 이디엄
설립시기	2013. 3
주소	서울특별시 마포구 새창로 7 SNU 장학빌딩 1601호
사업분야	• 빅데이터 분석 컨설팅 서비스 • 빅데이터 플랫폼 기술 개발 • 빅데이터 분석 소프트웨어 개발
자본금	15,000만원
임직원수	21명 - 개발(15명), 기술(3명), 영업(2명), 관리(1명)
특허	• 제 10-1112568 실시간 역인덱스 생성 및 검색 기술 • 제 10-1459018 이벤트 처리 시스템의 이벤트 처리 방법 • 제 10-1487859 자바 프로그램 실행 시 유저 데이터그램 패킷을 수집하는 방법 • 제 10-1542526 빅데이터 입력 및 처리 방법 • 제 10-2025-0113619 메시지의 필드 인덱싱 방법 (출원)



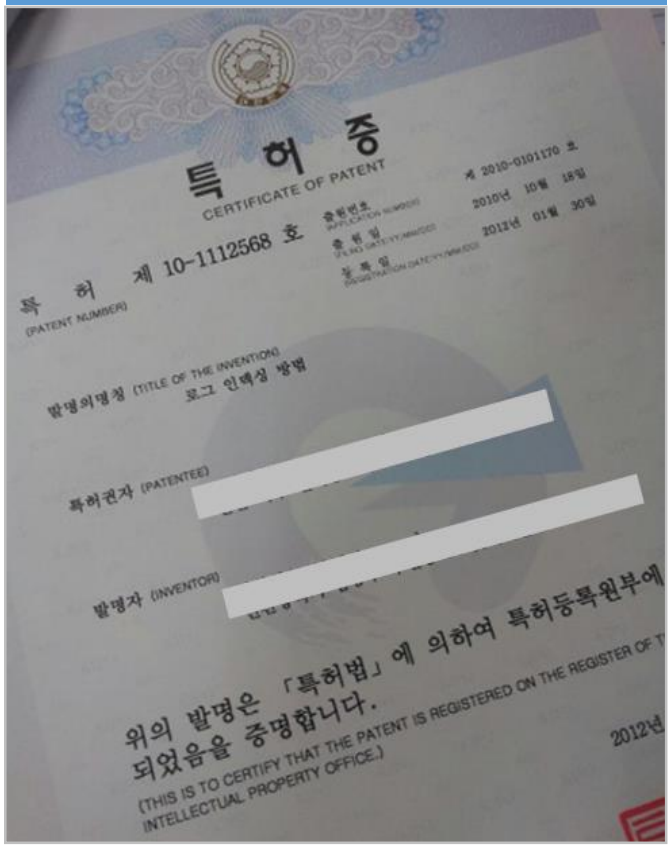
(주)이디엄은 고객이 데이터에 기반하여 효율적인 의사결정을 할 수 있도록, 자체 기술로 빅데이터 솔루션을 개발 및 공급하는 회사로서, 실시간 빅데이터 수집, 검색, 분석 및 시각화 등 전체 과정 통합구축을 지원합니다.

4. 회사 소개 : (2) 연혁

연혁	2013	멀티코어 기반 실시간 고속 풀텍스트 엔진 개발 완료 (2013. 4)
		금융결제원, 정부통합전산센터, 농협생명, 제품 공급 (2013. 6 ~ 7)
		경찰청, SKT, 병무청 제품 공급 (2013. 9~12)
	2014	LG유플러스 실시간 서비스 품질 모니터링 시스템 구축 (2014. 1 ~ 2)
		미래에셋증권 금융이상거래탐지시스템 구축 (2014. 1 ~ 2)
		특허등록 '이벤트 처리 시스템의 이벤트 처리 방법' (2014. 10)
		SK플래닛 통합로그관리시스템 구축 (2013. 11 ~ 2014. 12)
		LG백오피스 이통메시징 실시간 분석통계시스템 구축 (2014. 12 ~ 현재)
		대구은행 금융이상거래탐지시스템 구축 (2014. 12~ 현재)
	2015	키움증권 통합로그관리시스템 구축 (2014. 1 ~ 3)
		현대중공업 통합관제 구축 (2014. 4 ~ 6)
		SK플래닛 시럽페이 FDS 구축 (2014. 5~ 7)
		한솔넥스지 차세대 통합관제 구축 (2015. 8 ~ 현재)
		KBS 통합관제 구축 (2015. 9 ~ 현재)

4. 회사 소개 : (3) 특허 및 보도자료

특허 제 10-1112568 실시간 역인덱스 생성 및 검색 기술



특허 증

CERTIFICATE OF PATENT

특 허 제 10-1112568 호

(PATENT NUMBER)

출원번호 제 2010-0101170 호

출 원 일 2010년 10월 18일

등 록 일 2012년 01월 30일

발명의명칭 (TITLE OF THE INVENTION)

로그 인덱싱 방법

특허권자 (PATENTEE)

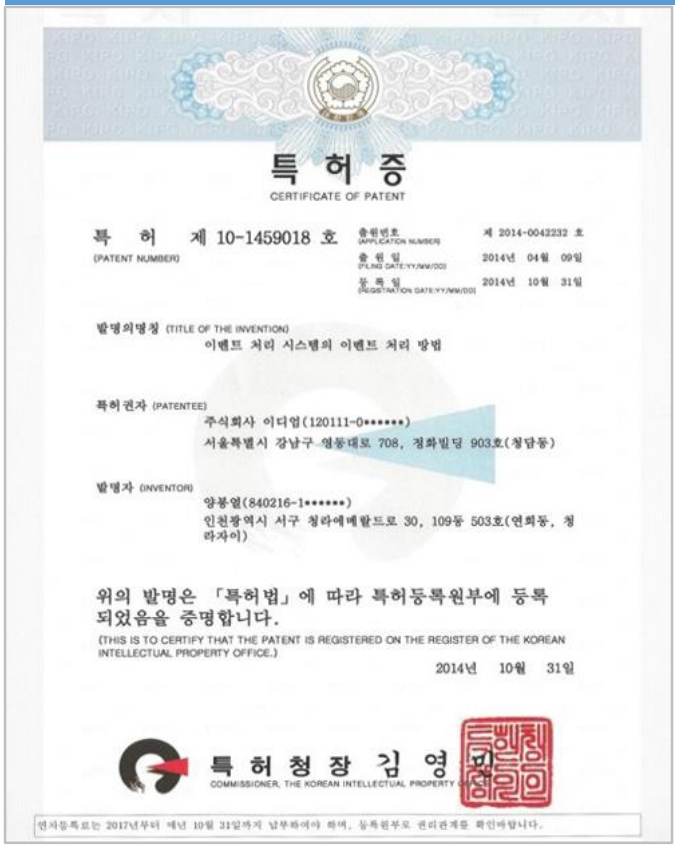
발명자 (INVENTOR)

위의 발명은 「특허법」에 의하여 특허등록원부에 등록되었음을 증명합니다.

(THIS IS TO CERTIFY THAT THE PATENT IS REGISTERED ON THE REGISTER OF THE KOREAN INTELLECTUAL PROPERTY OFFICE.)

2012년

특허 제 10-1459018 이벤트 처리시스템의 이벤트처리 방법



특 허 증

CERTIFICATE OF PATENT

특 허 제 10-1459018 호

(PATENT NUMBER)

출원번호 (APPLICATION NUMBER) 제 2014-0042232 호

출 원 일 (FILING DATE:YY/MM/DD) 2014년 04월 09일

등 록 일 (REGISTRATION DATE:YY/MM/DD) 2014년 10월 31일

발명의명칭 (TITLE OF THE INVENTION)

이벤트 처리 시스템의 이벤트 처리 방법

특허권자 (PATENTEE)

주식회사 이디엄(120111-0******)

서울특별시 강남구 영동대로 708, 정화빌딩 903호(청담동)

발명자 (INVENTOR)

양봉열(840216-1******)

인천광역시 서구 청라에메랄드로 30, 109동 503호(연희동, 청라자이)

위의 발명은 「특허법」에 따라 특허등록원부에 등록되었음을 증명합니다.

(THIS IS TO CERTIFY THAT THE PATENT IS REGISTERED ON THE REGISTER OF THE KOREAN INTELLECTUAL PROPERTY OFFICE.)

2014년 10월 31일

특허청장 김 영

COMMISSIONER, THE KOREAN INTELLECTUAL PROPERTY

연차등록료는 2017년부터 매년 10월 31일까지 납부하여야 하며, 등록원부로 권리관계를 확인하십시오.

디지털 타임즈 - 2014. 06. 24

Zoom up 기업열전 이디엄

외산보다 5배빠른 '빅데이터 검색엔진' 강자

<그림제공>

회사개요

3일 걸리던 분석시간 초단위로 단축 기업 의사결정 지원

폭증하는 데이터 환경 대비… 실시간 데이터 분석 '강점'

질의 & 응답
Question and Answer

감사합니다
Thank you
-
LOGPRESSO