JL Infra (주) 제이엘 인프라 Co.,Ltd

For IoT & BigData
INFINIFLUX
THE WORLD'S FASTEST TIME SERIES DBMS

The World Fastest Time Series DBMS
for IoT and Big Data

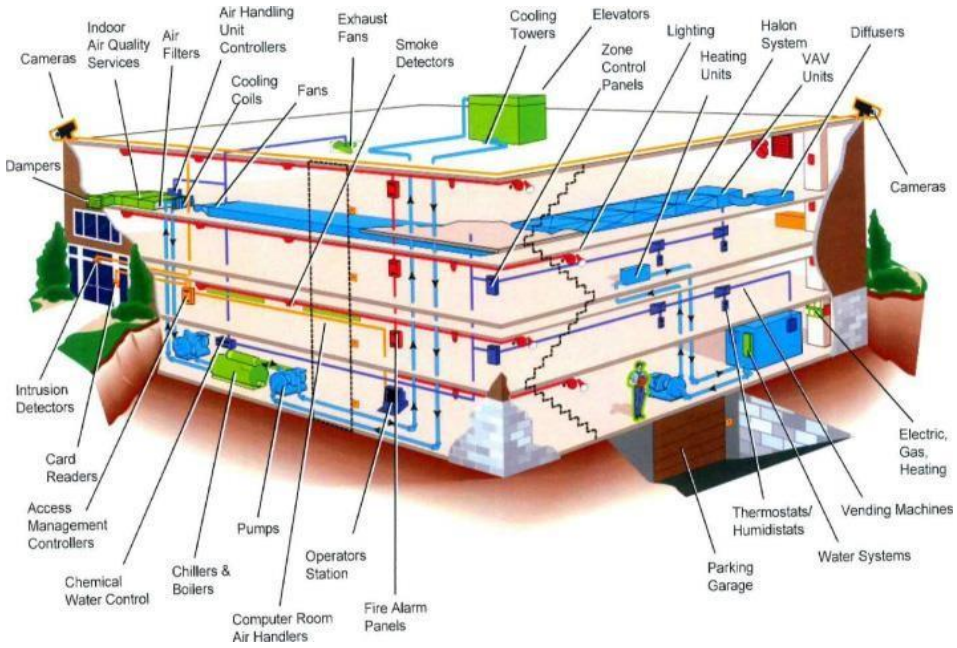
InfiniFlux Corp.

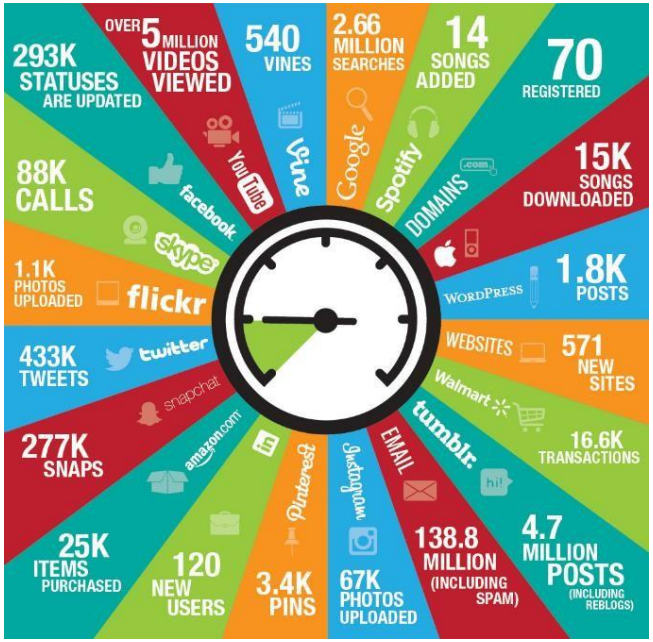
Table of Contents

The World Fastest DBMS ***InfiniFlux***

1. IoT 빅데이터 시대

데이터 소스의 폭증





- IoT 디바이스의 숫자가 **매년 57%씩 증가**
- 2020년에는 **500억 개**가 될 것으로 전망
- Internet 환경의 Human based 로그 데이터도 증가

실시간 분석필요성

- 실시간 모니터링
- 비정상 상황에 대한 즉시 대응
- 운영 인텔리전스 (Operational Intelligence)
- 실시간 엣지 (Edge) 분석

2. 시계열 빅데이터

머신 데이터, 로그 데이터, 센서 데이터

```
Dec 17 11:16:51 ip-172-31-15-249 sshd[18548]: input_userauth_request:
Dec 17 11:16:51 ip-172-31-15-249 sshd[18548]: Received disconnect from
Dec 17 11:16:52 ip-172-31-15-249 sshd[18550]: Invalid user admln from
Dec 17 11:16:52 ip-172-31-15-249 sshd[18550]: input_userauth_request:
Dec 17 11:16:52 ip-172-31-15-249 sshd[18550]: Received disconnect from
Dec 17 11:16:53 ip-172-31-15-249 sshd[18552]: Invalid user jenkins from
Dec 17 11:16:53 ip-172-31-15-249 sshd[18552]: input_userauth_request:
Dec 17 11:16:54 ip-172-31-15-249 sshd[18552]: Received disconnect from
Dec 17 11:16:55 ip-172-31-15-249 sshd[18554]: Invalid user hadoop from
Dec 17 11:16:55 ip-172-31-15-249 sshd[18554]: input_userauth_request:
Dec 17 11:16:55 ip-172-31-15-249 sshd[18554]: Received disconnect from
Dec 17 11:16:56 ip-172-31-15-249 sshd[18556]: Invalid user nagios from
Dec 17 11:16:56 ip-172-31-15-249 sshd[18556]: input_userauth_request:
Dec 17 11:16:56 ip-172-31-15-249 sshd[18556]: Received disconnect from
Dec 17 11:16:57 ip-172-31-15-249 sshd[18558]: Invalid user webadmin fro
Dec 17 11:16:57 ip-172-31-15-249 sshd[18558]: input_userauth_request:
Dec 17 11:16:57 ip-172-31-15-249 sshd[18558]: Received disconnect from
Dec 17 11:16:58 ip-172-31-15-249 sshd[18560]: Invalid user postgres from
Dec 17 11:16:58 ip-172-31-15-249 sshd[18560]: input_userauth_request:
Dec 17 11:16:58 ip-172-31-15-249 sshd[18560]: Received disconnect from
Dec 17 11:16:59 ip-172-31-15-249 sshd[18562]: Invalid user git from 190.
Dec 17 11:16:59 ip-172-31-15-249 sshd[18562]: input_userauth_request:
Dec 17 11:16:59 ip-172-31-15-249 sshd[18562]: Received disconnect from
Dec 17 11:17:01 ip-172-31-15-249 sshd[18564]: Invalid user git from 190.4
Dec 17 11:17:01 ip-172-31-15-249 sshd[18564]: input_userauth_request: i
Dec 17 11:17:01 ip-172-31-15-249 sshd[18564]: Received disconnect from
Dec 17 11:17:02 ip-172-31-15-249 sshd[18566]: Invalid user git from 190.4
Dec 17 11:17:02 ip-172-31-15-249 sshd[18566]: input_userauth_request:
```

10,000 EA * 1초 * 100 Bytes

9억 건

80GB

3,145억 건

3TB

시계열 빅데이터

시계열 특성을 지원하는 최적화된 제품 필요

데이터 특성 (정형/반정형)

시간흐름에따라발생하는 로그,이벤트정보

단순파일저장

대용량로그이벤트활용 방법부재원인

ID,상태정보 포함

해당데이터소스의ID 및 상태정보를반드시포함

시간데이터

시간을기준으로 각종통계, 분석수행

매우빠른 생성속도

동일패턴, 지속적으로 빠르게생성

모니터링

현상및데이터추이이해

분석

과거이벤트확인및이해

방지

장애, 침입탐지

예측

미래비정상상황및대처가능

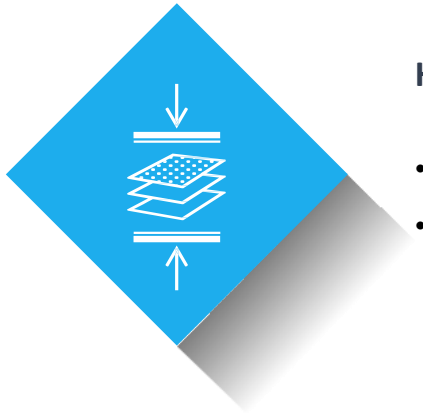
데이터 처리 요구 사항

Real-time writes

- 다양한 원시 데이터 소스 지원
- 초당 수만 ~ 수십만 건 저장

On-time Analytics

- 실시간 인덱싱
- 실시간 질의 처리



High Compression

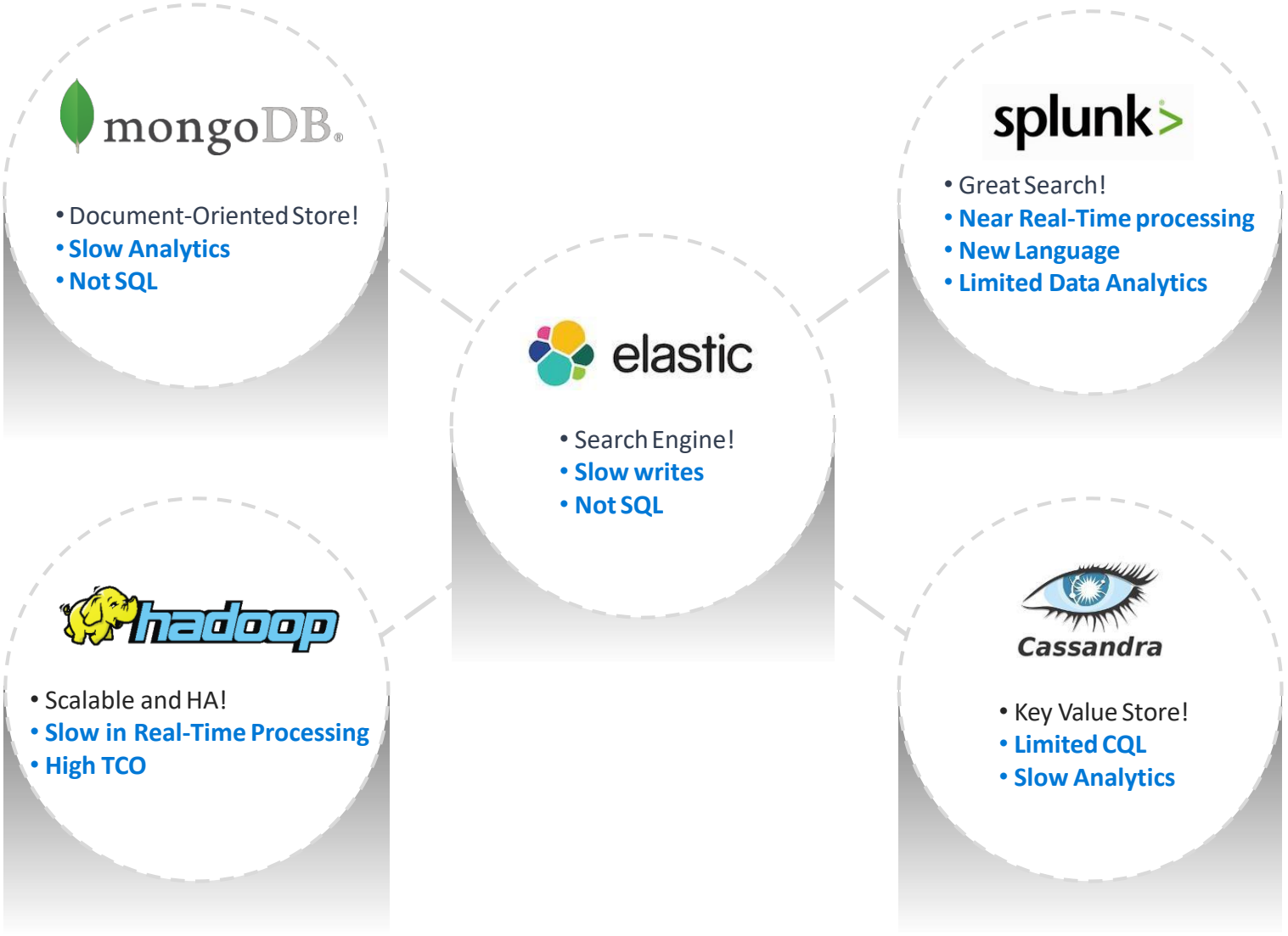
- 스토리지 공간 절약
- 시스템 자원 효율 증대

Highly Scalable

- 노드 추가 확장 기능
- 고가용성

기존 솔루션 한계 : RDBMS

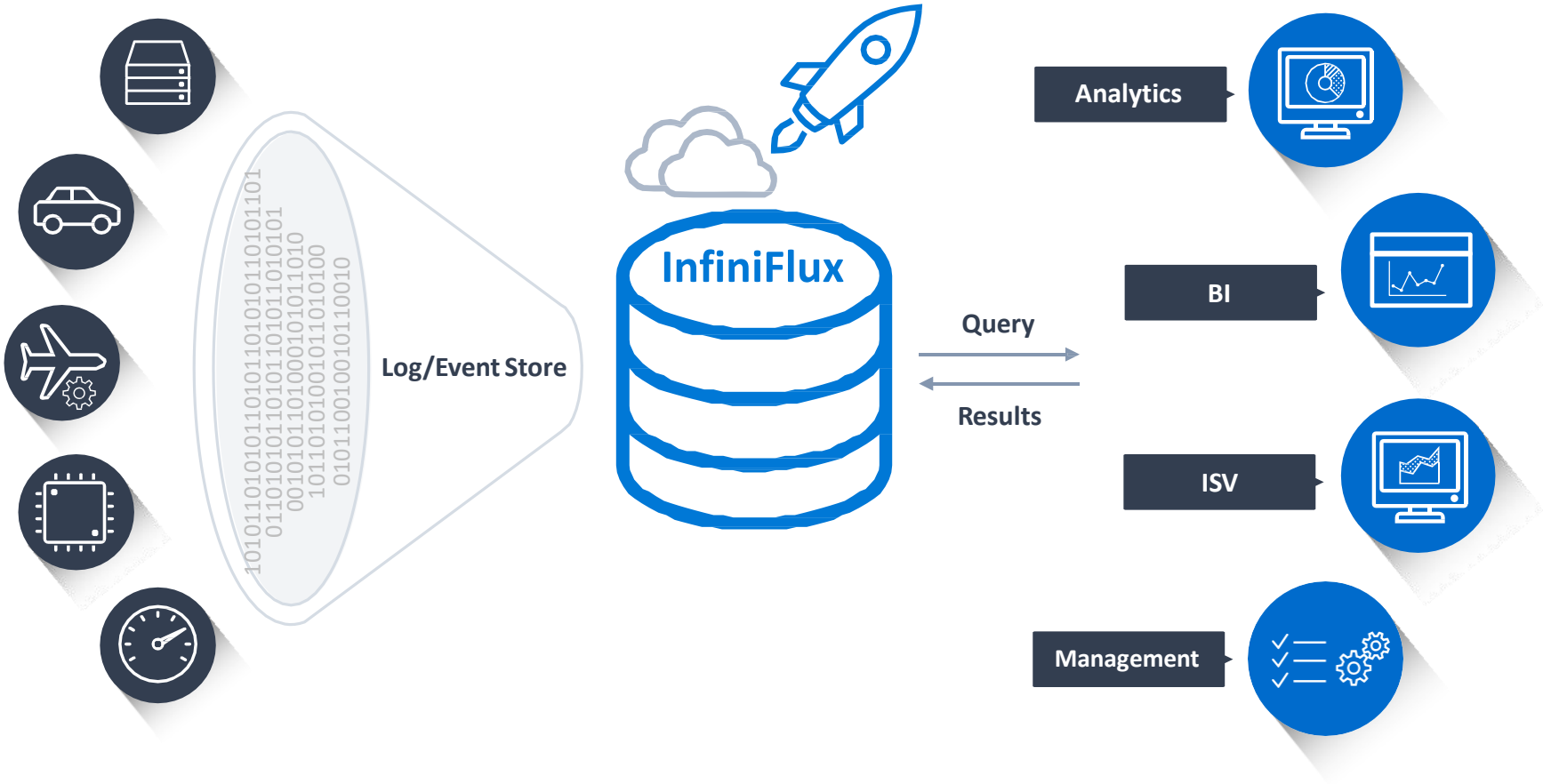
기존 솔루션 한계 : 타 솔루션의 한계



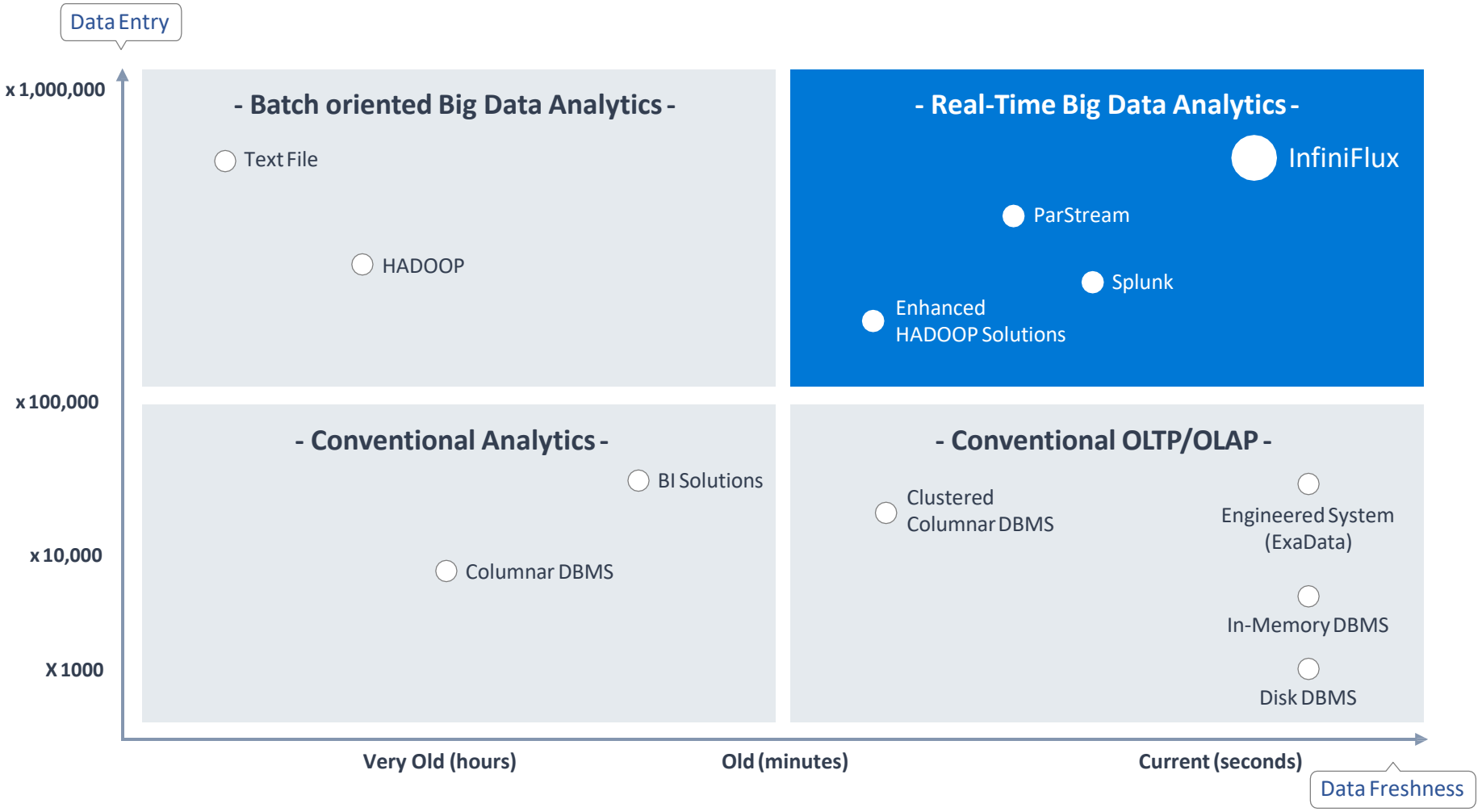
3. 제품 개요

Fastest Time Series DBMS

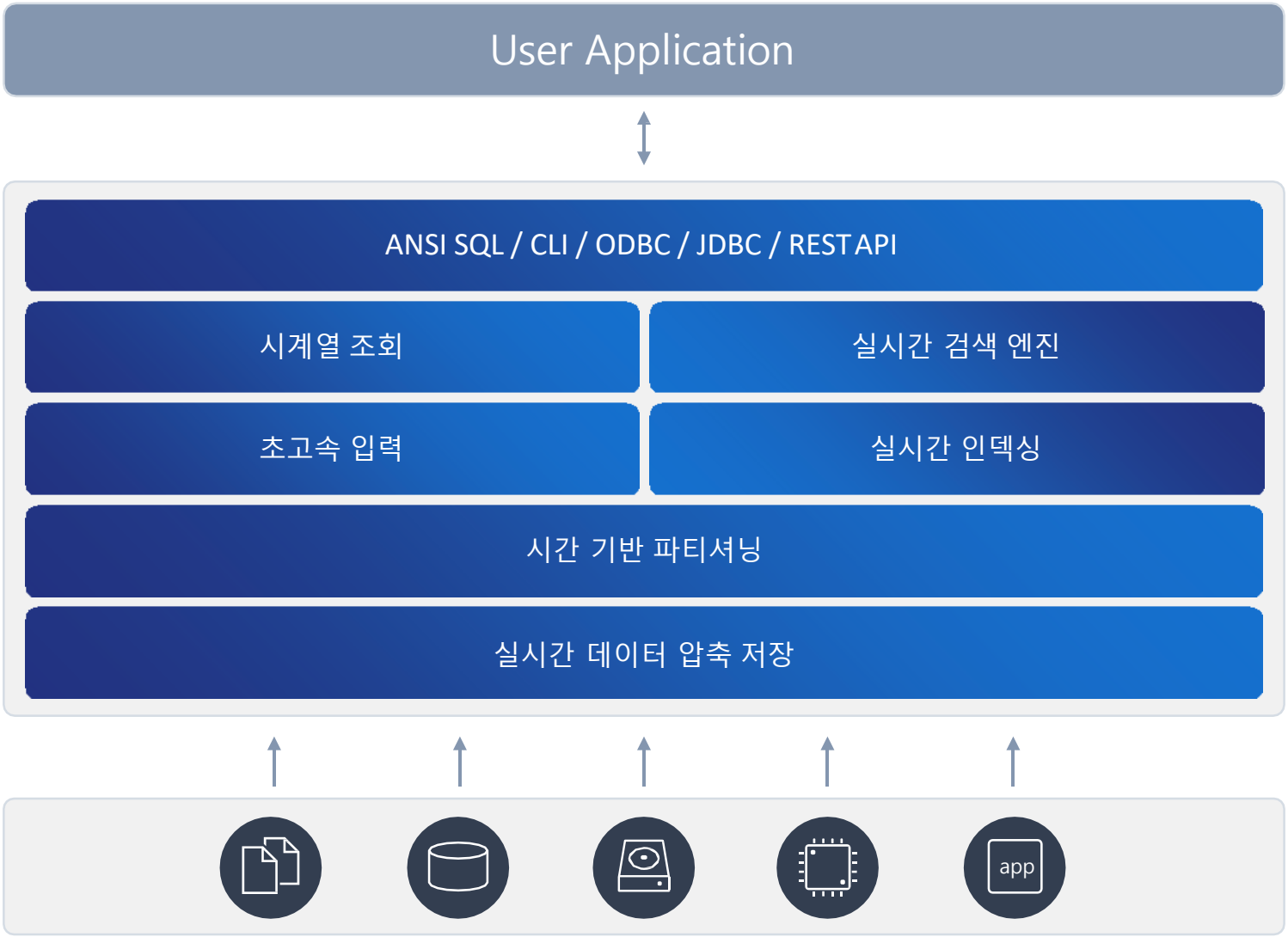
Fast Write & Real-Time Search



Product Positioning

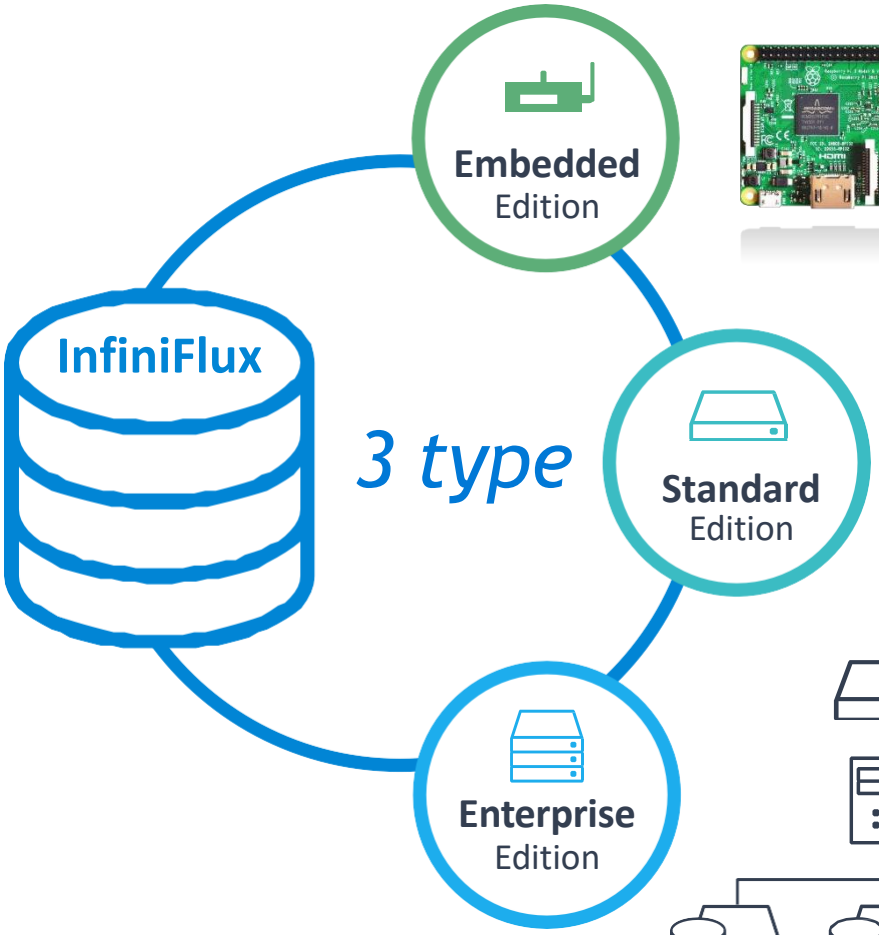


Product Architecture



4. 제품군

3 Type Editions



Real-time Edge Analytics for
IoT sensor data

Fast data processing performance in a
single node

Big data technology revolution in
multi-nodes cluster

Embedded Edition

Edge Analytics를 위한 최고 솔루션

- **CPU** : ARM, x86 지원
- **O/S** : Linux(Redhat, CentOS, Fedora, Ubuntu)

  Real-time Linux (Wind river Linux)

  Windows (64bit 2017/Q1)
- **성능** : 20,000~ 70,000 EPS 입력
- 2017년 Q1 정식 버전 출시

Standard Edition

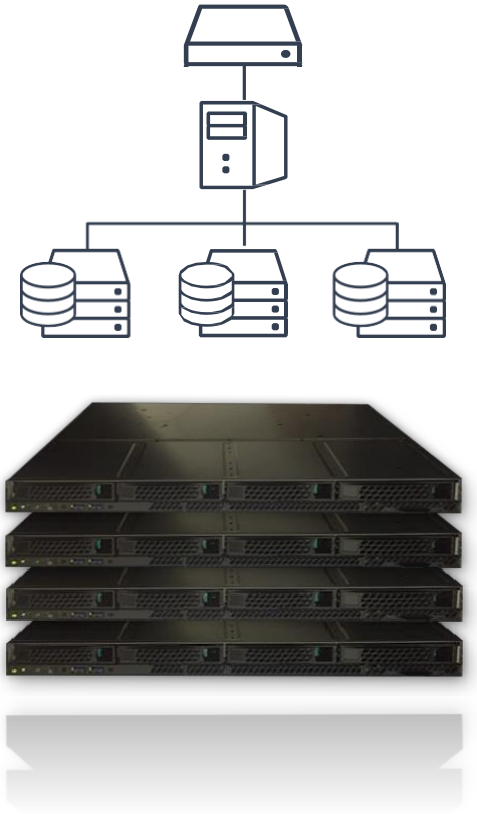
단일 노드에서 최고의 데이터 처리 성능 발휘

- 초당 수만 ~ 수십 만 건 입력 가능
- 데이터 처리 및 보안 관련 ISV 에게 최고의 솔루션
- 삼성 SECUI, 대아 TI, 퓨처시스템, 경기도평생교육진흥원
- 홈페이지(**http://www.infiniflux.com**) 정식 버전 다운로드 가능

Enterprise Edition

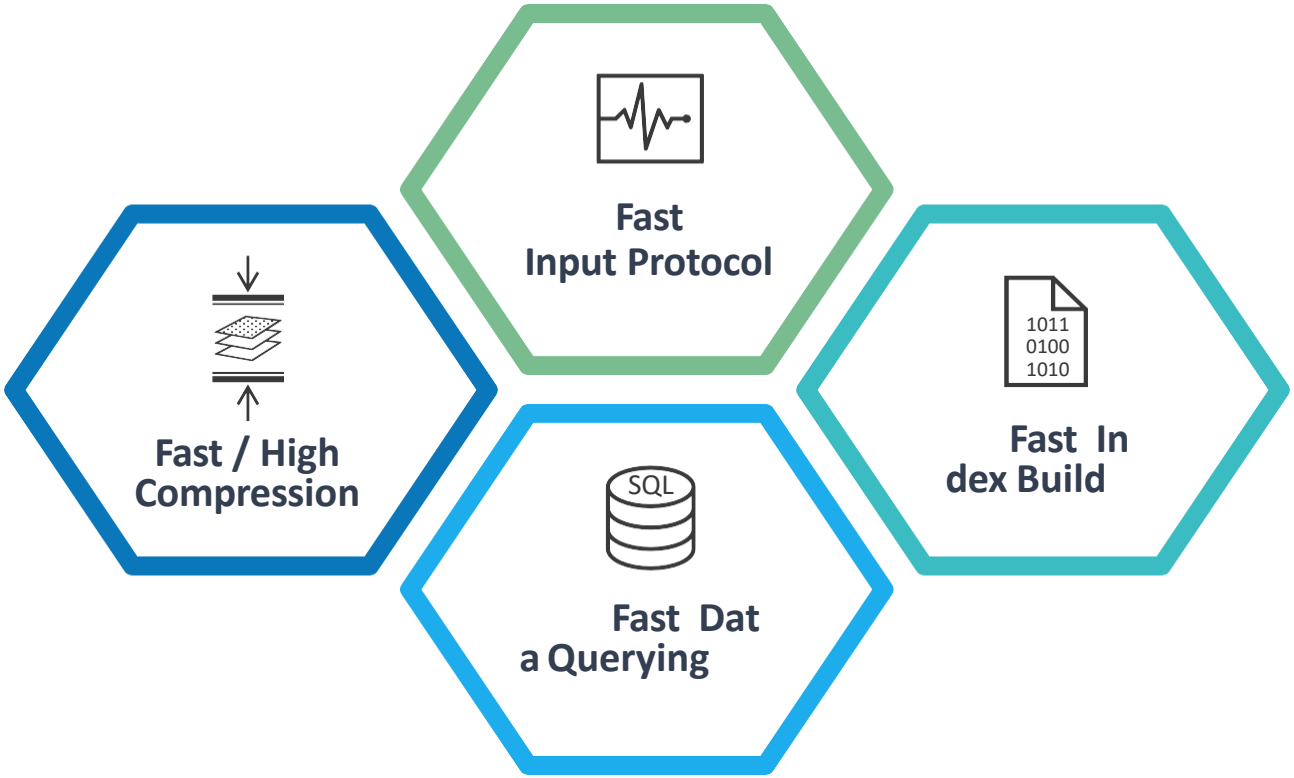
다중 노드 클러스터 확장 구성

- 단일 노드에서도 Scale-up 구성 가능
- 멀티 노드로 Scale-out 구성
- 최대 32 노드 까지 연결
- 선형적인 성능 증가
- 시계열 빅데이터 클라우드 시스템 구축 가능
- 2017년 상반기 정식 버전 출시 예정



5. 제품 특징

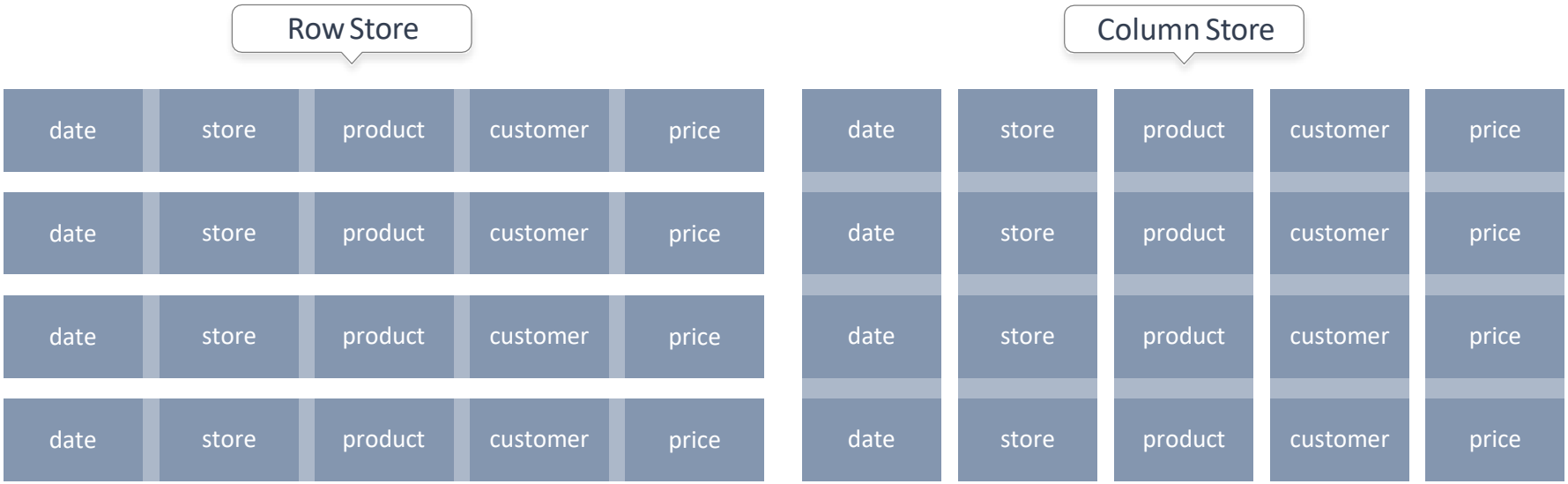
Fast Data Computing



Columnar Store

OLAP에 최적화된 Columnar Store

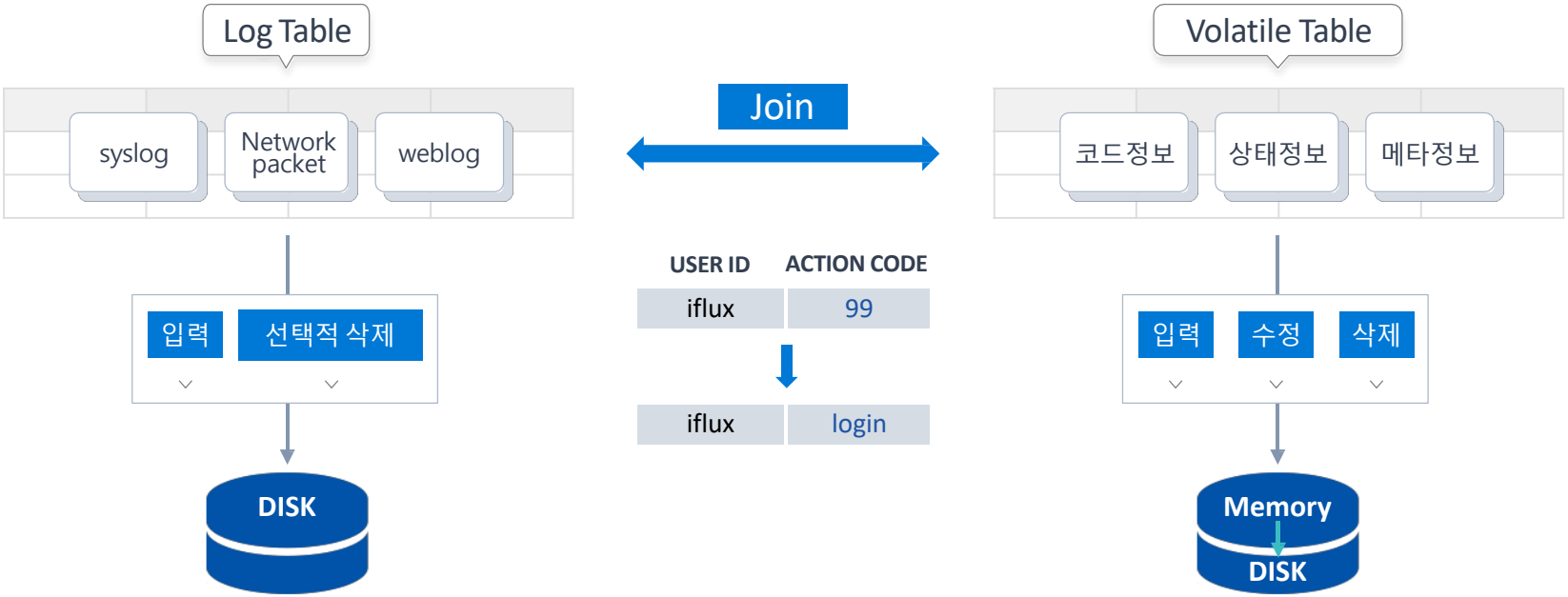
- 컬럼 단위로 저장, 컬럼의 값들을 서로 연속된 디스크 혹은 메모리 공간에 위치
- 서로 다른 레코드의 컬럼도 시스템에 큰 부하 없이 검색 가능
- 데이터 분석 성능은 ROW 기반에 비해 수십 배 빠름
- 데이터 압축 용이함



Log Table vs Volatile Table

로그 저장 테이블과 기준 정보 저장 테이블 제공

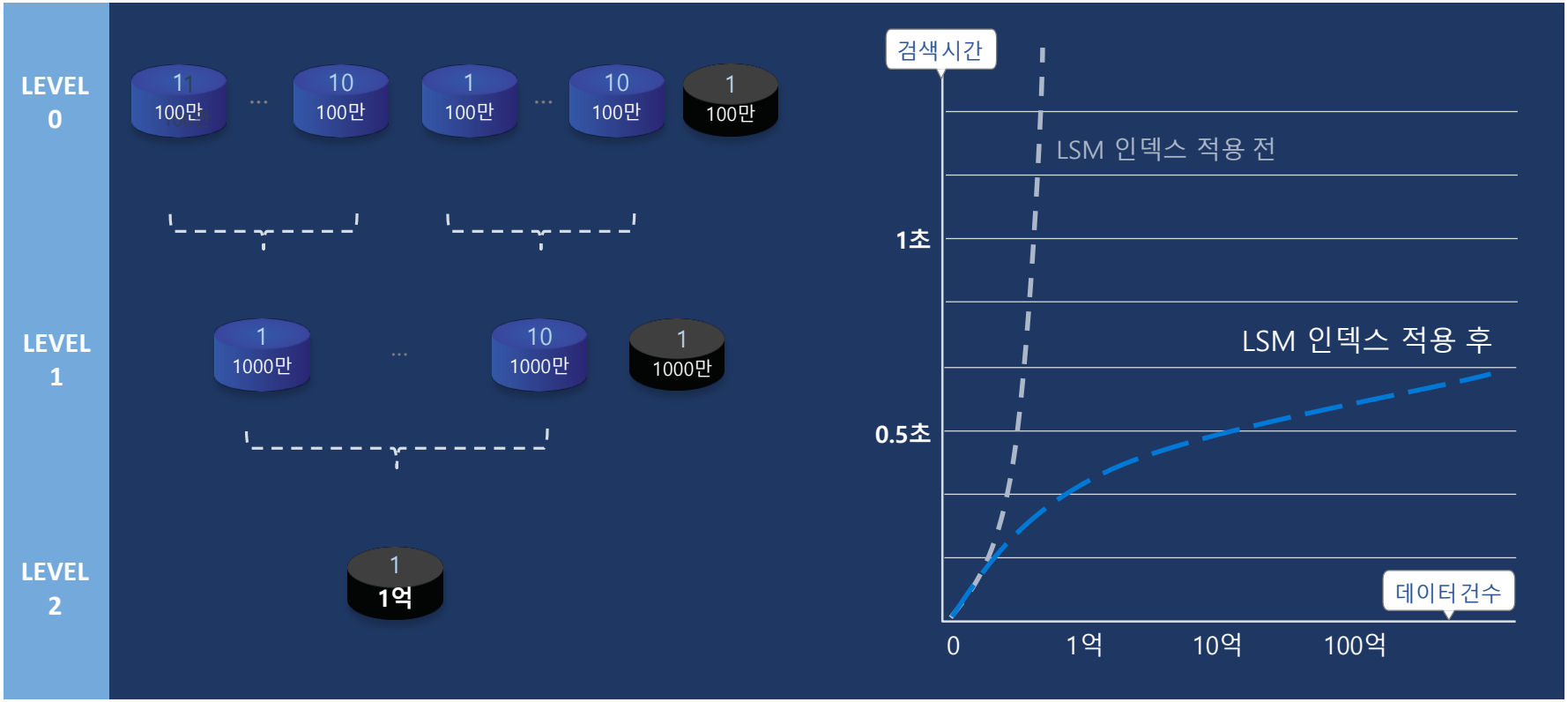
- Log Table은 초고속 데이터 입력과 부분 삭제 가능, 수정 기능 없음
- Volatile Table은 Primary Key 기반 실시간 입력, 수정, 삭제 가능
- 로그 테이블과의 조인(join)을 통한 데이터 분석 지원
- 메모리상에서 데이터를 처리하며 디스크에 저장 가능



LSM INDEX

대용량 데이터 처리에 적합한 인덱스

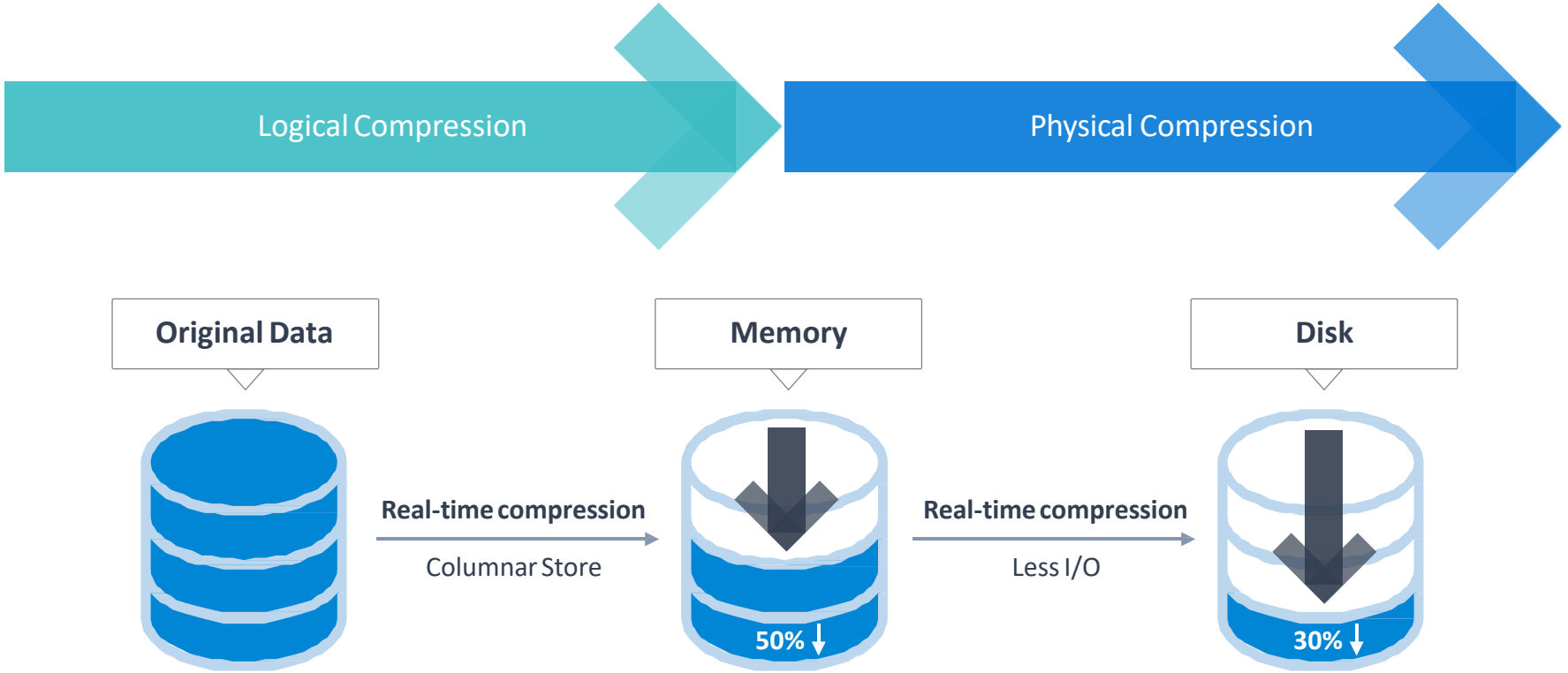
- **빠른 데이터 증가 환경에서 고속 검색 지원 기술**
- 로컬 인덱스와 글로벌 인덱스의 장점을 취해 만든 혁신적인 인덱스 기술
- 일반 서버 환경에서 10억 건 중 1건 검색시간 0.5초



Compression

고성능 데이터 압축으로 저장공간 절약

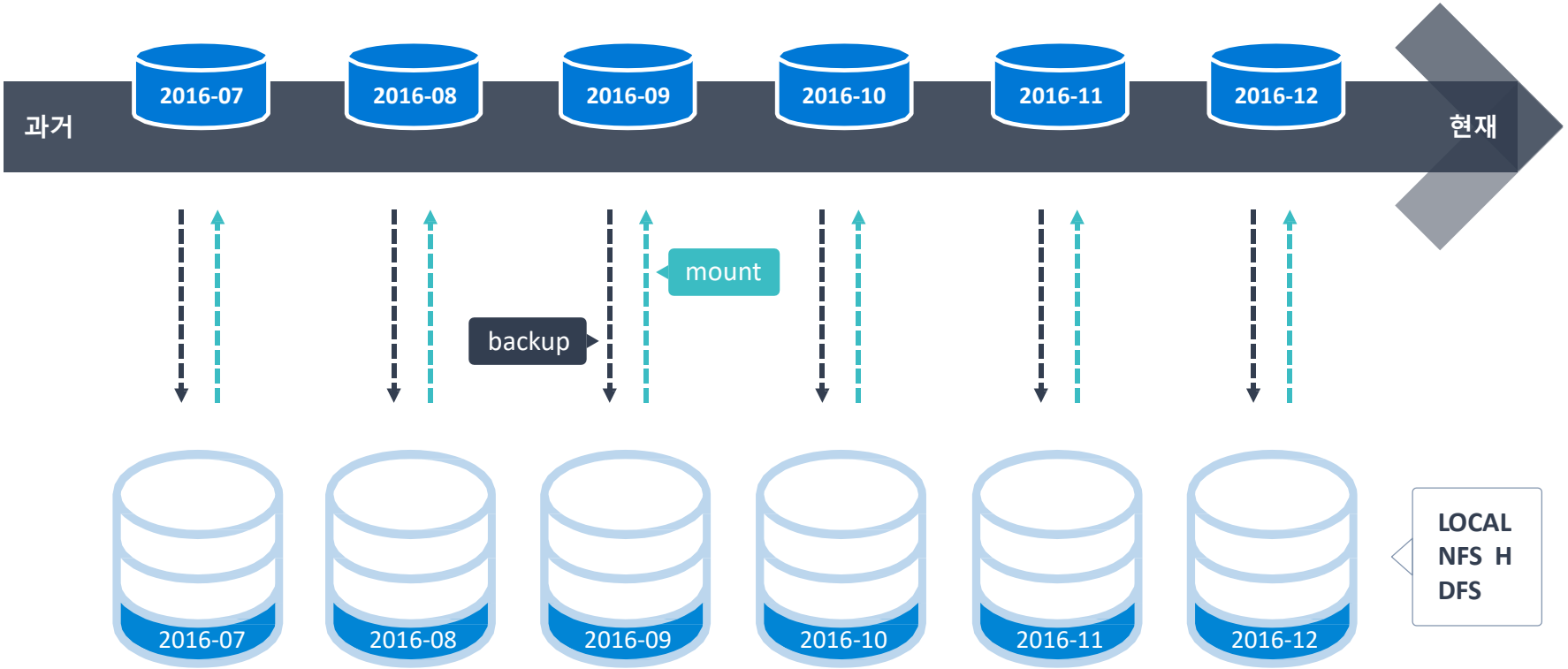
- 논리적, 물리적 2단계 데이터 압축 저장
- 원본 대비 수 배 ~ 수십 배 압축
- 시스템 부하 최소화 및 데이터 입력,검색 성능 향상



Backup & Mount

혁신적인 마운트 기능으로 빠른 백업 데이터 조회

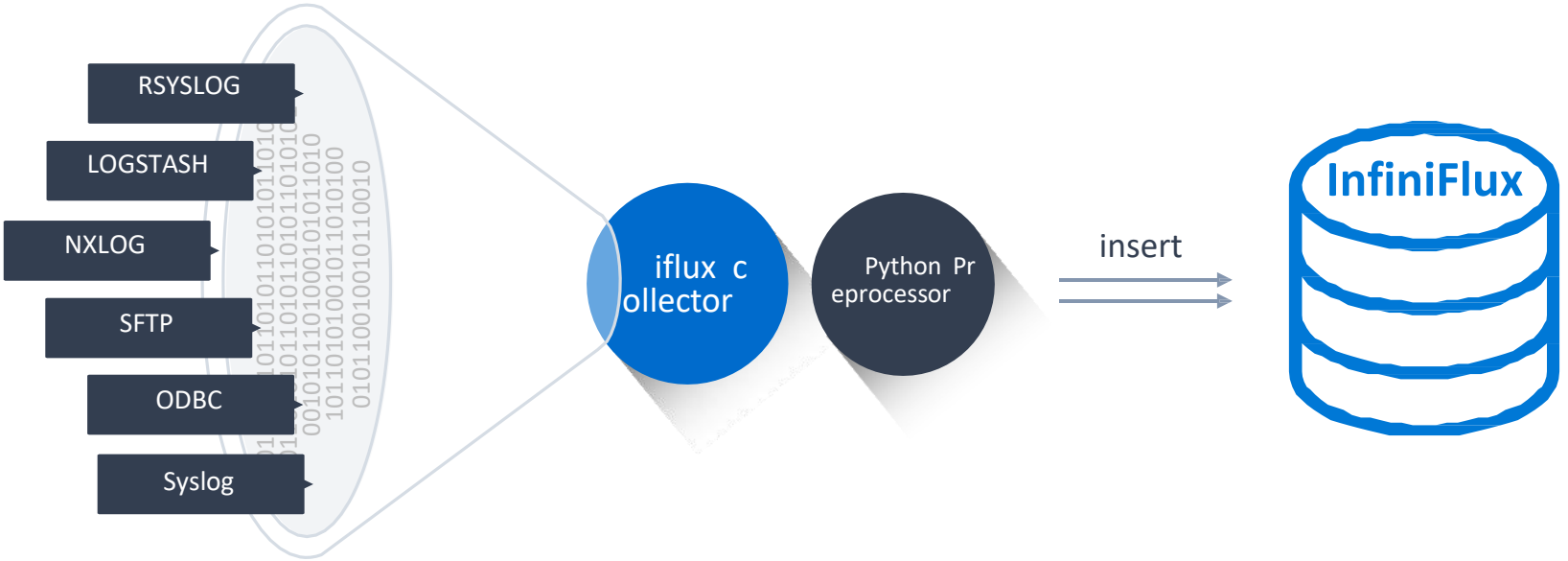
- **마운트(mount) : 백업된 데이터 정보를 데이터 로딩 없이 즉시 조회할 수 있는 기능**
- 데이터베이스 단위 및 테이블 단위 백업, 마운트 지원
- 로컬 디스크, NFS, HDFS 저장공간에 백업, 마운트 가능

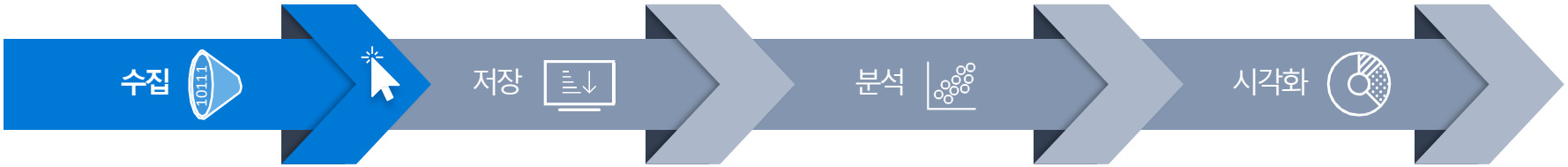


6. 제품 기능

로그 수집기

- 다양한 데이터 원천 소스 기반 로그 수집 기능 제공
- Python 언어 기반 실시간 데이터 전처리 변환 가능
- 오픈 소스 로그 수집기, SFTP, ODBC, Syslog 를 통한 로그 수집 지원



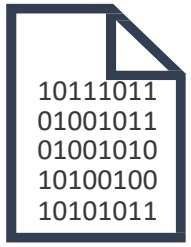


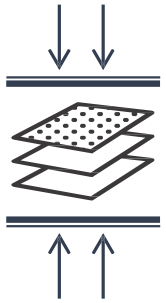
실시간 압축저장

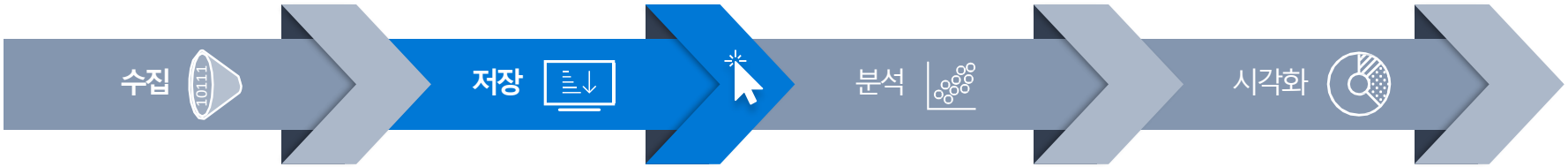
실시간 저장

실시간 인덱싱

실시간 압축







시계열 분석구문

표준 SQL

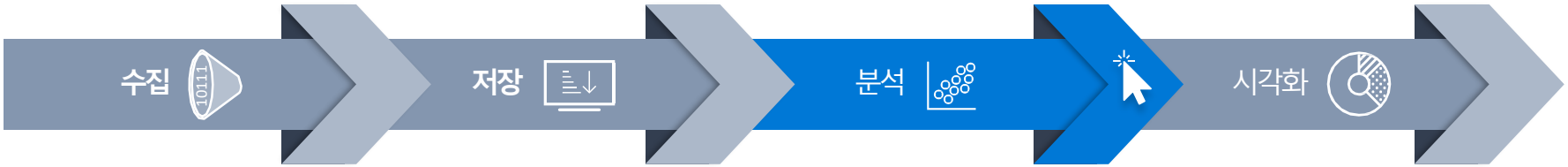
```
SELECT
    code,
    count(*) as count
FROM logtable  WH
ERE code = '200'
GROUP BY code
ORDER BY code DESC
LIMIT 10;
```

시계열 확장 구문

```
SELECT
    _arrival_time,
    code
FROM logtable
DURAT ION 10 minute;
```

확장 데이터 타입

- IP Address Type
  - IPv4
  - IPv6
- Unsigned Type
  - USHORT
  - UINTEGER
  - ULONG
- Binary Type
  - BINARY

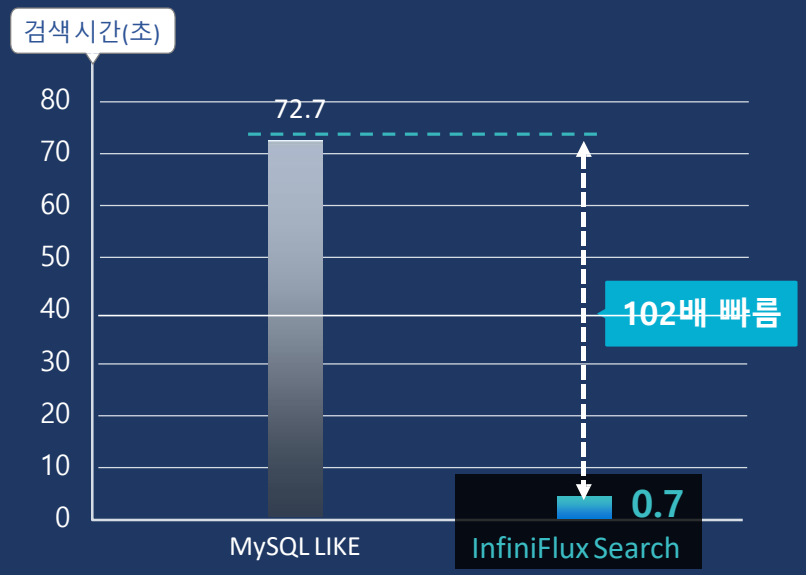


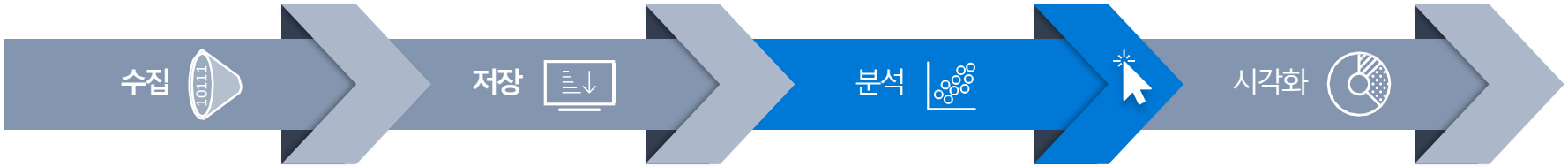
Full Text Search

- RDBMS의 LIKE 와 유사한 'SEARCH' 구문 제공
- KEYWORD INDEX 를 이용하여 빠른 검색 가능
- 영문 뿐만 아니라 UTF-8 형식의 한글, 중국어, 일본어 문자 검색도 가능

SEARCH

```
SELECT message  FRO
M textsearch  WHERE
message  SEARCH 'ER
ROR 500';
```





IP Address Type

- 제품 엔진 레벨에서 IPv4, IPv6 데이터 타입 지원
- Network Mask 형식 지원 및 편리한 연산자와 함수 제공
- 간단하고 빠르게 IP 주소에 대한 저장과 검색 가능

CREATE

```
CREATE TABLE
addrtable
(
   srcip ipv4,
   dstip ipv6
);
```

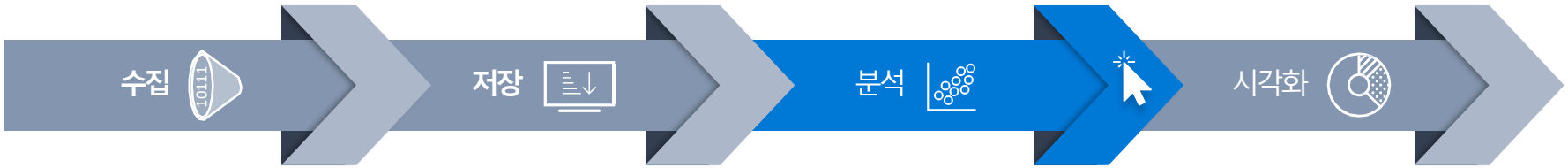
INSERT

```
INSERT INTO
addrtable
VALUES
(
  '127.0.0.1',
  '::127.0.0.1'
);
```

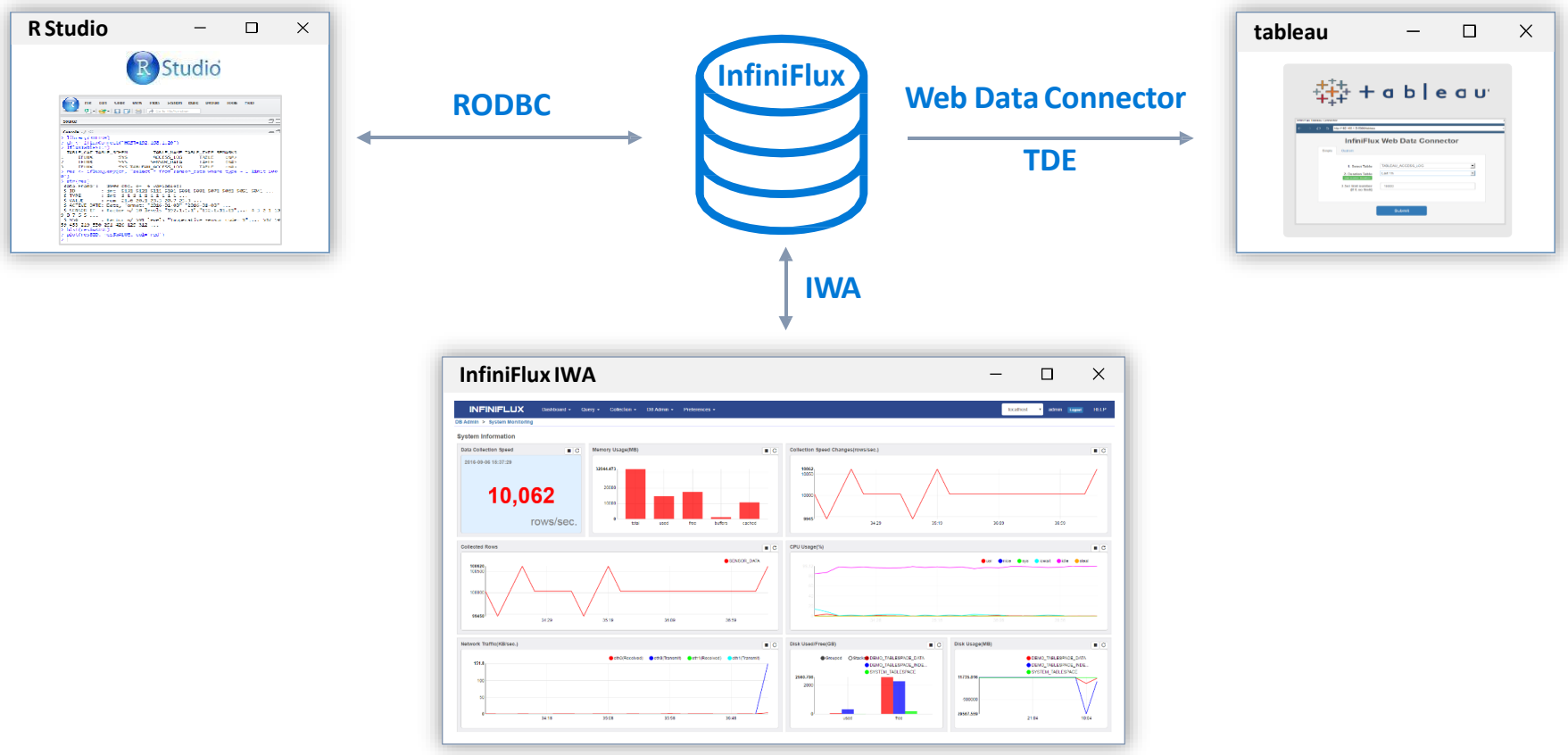
SELECT

```
SELECT srcip
FROM addrtable
WHERE srcip = '192.168.0.*';

SELECT srcip  FR
OM addrtable
WHERE srcip
          CONTAINED '19
          2.168.0.0/16';
```

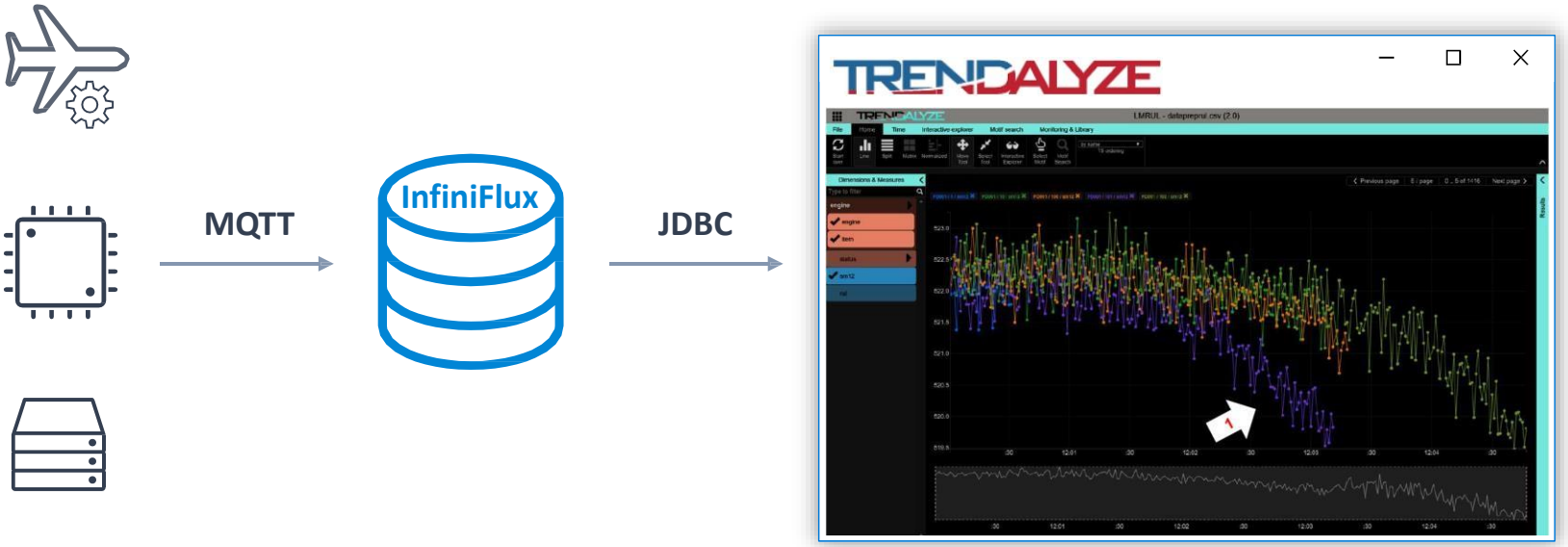


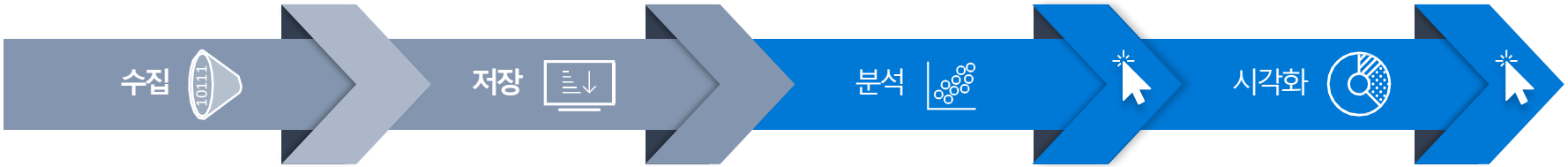
연동 및 시각화


R Studio
RODBC
InfiniFlux
Web Data Connector
TDE
tableau
IWA
InfiniFlux IWA
10,062
rows/sec.
수집
저장
분석
시각화

Trendalyze 연동

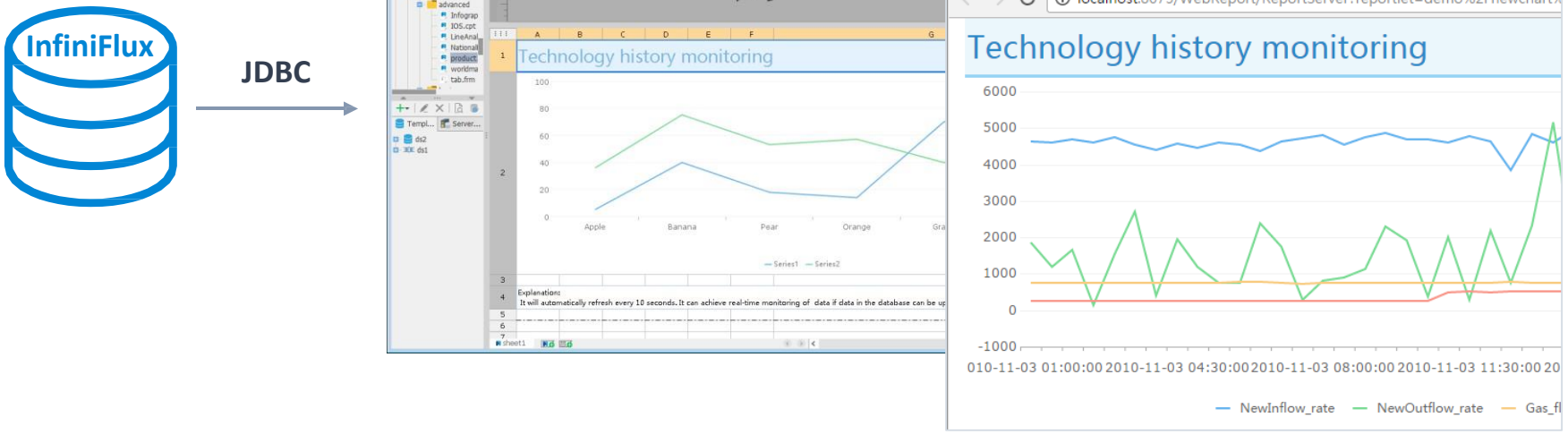
- Unlocking the value of time patterns
- Motif analysis, library and monitoring
- Lockheed Martin Case : 97% Preventive Failure Detection without modeling

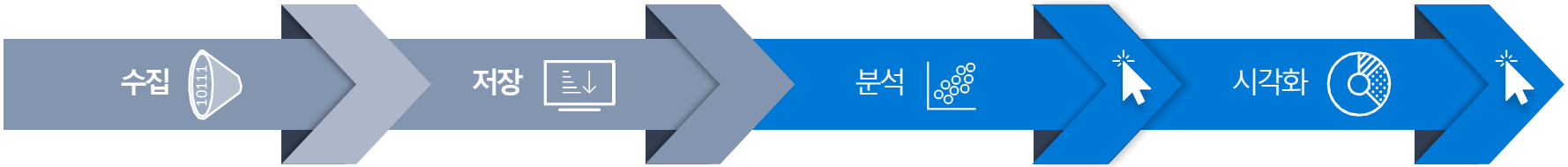




FineReport 연동

- Enterprise-oriented business intelligence and reporting tool with ‘No-code’ concept
- Excel-analogous and ‘Excel + Binding Data Column’ operation interface
- 4500+ cooperative clients / 1,000,000+ users





7. 성능 비교

성능 비교 환경

1억건, 13GB의 데이터, 각 제품의 데이터 입력 및 분석 성능 측정

- CentOS 6.6
- Intel(R) Core(TM) i7-4790 CPU @3.60GHz(4 core)
- 32GB memory
- SATA DISK

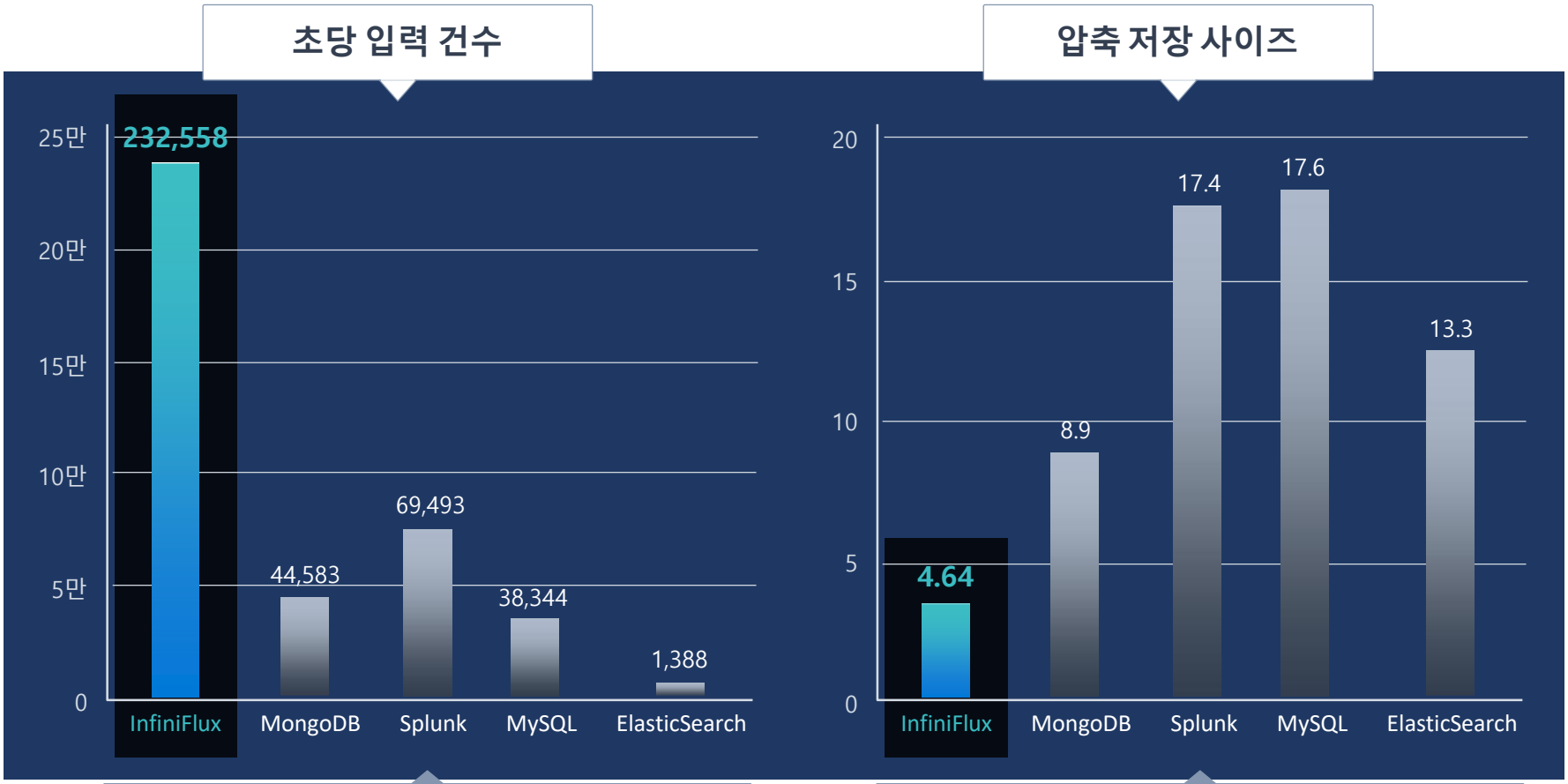
- InfiniFlux 3.1.1
- MySQL 5.2.12 MyISAM
- Splunk 6.4.0
- Elasticsearch 2.3.4
- MongoDB 3.2.6

DATA

필드	로그 생성시간	출발지 ip	출발지 port	도착지 ip	도착지 port	프로토콜 타입	로그 텍스트	상태 코드	데이터 크기
필드명	arrivaltime	srcip	srcport	dstip	dstport	protocol	eventlog	eventcode	eventsize
필드타입	datetime	ipv4	integer	ipv4	integer	short	varchar (1024)	short	long

성능 비교 결과

뛰어난 데이터 입력과 압축 성능 ⇉ **InfiniFlux**



데이터 입력 시간과 인덱스 생성 시간을 종합하여 계산함

InfiniFlux는 원본 크기보다 64.3%압축됨.(4.64GB/13GB)

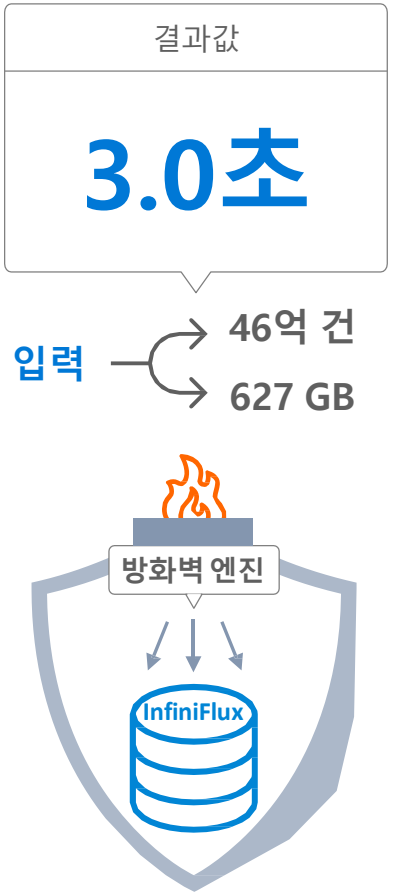
성능 비교 상세

		InfiniFlux 3.1.1	MongoDB 3.2.6	Splunk 6.4.0	MySQL 5.7.12	ElasticSearch 2.3.4
Data Loading (sec)		**430 (00:07:10)**	2243 (00:37:23)	1439 (00:23:59)	2608 (00:43:28)	72025 (20:00:25)
Inserted csv size (GB)				**13G**		
Data size (GB)		**4.64**	8.9	17.4	17.6	13.3
Compression ratio (%)		**64.3%**	31.5%	Uncompressed (133.8%)	Uncompressed (135.4%)	Uncompressed (102.3%)
Data search (sec)	Text search (33.9M)	**3.75**	167.7	394.82	64.84	7.06
Statistic (sec)	IP search (2.66M)	**0.81**	83.05	86.17	0.97	7.00
	Time search (268K)	**0.25**	78.03	2.66	0.34	4.37
	Count	**8.14**	104.25	344.34	63.45	3.11
	Sum	**12.1**	104.97	392.32	63.74	4.17
	Average	**12.5**	109.7	391.4	63.54	4.93
	Complex query	**5.35**	10.136	87.63	125.22	8.31

Firewal data 분석 사례

1개월 치 Firewall rule top 10의 일자별, rule별 건수 조회

```
SELECT TO_CHAR(dttime1,'YYYY-MM-DD') dt, policyid, COUNT(*)
FROM FIREWALL_DATA
WHERE policyid IN ( SELECT policyid
            FROM ( SELECT policyid, COUNT(policyid)
                FROM FIREWALL_DATA
                WHERE dttime1 >= TO_DATE('2016-11-01 00:00:00') AND
                    dttime1 <  TO_DATE('2016-12-01 00:00:00')
                GROUP BY policyid
                ORDER BY 2 DESC
                LIMIT 10)
            ) AND
    dttime1 >= TO_DATE('2016-11-01 00:00:00') AND
    dttime1 <  TO_DATE('2016-12-01 00:00:00')
GROUP BY dt, policyid
ORDER BY 2, 1;
```



(총 46억 건/ 627GB 입력, 테스트 환경 : 8 core CPU, 16GB Memory, SSD)

8. 고객 사례

적용 분야

대용량 데이터를 실시간 저장하고 빠른 분석이 필요한 모든 산업에 적용 가능

- 보안 로그 저장, 관리 시스템
- 네트워크 패킷 분석 시스템
- UTM 장비 탑재 로그 분석

- 제조 공정 수율 관리(MES)
- 제조 공정 위험 관리 및 방지
- 삼성전자 반도체 공장 MES

- 발전기 센서 감지 및 위험 경고
- 발전소 원격 관리 및 모니터링
- 스마트 팩토리

- 게임 로그 저장 및 관리
- 게임 아이템 추적
- 실시간 게임 사용자 행동 분석

- FDS(Fraud Detection System)
- 사용자 거래 내역 분석
- 실시간 서비스 로그 저장 관리

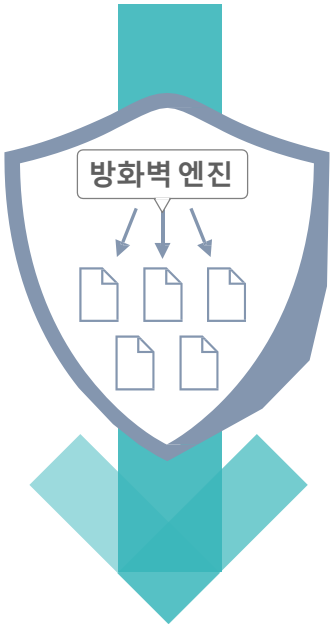
- 중계기, 장비 로그 저장 관리
- 위치 기반 광고 전송 시스템
- 맞춤형 서비스 제안 시스템

시큐아이 방화벽

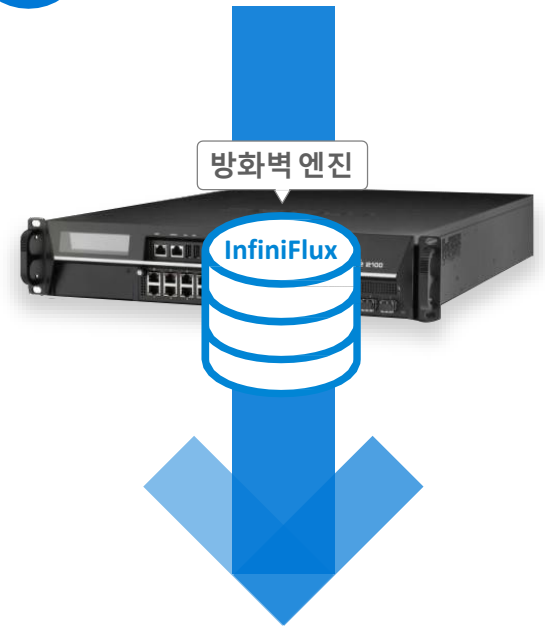
차세대 방화벽 장비 내에 InfiniFlux를 탑재하여 로그 분석

- 기존에는 방화벽 로그를 파일 형태로 저장하고 활용함.
- InfiniFlux를 이용한 대용량 로그 처리 및 로그 관리, 빠른 검색 및 리포팅 지원

로그를 **파일** 형태로 저장

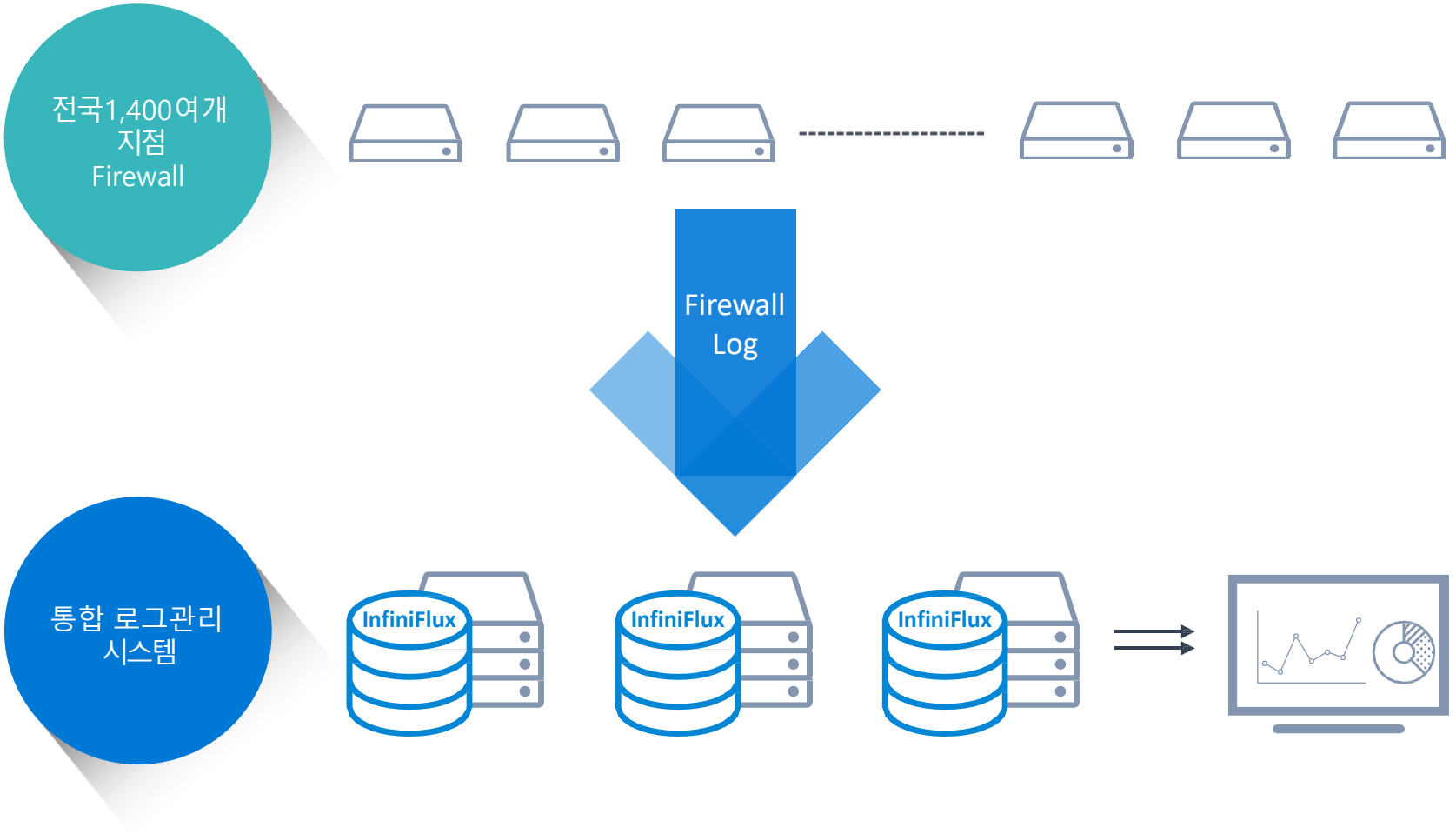


로그를 **InfiniFlux** 로 저장



퓨쳐시스템(금융권)

일일 45억 건, 1TB Firewall Log를 실시간으로 처리하기 위하여 InfiniFlux 도입



References

SECUI	경기도평생교육진흥원	Humanizing Genomics macrogen
방화벽 로그분석	G-MOOC 학습분석시스템	유전체(vcf) 분석 시스템
KOREA POST	ETRI 한국전자통신연구원 Electronics and Telecommunications Research Institute	KORAIL 한국철도공사
WEEDS 개인정보침해탐지	Cyber BlackBox Project	대아티아이 철도안전관제시스템

9. 회사 소개

INFINIFLUX

초고속 시계열 DBMS 개발 전문 기업

JL Infra (주) 제이엘 인프라 Co.,Ltd

For IoT & BigData
INFINIFLUX
THE WORLD'S FASTEST TIME SERIES DBMS

감사합니다

Email : info@JLinfra.co.kr

Tel : 1522-9217

Appendix

유틸리티

ifluxadmin

- InfiniFlux 서버의 구동, 상태파악, 종료
- 데이터베이스 생성, 복구, 삭제

ifluxsql

- Console 화면을 통해 SQL query 수행, 확인
- 편리한 SHOW 명령어 제공

ifluxloader

- InfiniFlux 데이터를 export & import
- csv format 지원

InfiniFlux Web Analytics

- Python flask 기반 웹 응용 프로그램
- 웹 UI 로 InfiniFlux 이용 및 dashboard 작성

데이터 입력

Application

- 어플리케이션을 개발하여 DB에 입력
- SDK(CLI, ODBC, JDBC)를 사용하여 개발

ifluxcollector

- ifluxcollector 를 이용 하여 Data 입력
- 로그 발생시 실시간 수집

SQL syntax

- SQL Query 를 사용하여 Data를 입력
- INSERT INTO, LOAD DATA 구문 사용

ifluxloader

- ifluxloader 를 이용하여 Data를 입력
- 대용량 파일 및 DB migration 용도로 사용

조회 및 삭제

Nano second timestamp 자동 저장

- 데이터 입력 순간 숨은 칼럼(_arrival_time)에 nano second 를 자동으로 저장
- 조회시 가장 최근에 입력된 데이터부터 시간의 역순으로 출력

DURATION 키워드 제공

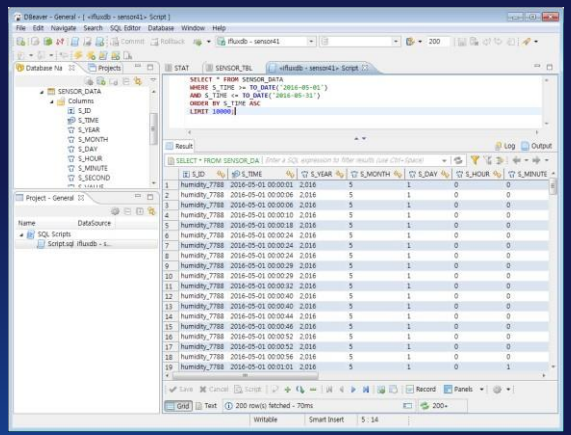
- 데이터 검색할 때 시간 범위를 쉽게 지정하기 위해서 제공되는 키워드
- 현재 시각 기준 10분 전까지 데이터의 합계를 구하는 경우
  : SELECT SUM(traffic) FROM T1 DURATION 10 minute;
- 현재 시점에서 한시간 이전 부터 10분간 데이터의 합계를 구하는 경우
  : SELECT SUM(traffic) FROM T1 DURATION 10 minute BEFORE 1 hour;

선택적 삭제 지원

- 일정 크기 이하의 데이터 유지를 위한 기능
- 지금부터 1일 동안의 데이터를 제외하고 모두 삭제하는 경우
  : DELETE FROM T1 EXCEPT 1 day;
- 2015년 6월 1일 이전의 데이터를 모두 삭제하는 경우
  : DELETE FROM T1 BEFORE TO_DATE('2016-08-01', 'YYYY-MM-DD');

연동 툴

DBeaver

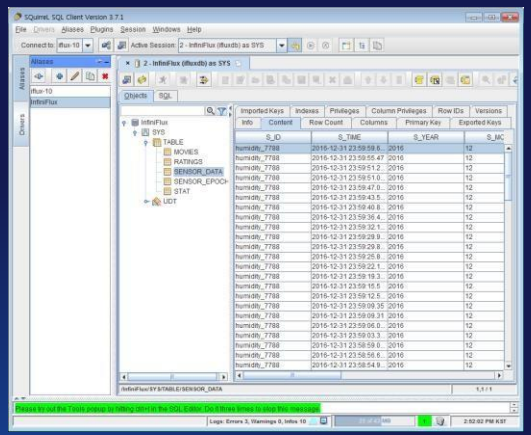


http://dbeaver.jkiss.org/

JDBC

JDBC

Squirrel SQL



http://squirrel-sql.sourceforge.net/