

MACHBASE

“ The Database for Things™ ”

The Fastest Time Series DBMS
for Industrial IoT Data



(주)마크베이스

이재훈 본부장

Table of Contents

The Fastest Time Series DBMS **Machbase**



- Global IoT Trend
- Time Series Database
- Industrial IoT
- Machbase for IIoT
- Machbase Use Cases
- Summary

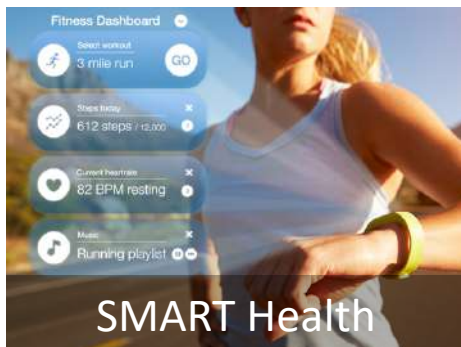
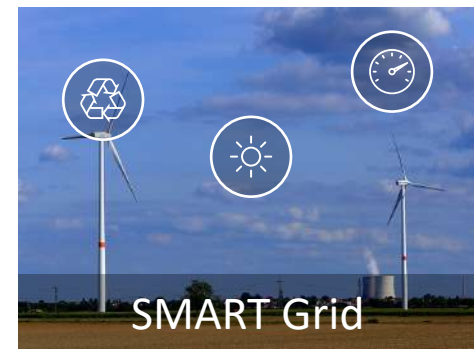
Global IoT Trend



Smart-X 시대의 도래

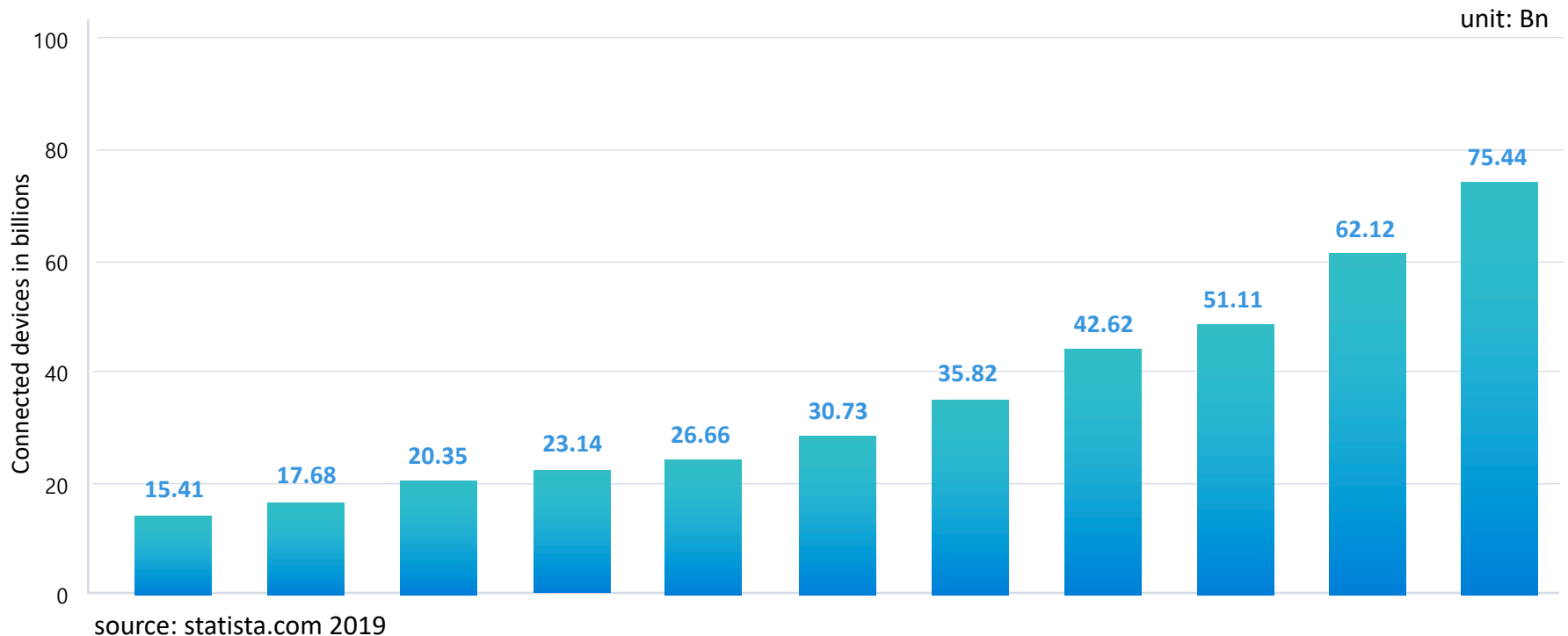


SMART X



IoT Device의 폭증

Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025



Strategic Technology Trend

* Source : gartner

Intelligent



- Applied AI & Advanced ML

Intelligence Things

- Intelligence Apps

- AI Foundations

Intelligent Things

- Intelligent Apps and Analytics

2019

- AI-Driven Development

Autonomous Things

- Augmented Analytics

Digital



- Digital Twins
- Virtual & Augmented Reality
- Blockchains and Distrinuted Ledgers

- Digital Twins

Cloud to the Edge

- Immersive Experience
- Conversational Platform

- Digital Twin

Empowered Edge

- Immersive Experience

Mesh



- Conversational Systems
- Mesh app and Service Architecture
- Digital Technology Platforms
- Adaptive Security Architecture

- Blockchain
- Event-Driven
- Continuous Adaptive Risk and Trust

- Blockchain

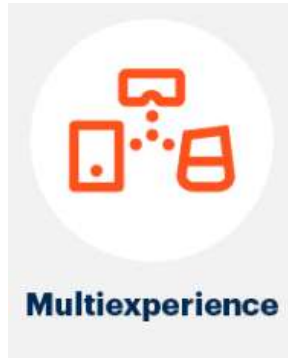
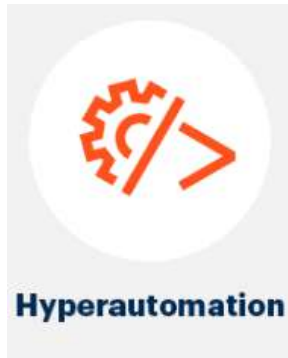
Smart Spaces

- Privacy and Ethics
- Quantum Computing

IoT 데이터 폭증 → Edge Computing → 초연결 지능화

Strategic Technology Trend 2020

● People-centric



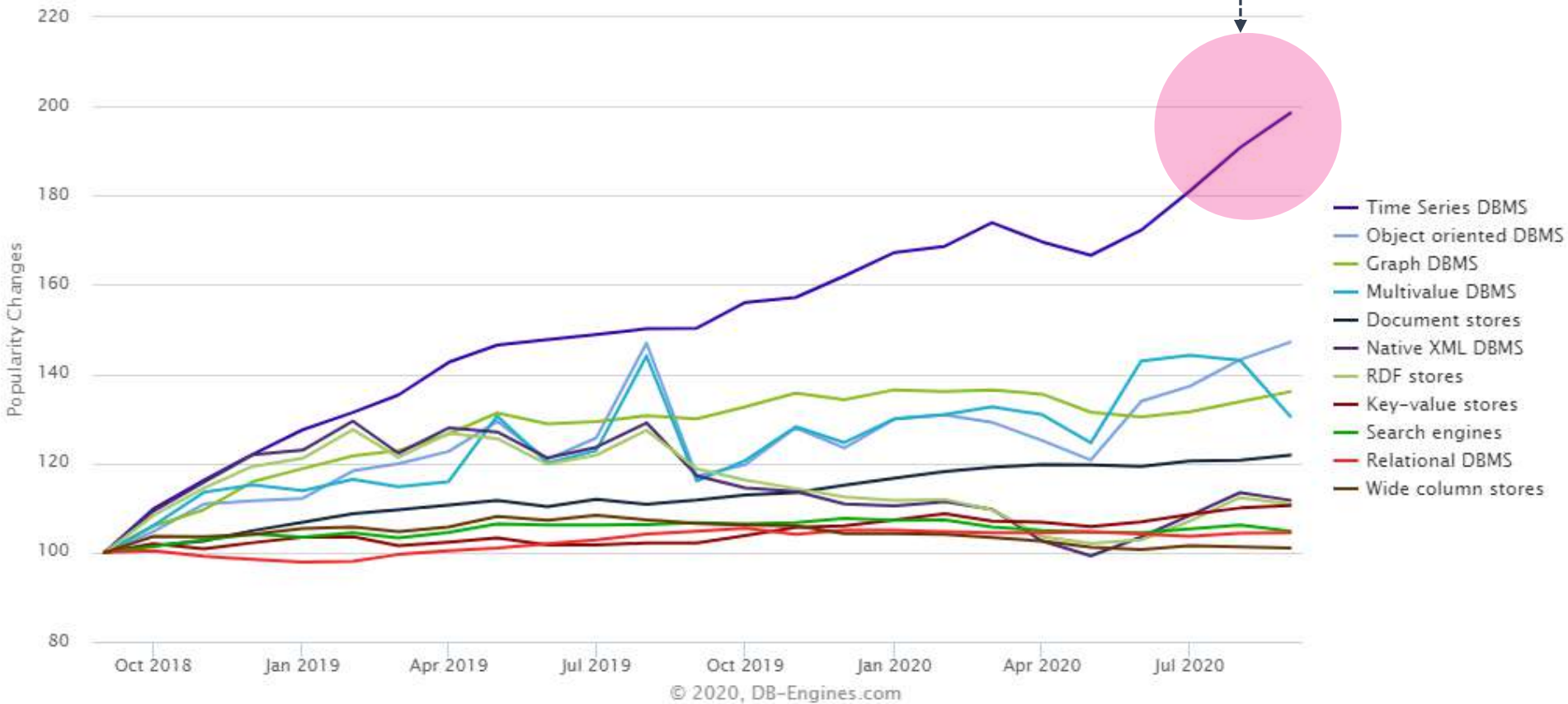
● Smart spaces



2019 Gartner Inc.

최근 24개월 DBMS Trend

Time Series DBMS



Time Series Database (TSDB)



Time Series Data ?



Time Series Data 특징

데이터 특성 (정형/반정형)

시간 흐름에 따라 발생하는
로그, 이벤트 정보

데이터 입력 위주

데이터 업데이트 없고
과거 데이터 순 삭제

ID 상태 정보 포함

해당 데이터 소스의 ID 및
상태 정보 포함

시간 데이터

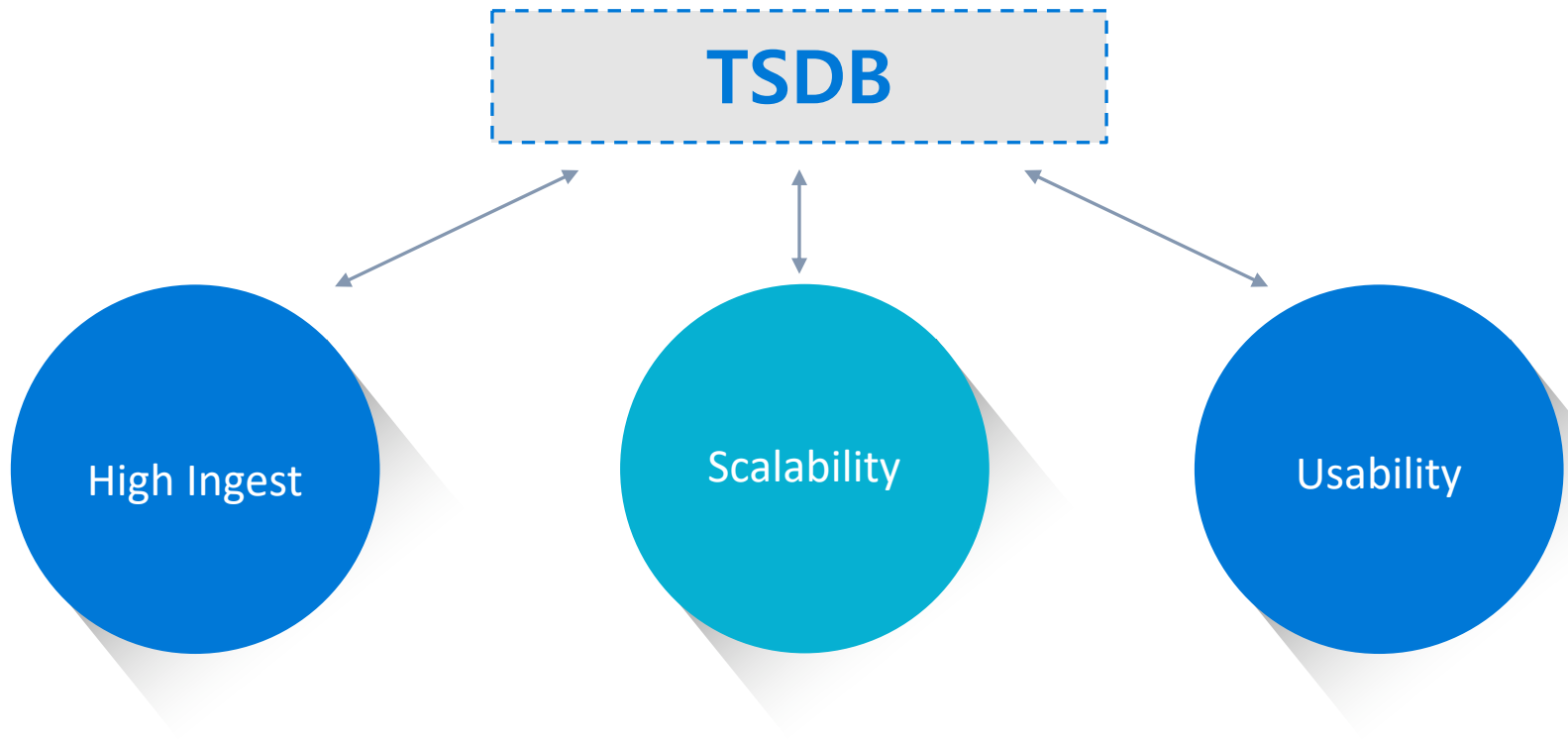
시간을 기준으로
각종 통계 분석 수행

매우 빠른 생성 속도

동일 패턴 지속적으로
빠르게 생성

Time Series DBMS(TSDB)

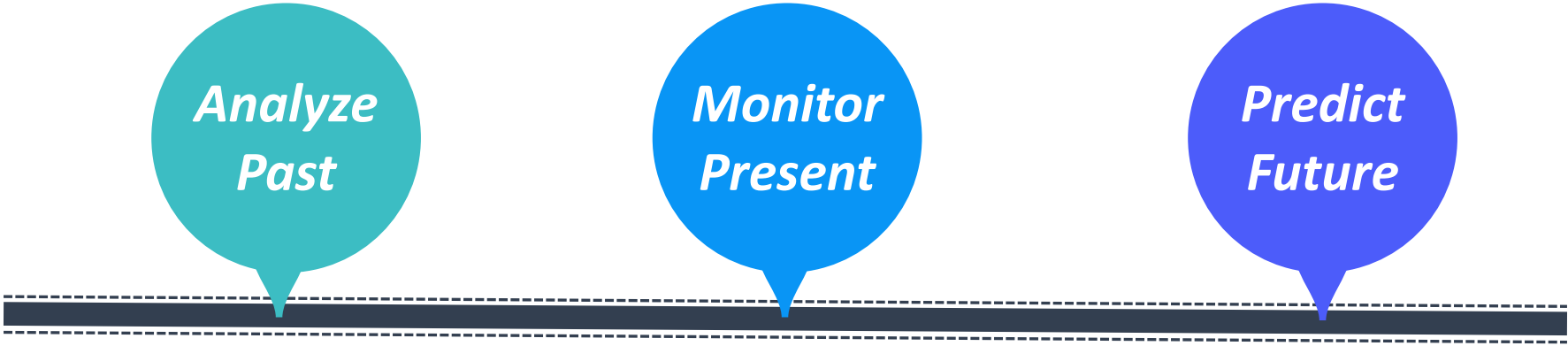
*A software system that is **optimized for storing and serving time series** through associated pairs of time(s) and value(s) - wikipedia*



Time Series DB 종류



Time Series DB 용도



*Analyze
Past*

*Monitor
Present*

*Predict
Future*

Historical Analysis
Check a specific
data point

Real-time Trend monitoring
Anomaly detection
Alarm & Notification

Predictive maintenance
Reducing downtime
Machine learning

Industrial IoT (IIoT)



OT와 IT의 융합

Visualization & Analytics

Tableau

QlikView

Spotfire

SAS

MSFT BI

ECMiner

Custom Applications

Enterprise Data Warehouse / Data Mart / Data Lake

ORACLE
EXADATA

Microsoft
SQL Server

TERADATA

SAP HANA

Convergence!

hadoop

Operational Intelligence

RTDB

PI System

Historian

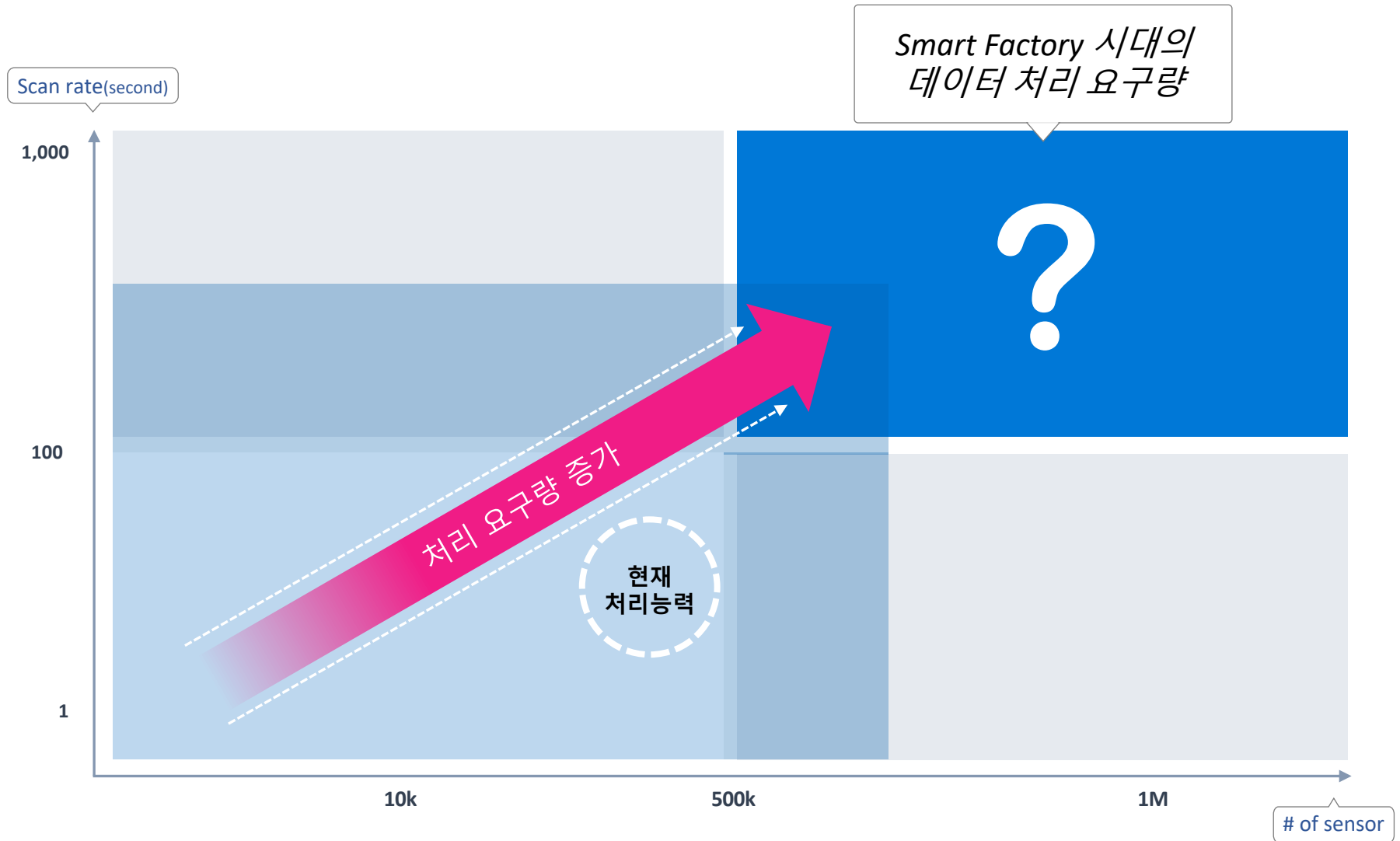
PHD

dataPARC

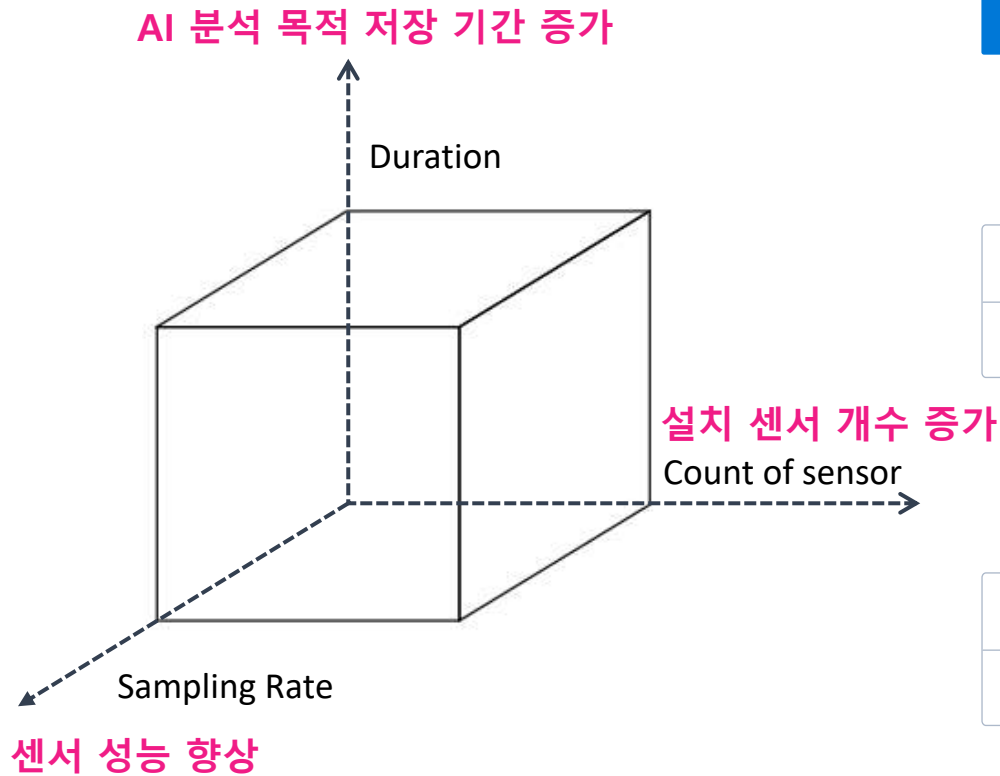
Predix

InfoPlus

Industrial IoT 데이터 추이



IIoT 데이터의 증가



1,000 개 * 100 ms * 40 Bytes

9억 건

32 GB



- day -

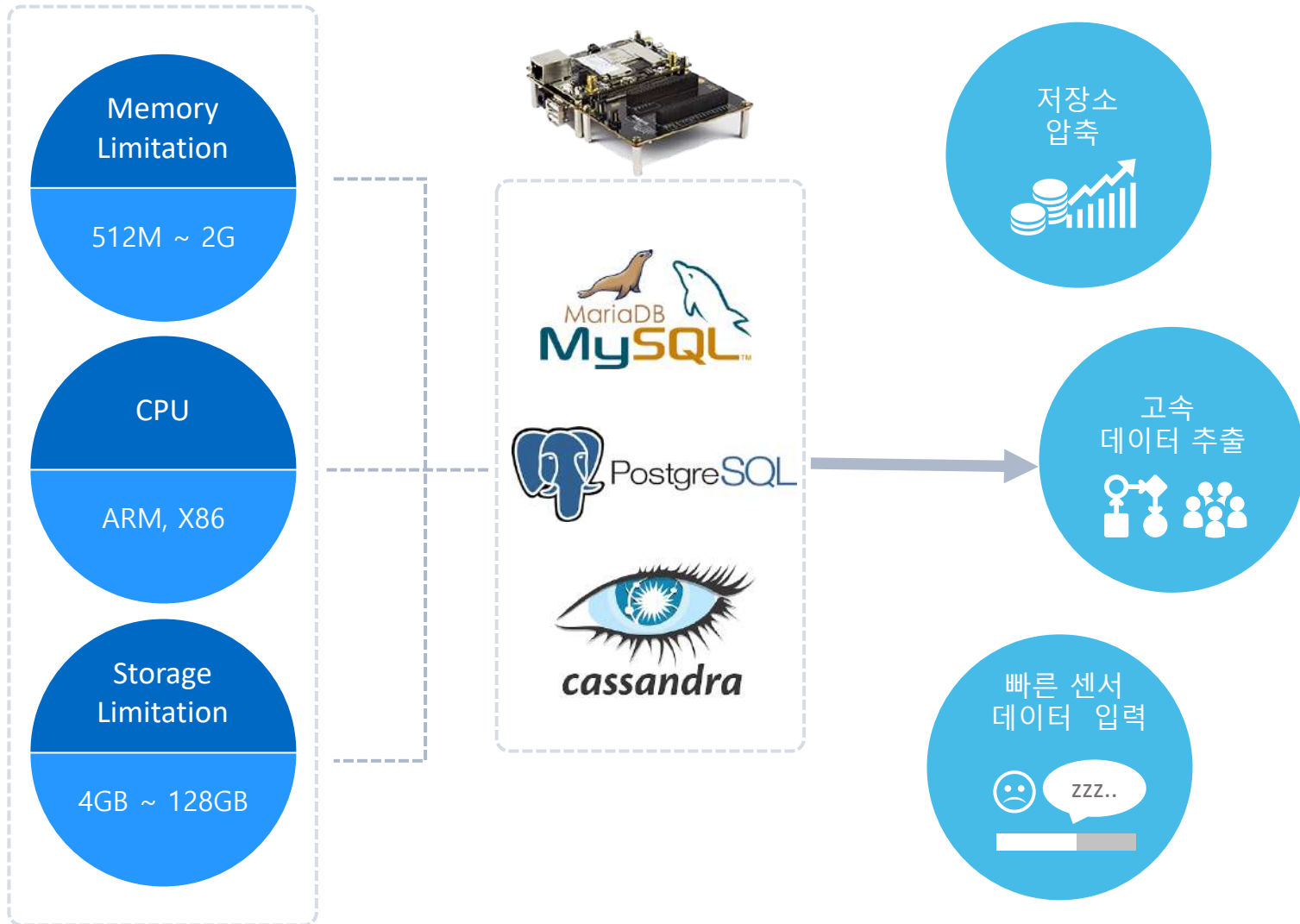
3,154억 건

11 TB

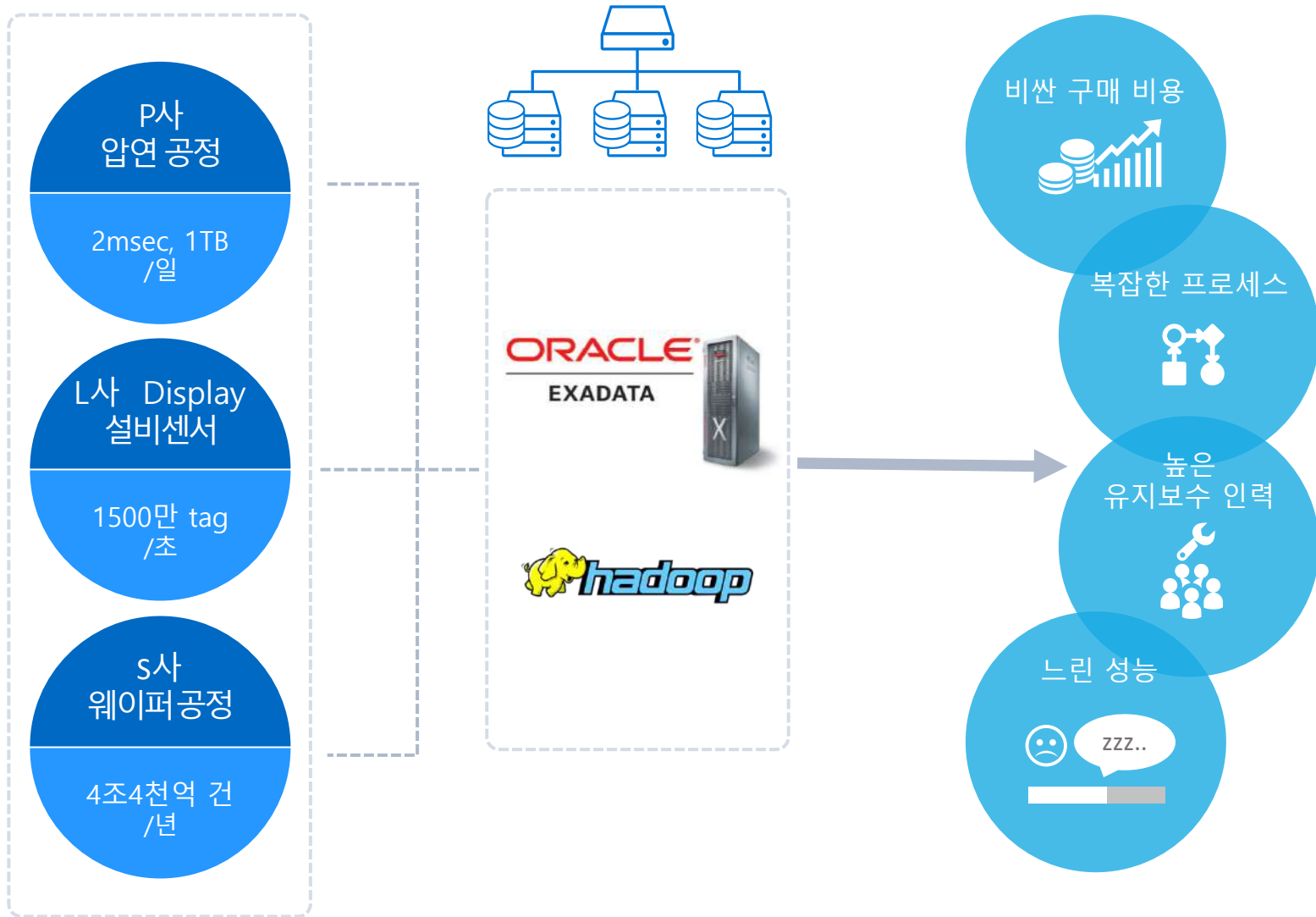


- year -

현장의 요구사항 (Edge Level)

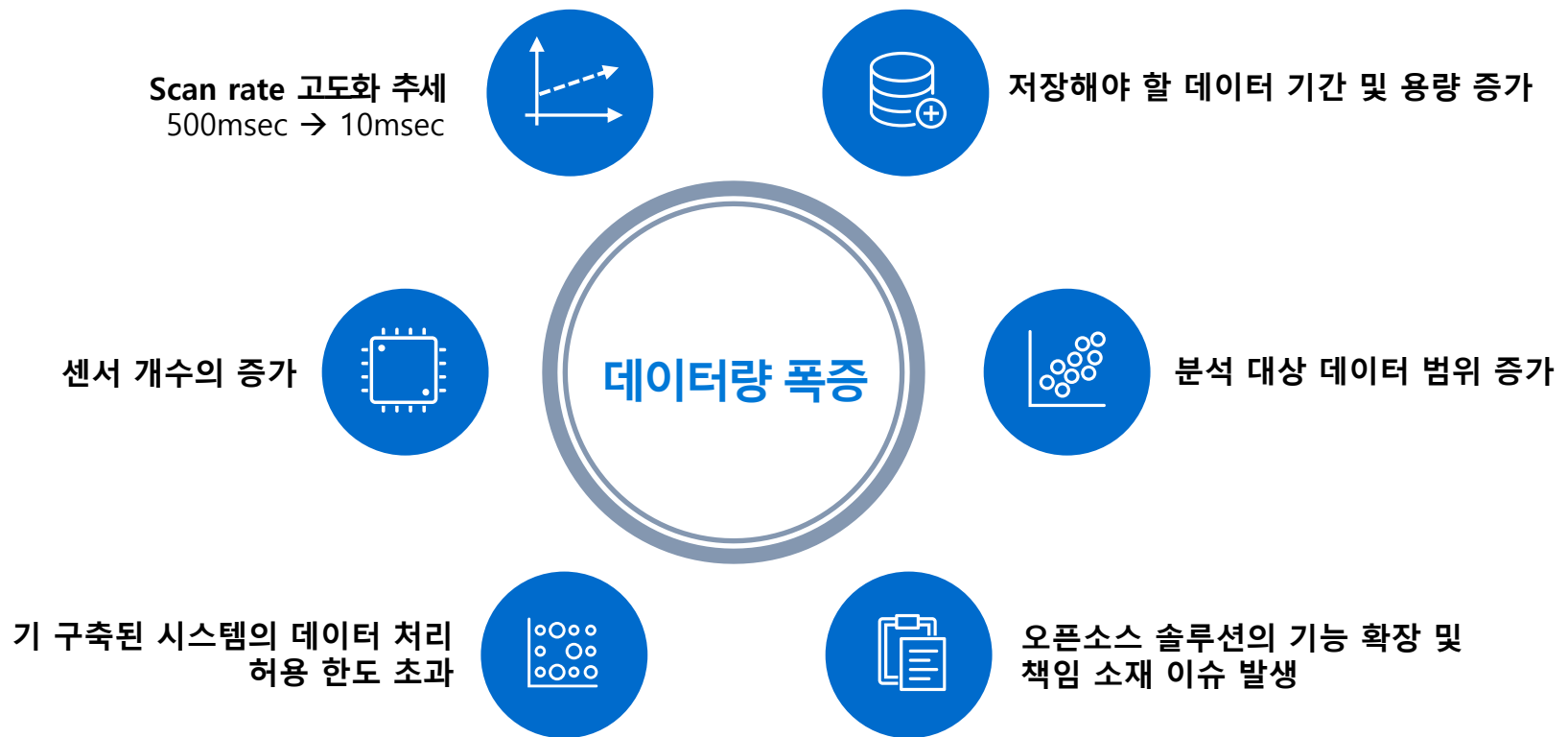


현장의 요구사항 (Enterprise Level)



IIoT 데이터 처리 환경 변화

제조 공정 내 Thing 연결 환경 및 데이터 특성에 맞는 저장 및 분석 SW 필요



제품 요구 사항

“IIoT의 특성을 고려한 DBMS 기본 구조 필요”

고속 입력 : 초당 수십만 건 이상의 입력 가능 아키텍처

실시간 압축 : 저장소 효율성 극대화를 위한 압축 알고리즘

센서 스키마 : IIoT의 센서를 대표하는 저장 구조 지원

고속 시계열 질의 : 시간 범위에 대한 빠른 검색 성능

실시간 통계 : 실시간 통계 기능 지원

Scale-out & High Availability : 대규모 센서 빅데이터 지원

Embedded Architecture : 에지 컴퓨팅(Edge Computing)의 필수 요소

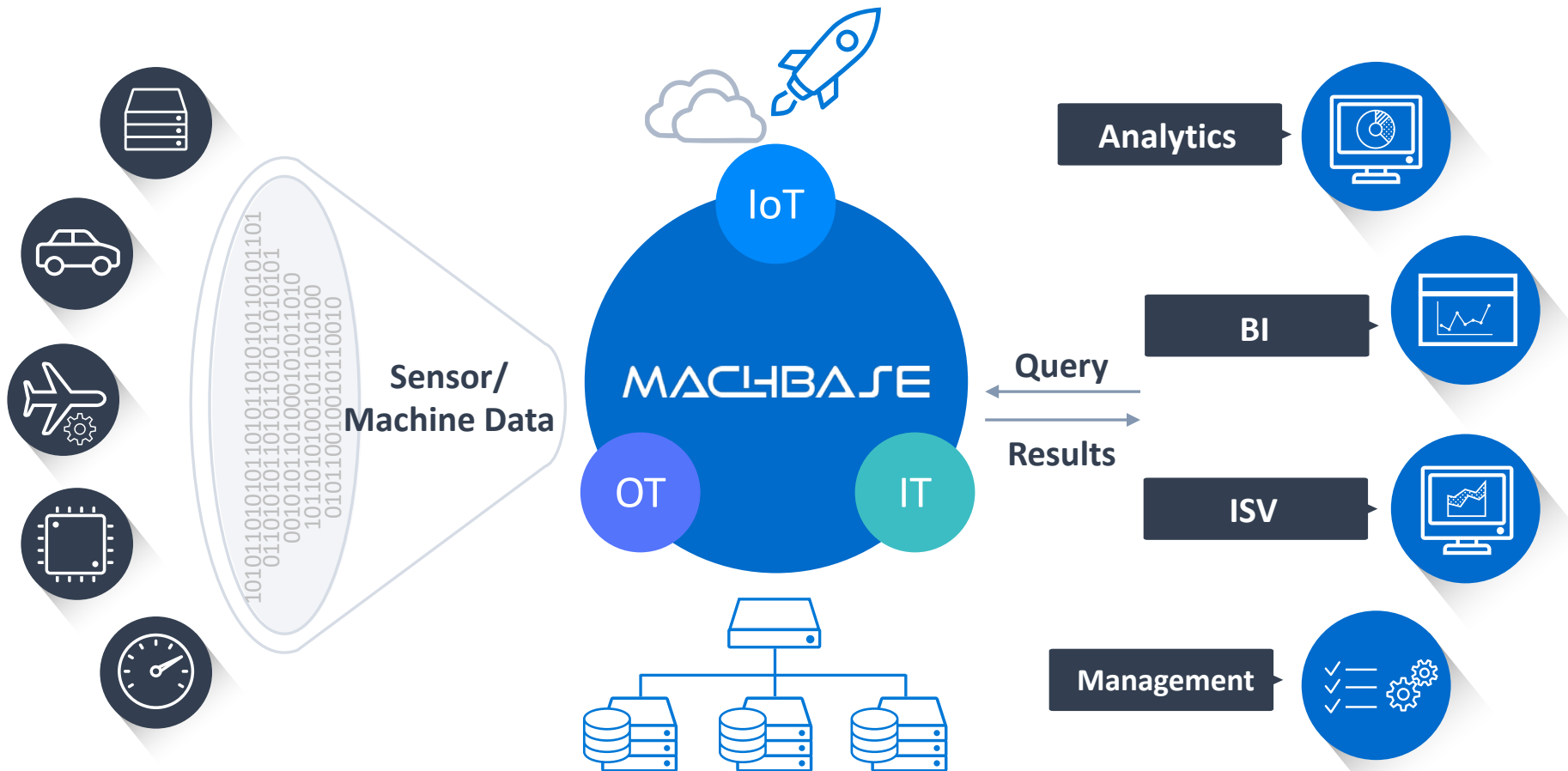
SQL support : 개발 편의성 및 기존 지식 재활용

Machbase for IIoT

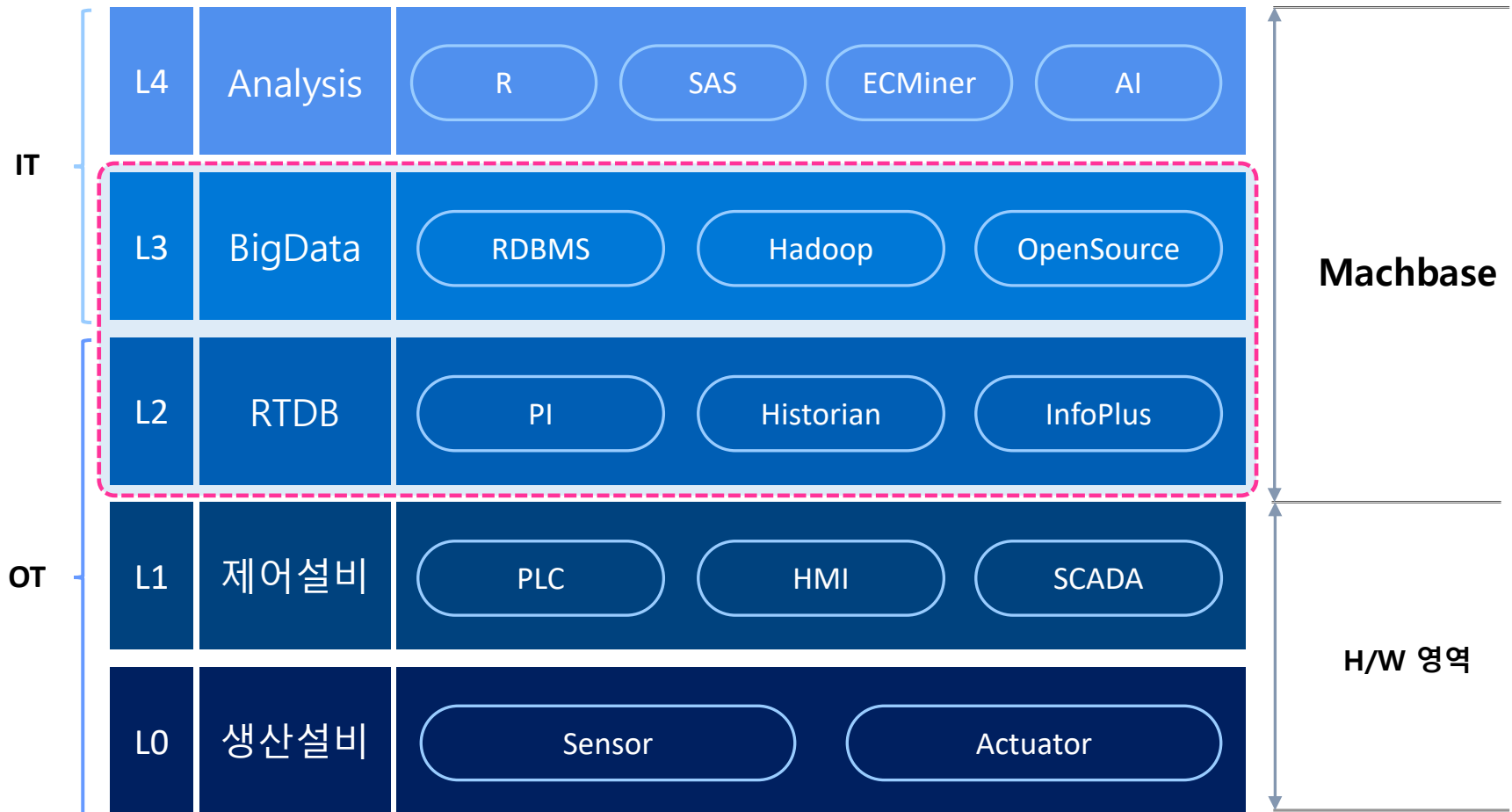


The Database for Things

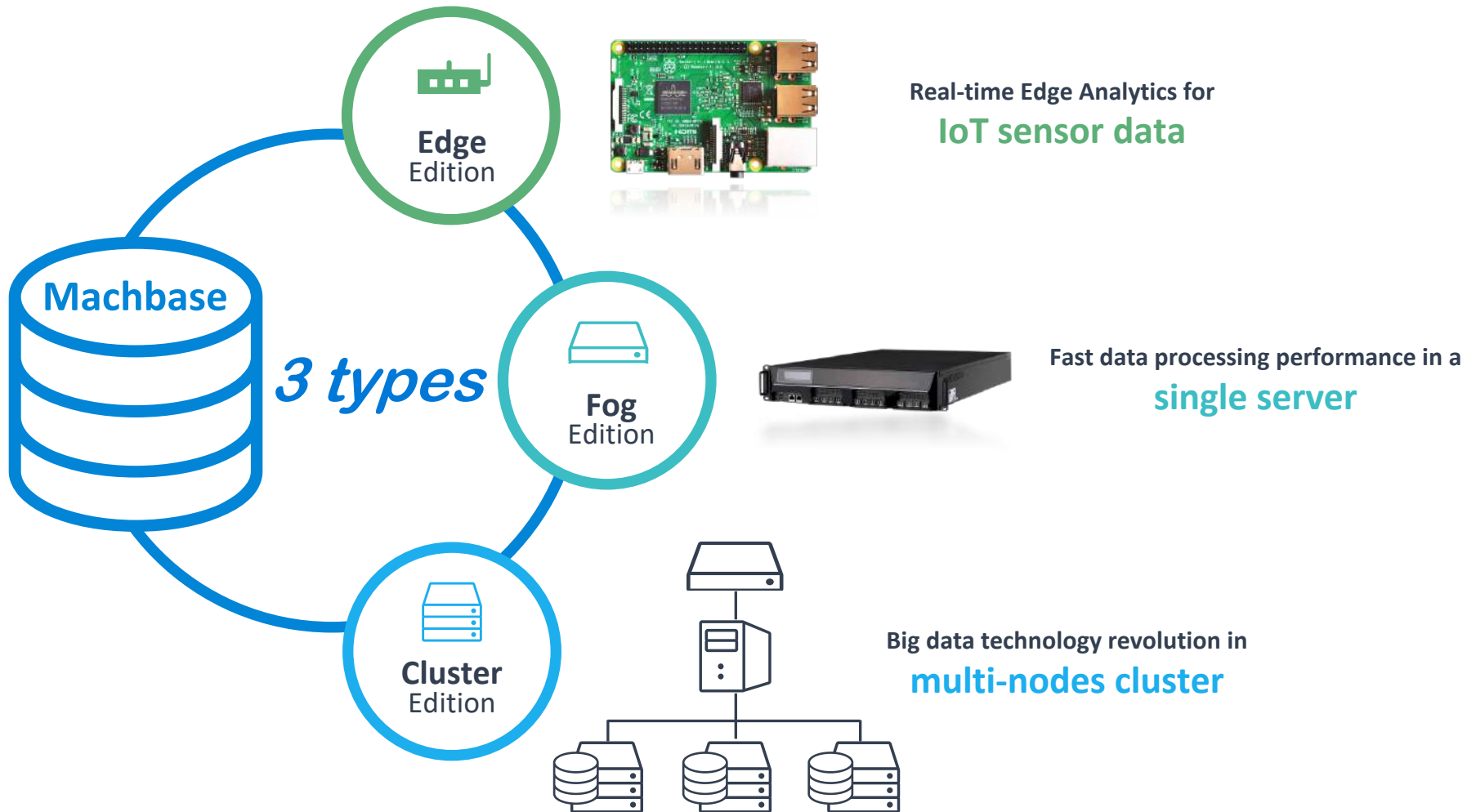
스마트 공장 설비에서 발생하는 센서 데이터를 실시간으로 저장, 분석하는 **시계열 DBMS**



Machbase Position



Machbase 제품군





Edge Edition

Best solution for Edge Analytics

- **CPU** : ARM, ATOM(x86) series
- **O/S** : Linux(Redhat, CentOS, Fedora, Ubuntu)
Real-time Linux (Windriver Linux)
Windows 64bit
- **Performance** : 초당 수천에서 수십만 건 저장 가능
- nVidia Jetson (64 bit, Linux, ARM CPU) 지원



Raspberry PI 3



Raspberry PI 4



nVidia Jetson Tx2



LattePanda

Fog Edition

Fast data process performance in a single server

- 단일 서버에서 최고의 성능 발휘
- 초당 수만 건에서 수십만 건 데이터 저장 가능
- 리눅스, 윈도우 64 bit 지원
- 고성능 시계열 DB를 필요로 하는 ISV의 최적 솔루션



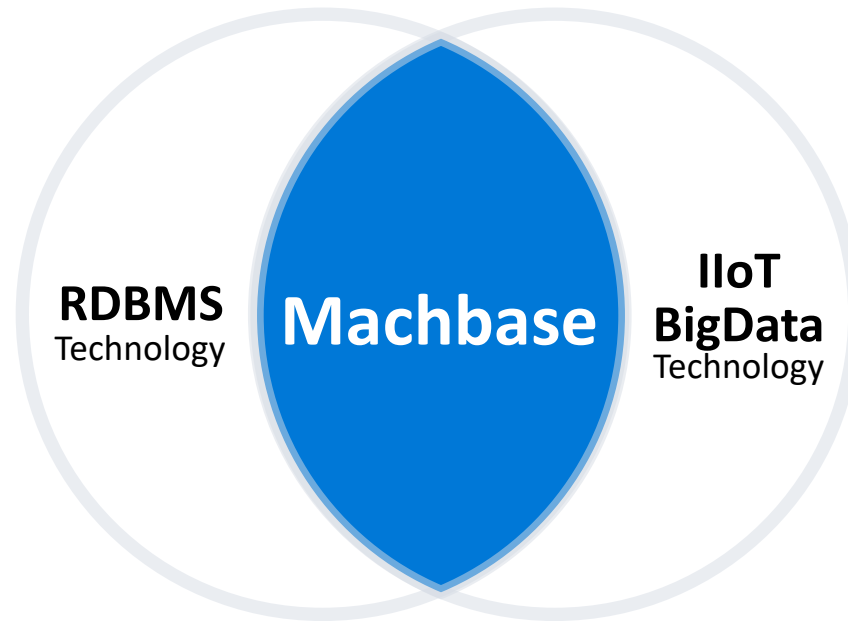
Cluster Edition

Multi-nodes Cluster Scalability

- 다중 노드를 통한 분산 저장, 분산 쿼리 지원
- 고가용성, 고확장성 지원
- 노드 증가에 따른 성능 향상
- 시계열 데이터용 클라우드 시스템 구성 가능



Machbase 혁신성



데이터베이스

- 안정성, 편의성, 사용자 특성 감안
- 소프트웨어 개발 및 관리 편리
- 느린 성능, 소규모 데이터 처리 적합
- 노드 및 데이터 증가에 따른 확장성 부족



빅데이터 솔루션

- 대용량 처리, 클러스터링, CAP 이론
- 대용량 데이터 입력 및 처리에 특화
- 데이터베이스 사용자 편의성 부족
- 학습 및 유지보수 관리 매우 취약

Machbase 기술적 차별성

초고속 데이터 입력 기술

- 시계열 데이터 입력 및 파티셔닝 기술 (미국 특허)
- 병렬 클러스터를 통해 선형적 성능 증가
- 초고속 데이터 전송 프로토콜 구현 (미국 특허)
- 실시간 데이터 압축 기술 (2단계 압축)

멀티 노드 클러스터링 기술

- 클러스터 노드 관리 기술
- 클러스터 기반 초고속 SQL 처리 기술
- 데이터 복제 및 데이터 밸런싱 기술 (미국 특허)
- 고가용 보장 및 노드 확장 기술

초대용량 데이터 검색 기술

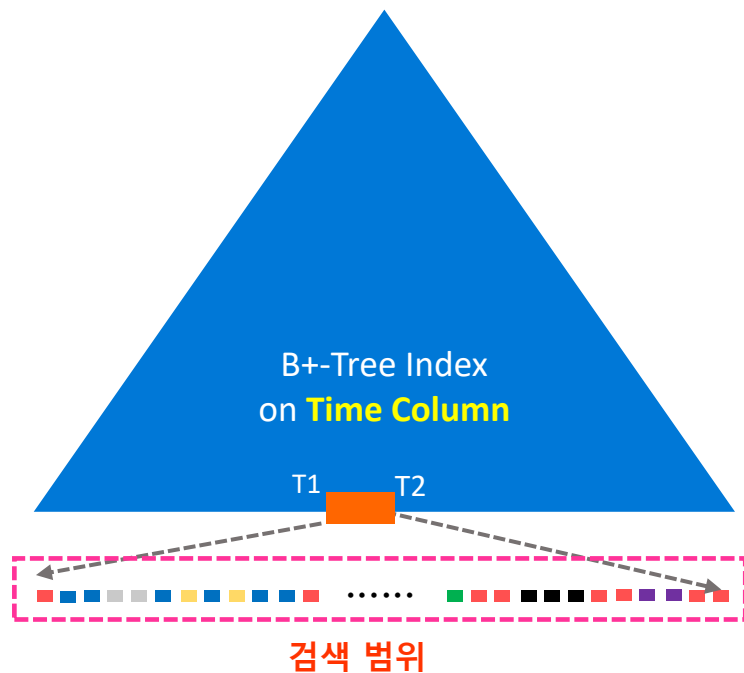
- 센서 데이터를 위한 Tag 구조 및 알고리즘(특허)
- 세계 최고 성능의 SQL 기반의 RTDB 기능 지원
- Rollup Table을 통한 실시간 통계 기능 지원
- 빅데이터를 위한 LSM 인덱스 지원
- 병렬 SQL 처리 기술 (미국 특허)

초고속 인덱스 생성 기술

- 실시간 센서 데이터 Tag 인덱스
- 실시간 비트맵 인덱스
인덱스별 초당 백만 건 이상 처리
- 파티셔닝 기반 병렬 인덱스 생성 기술 (미국 특허)

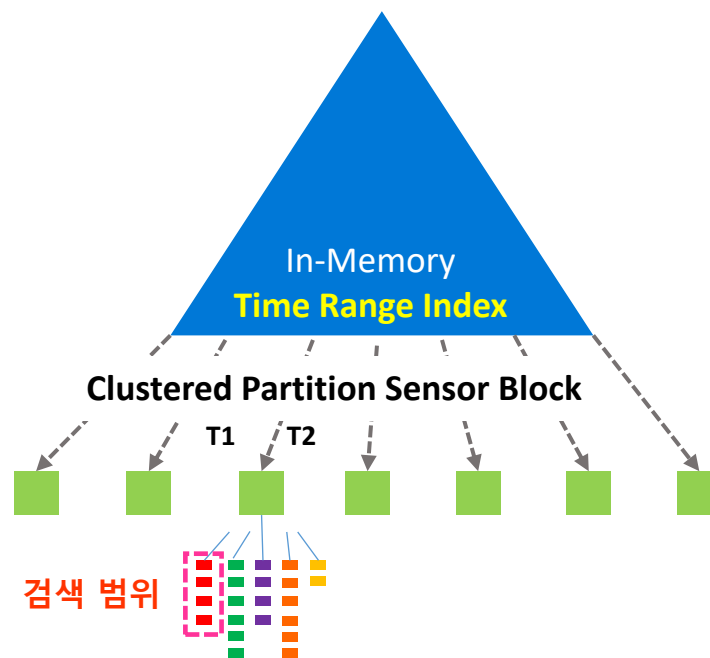
센서 데이터 처리 기법 비교

Tree 기반 기존 기술



- 시간 순서 센서 데이터는 서로 섞임
- 한 개 센서 데이터 검색 범위 매우 증가
- 데이터 량, 센서 개수 증가 시
입력/검색 성능은 선형적으로 저하
- 모든 Tree 기반 구조는 동일 약점 보유

마크베이스 Tag Table 기술



- 각각의 센서 데이터가 Clustering 됨
- 특정시간범위, 센서 데이터 접근시 효율
- 데이터 량, 센서 개수 증가에도
입력/검색 성능은 미비하게 저하
- 수백억 건 저장 기준 밀리 초 추출 성능

특화된 테이블 제공

Memory

Volatile Table

- 휘발성 데이터
- Snapshot 데이터
- PK기반 입력, 수정, 삭제

Lookup Table

- 기준 정보 데이터
- 디스크 저장 유지
- PK기반 입력, 수정, 삭제

Disk

Log Table

- 시계열 로그 데이터
- 이력(Historical) 데이터
- 입력, 선택적 삭제

Tag Table

- 센서 데이터
- 자동 통계 데이터 생성
- 입력, 선택적 삭제

TPC.org 공식 표준 IoT DBMS

TPC (세계 성능 평가 협회)
(Transaction Processing Performance Council)

<http://www.tpc.org/tpcx-iot/>

Full Members

- ✓ HP, IBM, 오라클, 인텔, 마이크로소프트 등 글로벌 IT 업체가 모여 만든 **성능 검증 비영리단체**로 TPC 인증은 **국제 표준**으로 통용되며, 제품 신뢰성을 확보하는 평가 모델로 전세계가 인정
- ✓ TPC 회원사인 **해외 주요 HW 기업**과 **DBMS 벤더**가 협력하여 공식 성능 평가 자료를 공개
- ✓ TPCx-IoT분야는 **IoT 데이터 처리 DBMS 성능**을 평가하는 스펙
- ✓ 정부의 지원(TTA)을 받아 TPC와 함께 테스트 및 등록 진행
- ✓ TPC 공식 검증을 통해 **세계 최고의 성능을 국제적으로 인정**

TPCx-IoT 세계 1위 DBMS

Rank	Company	System	Performance (IoTps)	Price/IoTps	Watts/IoTps	System Availability	Database	Operating System
1		Supermicro A+ Server 2014TP-HTR	2,480,917	.18 USD	NR	04/14/20	Machbase 5.7.13	Red Hat Enterprise Linux Server 7.7
2		Supermicro A+ Server 2014TP-HTR	2,199,052	.20 USD	NR	03/30/20	Machbase 5.7.13	Red Hat Enterprise Linux Server 7.7
3		Machbase 5.7.3	1,043,276	.34 USD	NR	11/11/19	Machbase 5.7.3	CentOS Linux 7.6.1810
4		Lenovo ThinkSystem SR655	742,256	.26 USD	NR	09/18/19	HBase 2.1.4. on Cludera Distribution for Apache Hadoop 6.3.0	Red Hat Enterprise Linux Server 7.6
5		Supermicro A+ Server 2014TP-HTR	482,200	.29 USD	NR	08/07/19	HBase 2.0.0. on Cludera Distribution for Apache Hadoop 6.0.1	Red Hat Enterprise Linux Server 7.6
6		Dell Power Edge R7415	354,811	.75 USD	NR	05/24/19	HBase 2.0.0. on Cludera Distribution for Apache Hadoop 6.0	Red Hat Enterprise Linux Server 7.5
7		Supermicro A+ Server AS-2124BT-HTR	253,178	.54 USD	NR	08/07/19	HBase 2.0.0. on Cludera Distribution for Apache Hadoop 6.0.1	Red Hat Enterprise Linux Server 7.6
8		Dell Power Edge R6415	165,054	.79 USD	NR	11/30/18	HBase 2.0.0. on Cludera Distribution for Apache Hadoop 6.0	Red Hat Enterprise Linux Server 7.5
9		Cisco UCS Mini	142,493	.94 USD	NR	11/20/17	HBase 1.2.1. on Cludera Distribution for Apache Hadoop 5.10.0	Red Hat Enterprise Linux Server 7.2

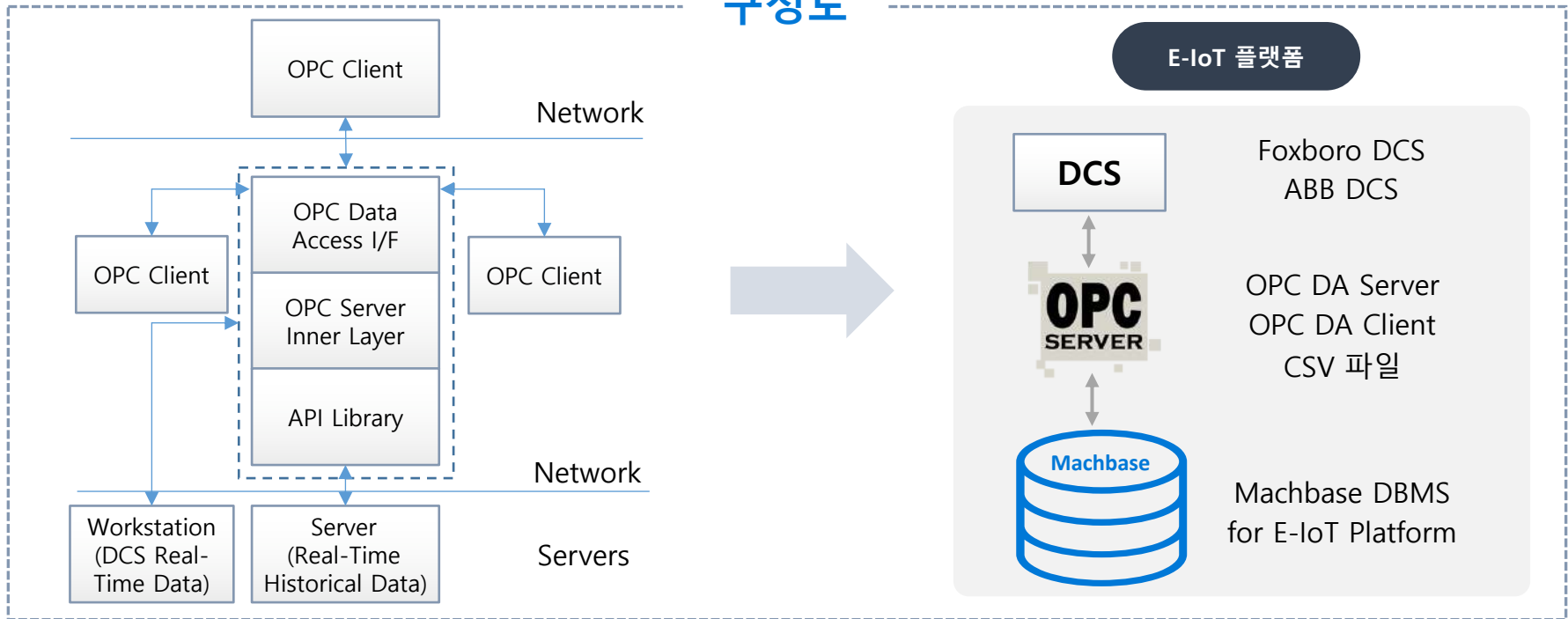
Machbase Use Cases



제지업체 E-IoT 플랫폼

제지 공정 내 설비 최적 운전 사전 단계로 Tag 고속 수집

구성도



배경

- DCS - OPC DA Client 수집 불가
- 기존 시스템 내 RDBMS 탑재
→ Tag 1000 개/초 수집이 어려움
- OPC UA 가 아닌 프로토콜에서의 수집 방안 연구

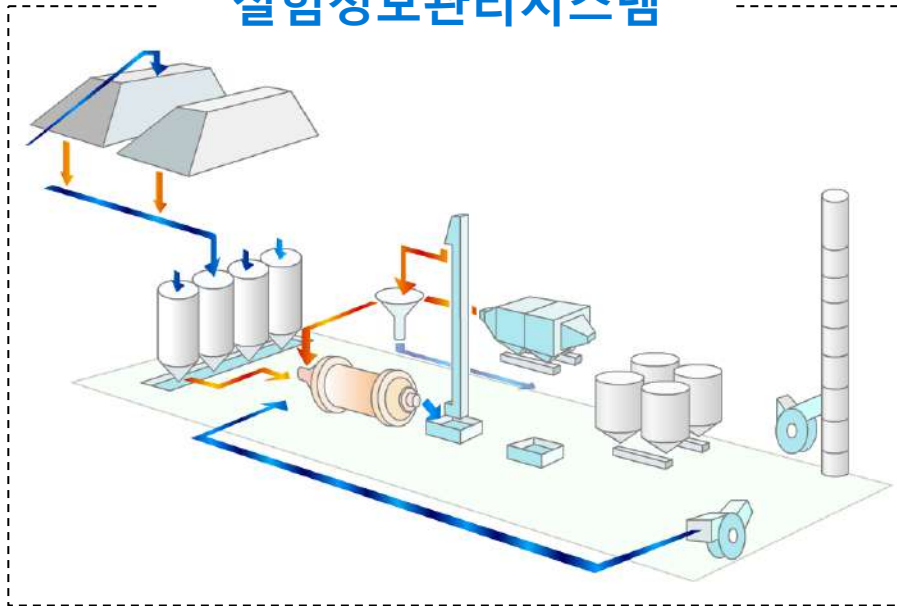
적용 효과

- OPC DA 서버로 부터 생성된 CSV파일 고속 수집
- 초당 1천 Tag 이상 수집 및 1초 단위로 데이터 입력 진행
- 수집한 센서데이터의 Tag 값에 대해 재접속 시 자동 입력

시멘트 공정 품질데이터 모니터링

시계열 DB 를 통한 공정 요소 트렌드, 알람, 이벤트, 조치

실험정보관리시스템



연속 공정 내 품질 데이터 집계

시계열 데이터베이스 + 실시간 모니터링

도입 배경

- 압축감도기, 동결 측정기, 밀도 측정기 등 수십 개 장비 내 선택 데이터에 대한 종합화면 구성 필요
- 검사항목 별로 샘플링한 검사 결과에 대한 자동화데이터 추출 및 모니터링 니즈

Needs

- 실험 데이터는 MS SQL 내 쌓고 있으나 각 공정별 품질 데이터 변화에 대한 실시간 모니터링이 어려워
- 수집 주기를 15분에서 초 단위로 변경하여 데이터 정밀도 향상 필요

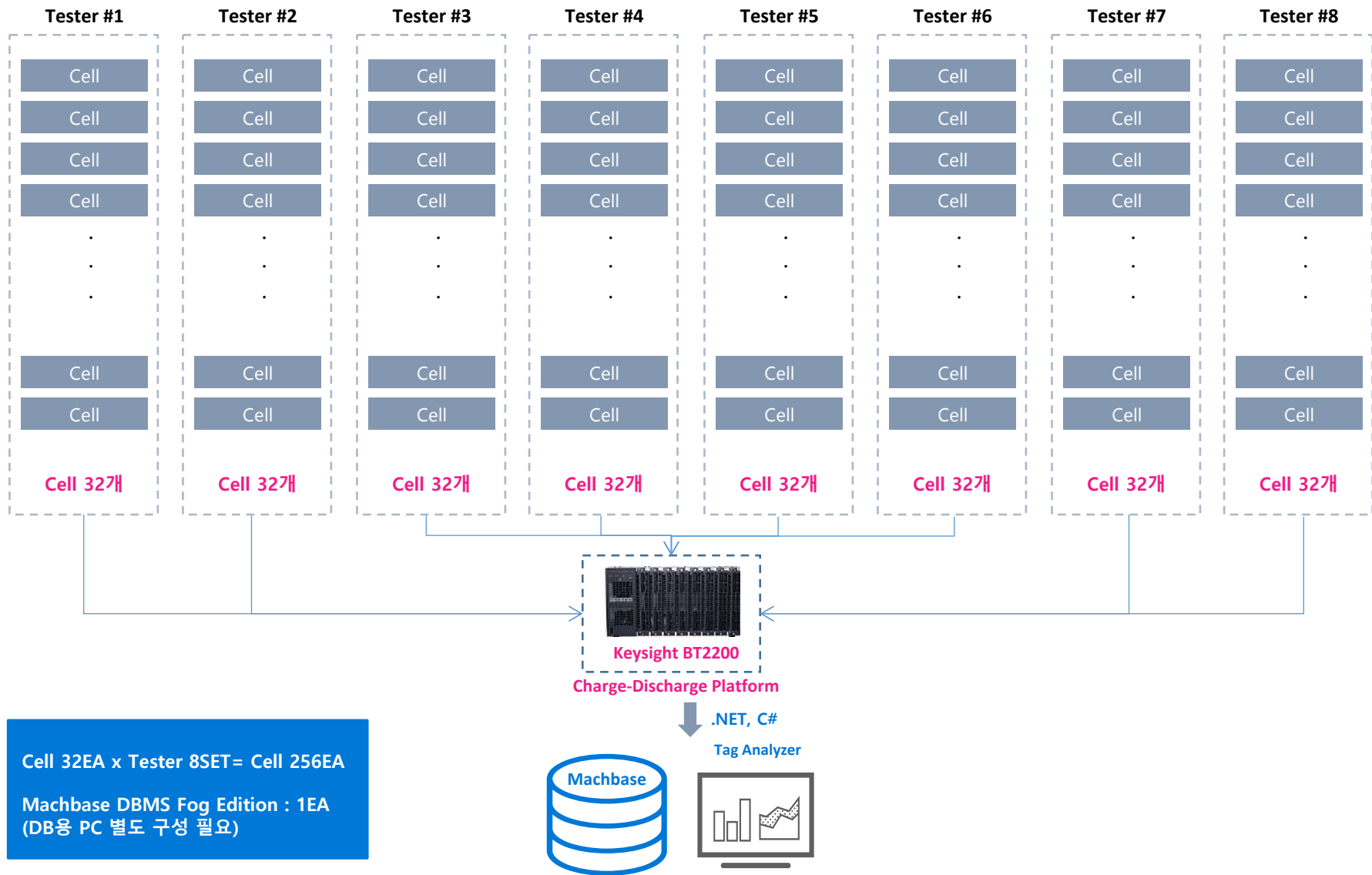
Pain Point

- 특수한 공정 내에서 이전 두 시간에 한 번씩 체크
- 아날로그데이터 수기 입력
- 최종 공정에 대한 결과치만 있고 중간 공정에는 없어

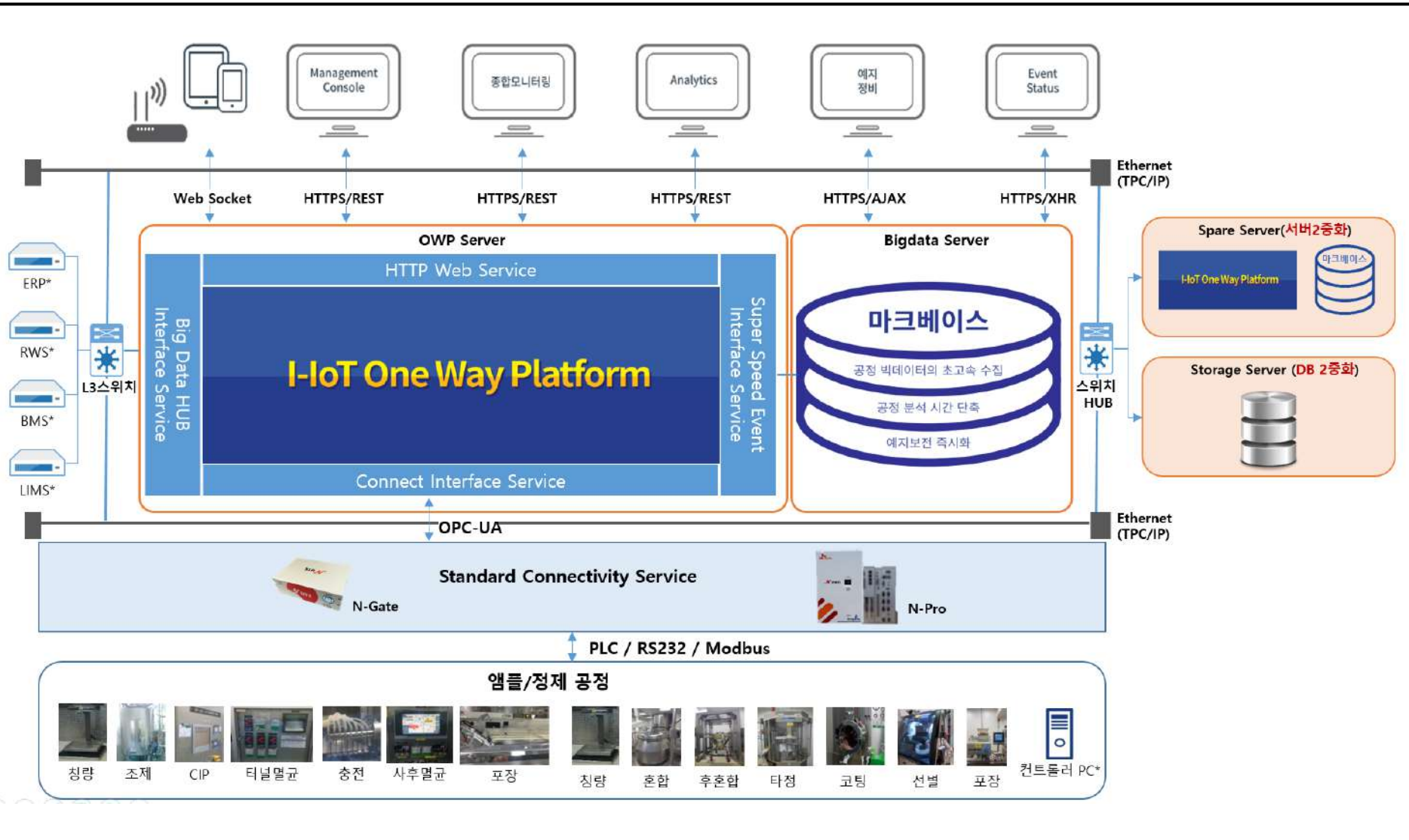
적용 효과

- 원료 (석회석) 품위 데이터에 대하여 초 단위로 데이터 수집
- 데이터 적재 시 데이터량 및 품위 변경에 따른 샘플링 결과와 장비별, 공정별로 시시각각 모니터링 가능해짐

리튬 이온 배터리 품질 데이터 수집



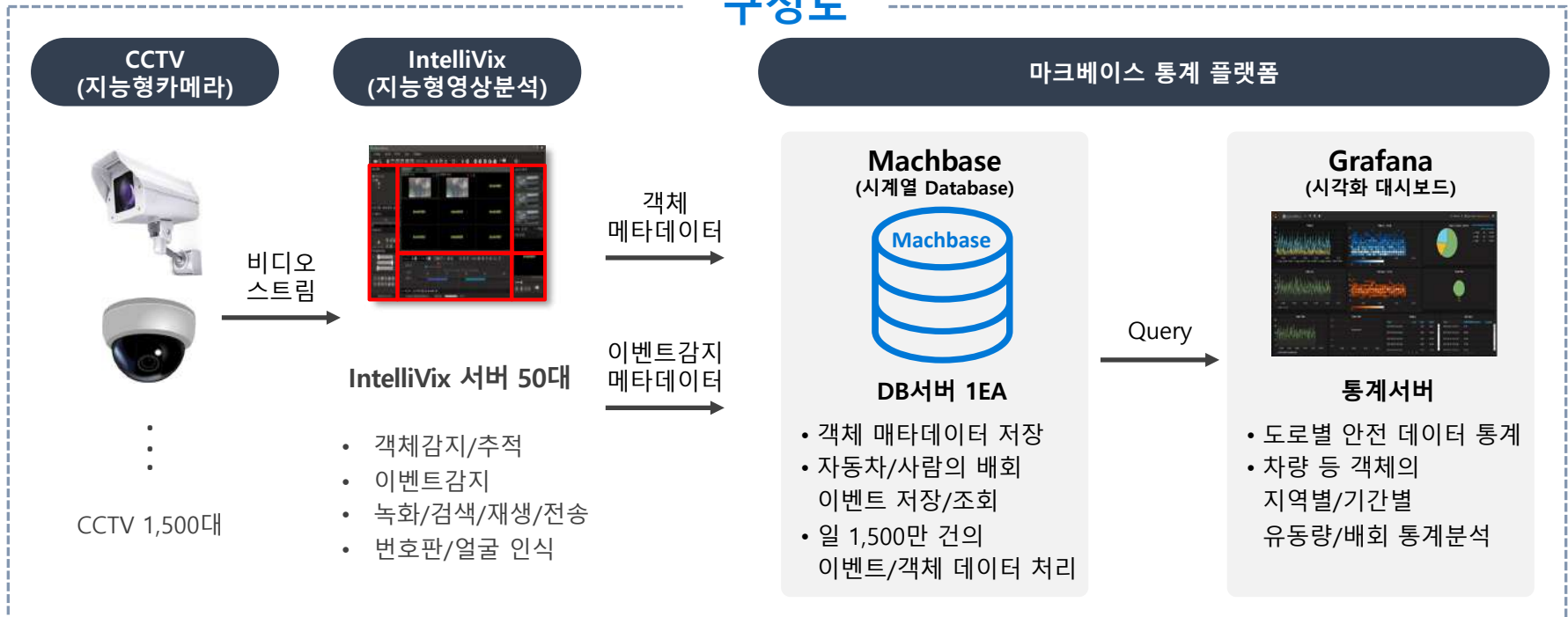
제약사 스마트팩토리



지능형 영상정보 저장 분석

오산시 지능형 CCTV 영상 정보 저장 및 통계 분석 플랫폼

구성도



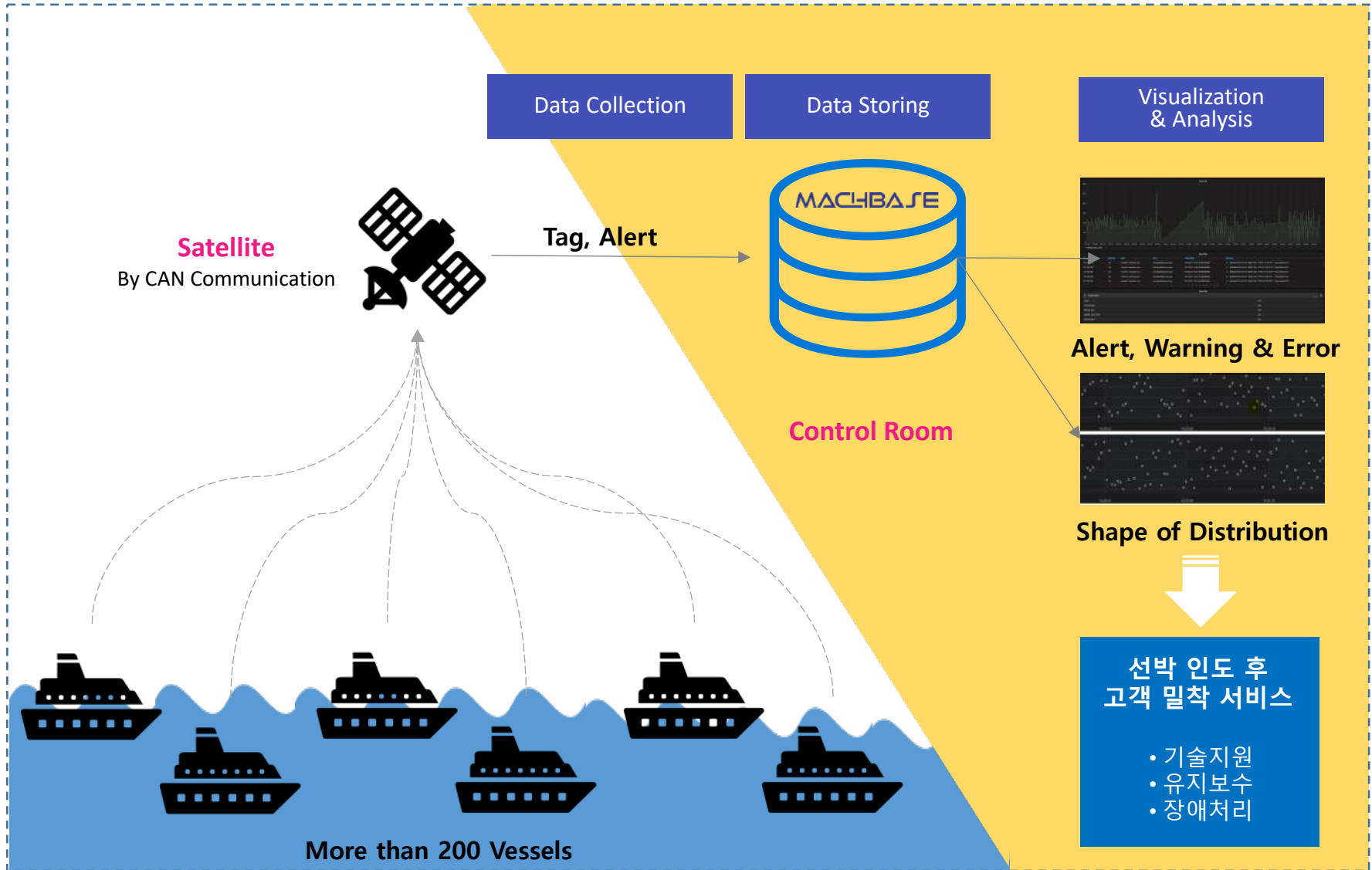
배경

- 인텔리빅스 내부 DB로는 모든 이벤트/메타 데이터 저장을 못함
- 속도 및 DB Size 제약으로 최근 1개월 내의 데이터만 보관
- 월/연인원, 차량/인구 유동량 등 통계자료 확보 못함

적용 효과

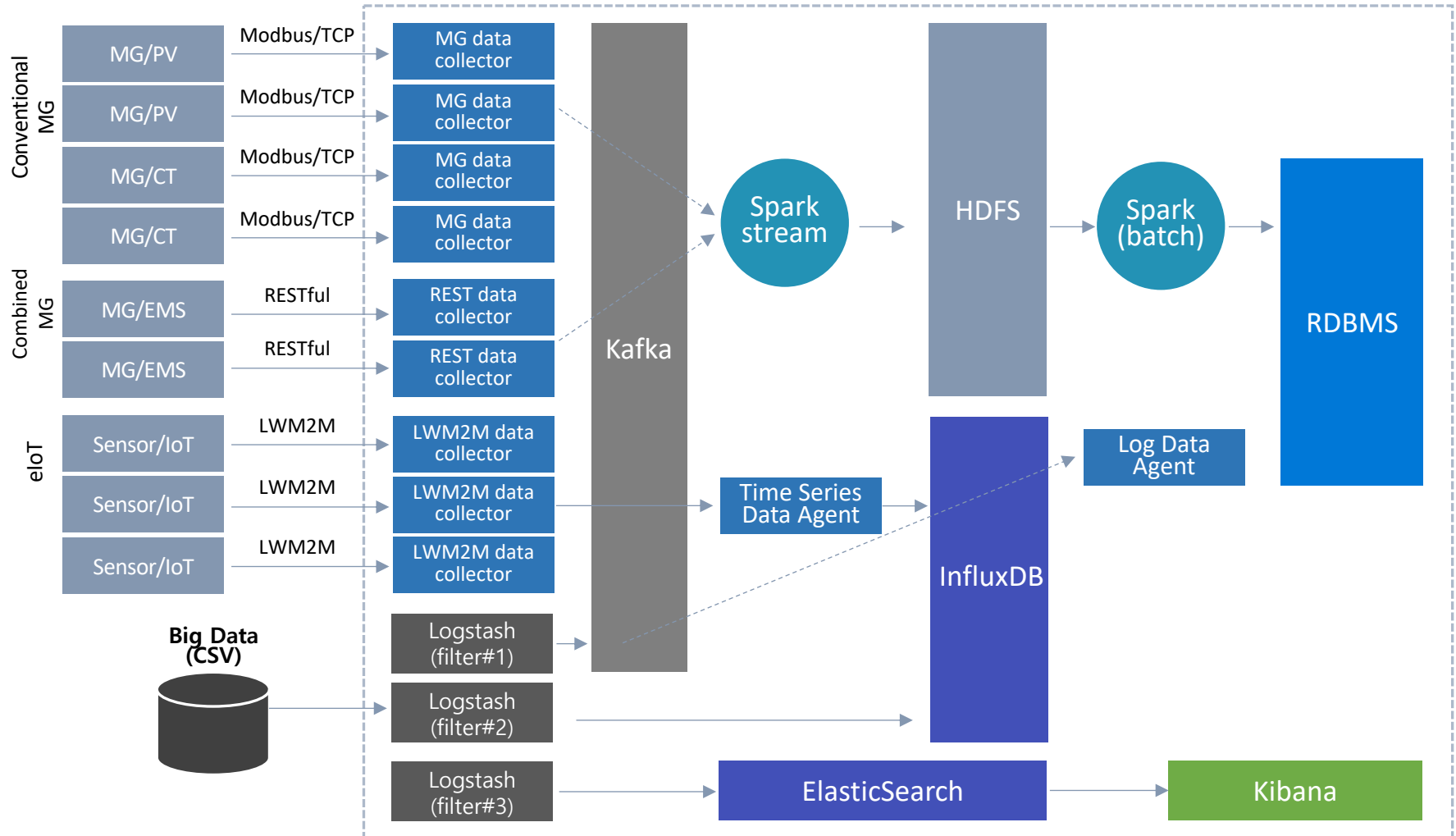
- 1년 이상의 전체 데이터 보관 가능하며, 향후 확장하여 최대 5년간 보관 예정
- 교차로 등의 사건/사고 추이 분석을 위한 기초자료 제공 및 대응 방안 수립을 위한 통계분석 활용
- 월별, 연별 유동량 및 배회 등 이벤트 패턴분석 기반 마련

선박 운전정보 모니터링 시스템



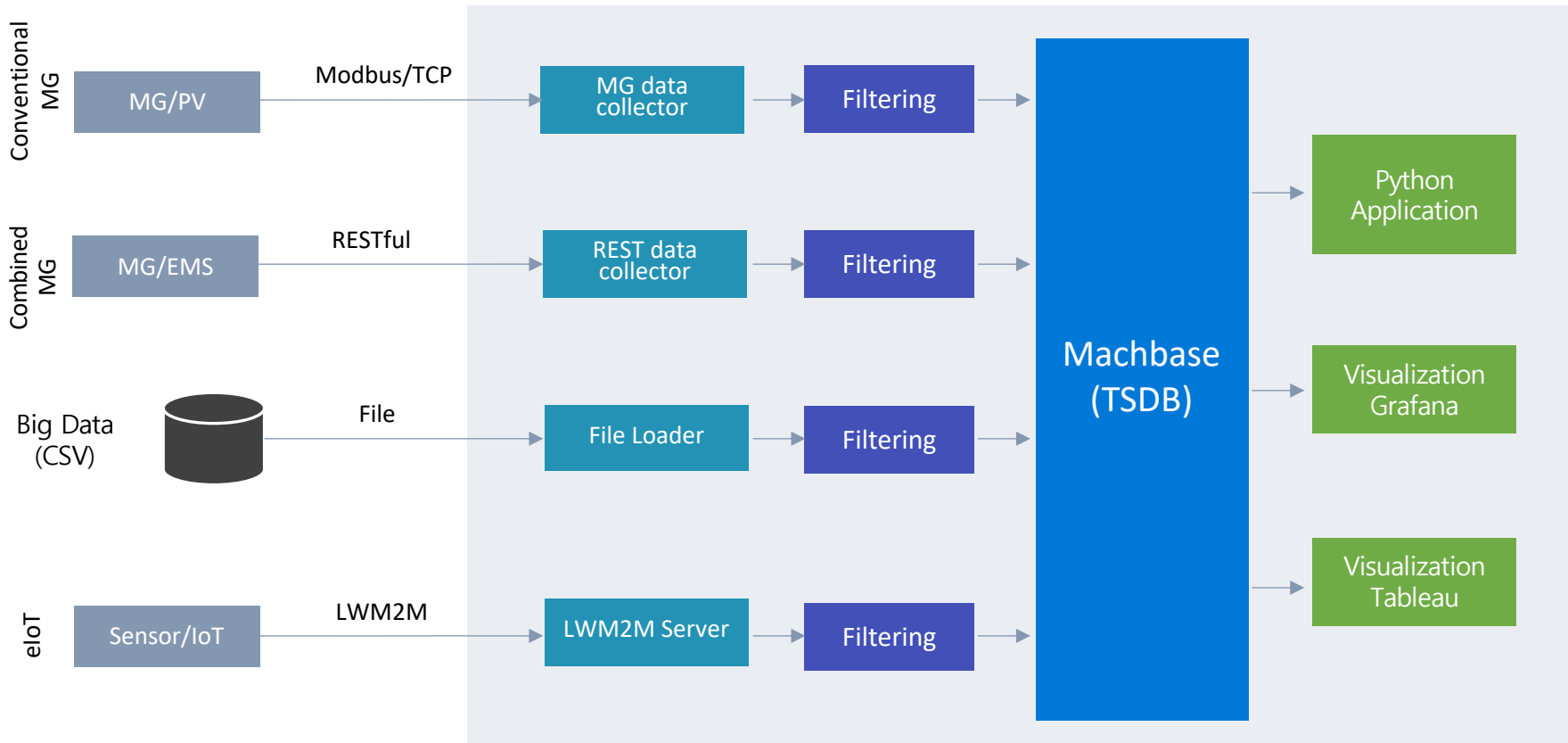
에너지 빅데이터 플랫폼(Before)

Hadoop Ecosystem Architecture



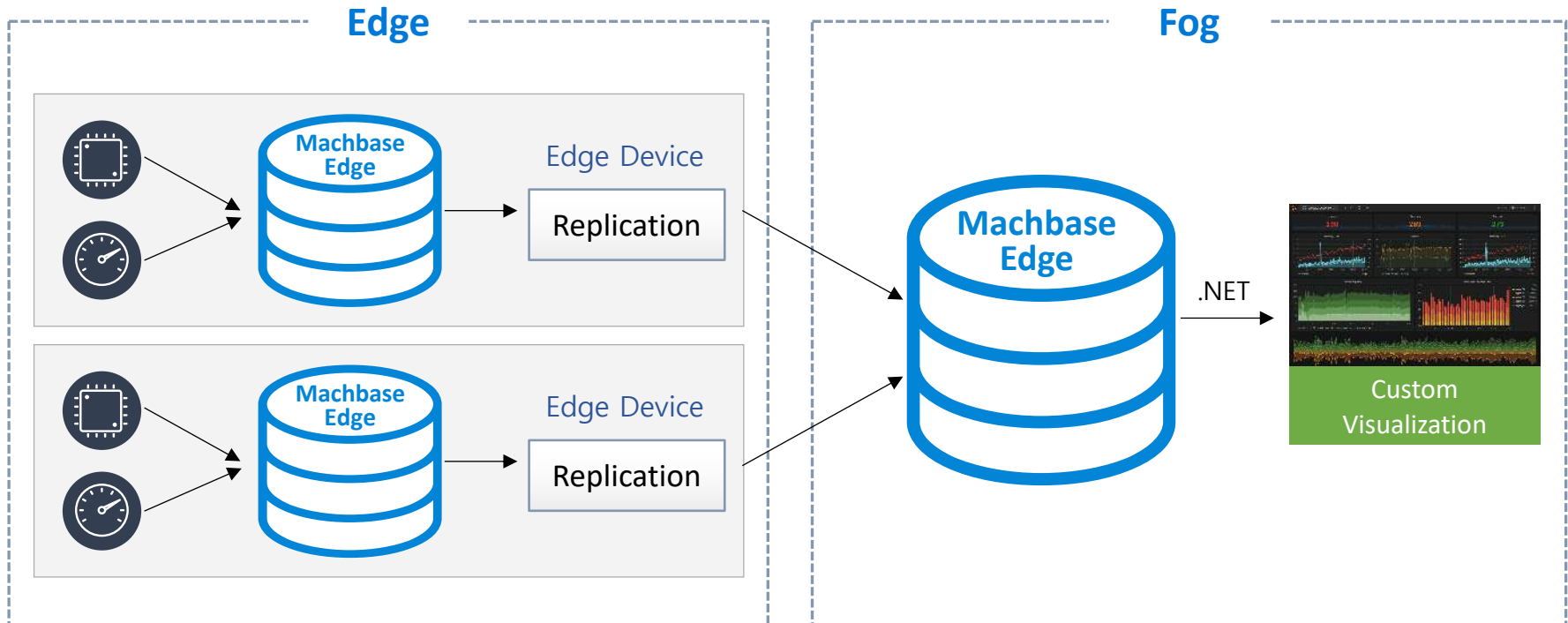
에너지 빅데이터 플랫폼(After)

Machbase Database Architecture



Edge Device Embedding

Edge Device 내 마크베이스 임베딩하여 Edge Computing 구현



배경

- 기존에는 Edge없이 Fog서버의 RDB에 데이터 저장
- RDB로는 raw 데이터 저장 못하고 avg값만 저장
- Fog에서 데이터 검색 성능 저하

적용 효과

- Edge Device내 Machbase 임베딩, 1차 데이터 수집
- 센서 개수 증가에 대응 가능한 데이터 수집 구조 구현
- 현장에 Edge Computing 적용을 위한 검증 완료

Summary



Summary

Smart-X

- Smart-X 시대 센서 데이터 처리에 대한 요구 사항 증대

Edge
Computing

- Edge Computing 이 IIoT 의 중심 기술로 대두되고 있음

TimeSeries
DBMS

- 센서 데이터 영역을 위한 최고 성능의 신개념 DBMS 필요

Machbase

- 마크베이스는 센서 데이터를 위한 최적의 DBMS 솔루션

**Human
Data**

**Oracle
MS-SQL**

**Machine
Data**

MACHBASE