

# 엣지 기반 경량화 플랫폼을 위한 데이터 수집 시스템의 데이터베이스 비교 및 평가

<sup>1</sup> 조우진, <sup>2</sup> 임채영, <sup>3\*</sup> 구재회

## Comparison and Evaluation of Data Collection System Database for Edge-Based Lightweight Platform

<sup>1</sup>Woojin Cho, <sup>2</sup>Chae-young Lim, <sup>3\*</sup>Jae-hoi Gu

### 요약

에너지 관리 시스템 중 공장 에너지 관리 시스템은 제 3차 에너지 기본 계획과 전세계적 에너지 비용 증가, 환경·기후 문제등에 의해 급격히 성장하고 발전하고 있다. 하지만 이러한 공장 에너지 관리 시스템은 한정된 공간 등의 공장의 특성 상 에너지 관리 시스템의 데이터 수집을 위해 필수적인 데이터 수집 시스템의 구축이 공간적, 환경적, 에너지적으로 어려움을 지닌다.

본 논문은 이러한 공간, 환경, 에너지 관련 문제를 개선하기 위해 엣지 기반 경량화 플랫폼으로 데이터 수집 시스템을 구축하여 엣지 디바이스에서 데이터베이스 구동에 대한 비교 및 평가를 진행한다. 이를 평가하기 위하여 기존 실증 중에 있었던 공장의 특성을 활용하여 CDI Benchmark라는 평가 도구를 개발하여 평가를 진행한다. 평가 결과 MySQL과 같은 RDBMS는 대량 데이터 삽입 시 구동 문제가 발생하여 사용에 부적합 하며, InfluxDB의 경우 높은 압축률을 지닌 압축 알고리즘으로 인해 평가에 따라 MyRocks보다 최대 6배 작은 데이터베이스 크기를 지니지만, MyRocks의 경우 InfluxDB에 비해 최대 80배가량 적은 시간을 소모하는 것을 확인하였다.

### Abstract

Factory energy management system is rapidly growing and evolving due to factors such as the 3rd Basic Energy Plan and global energy cost increases, as well as environmental issues. However, implementing an essential data collection system for energy management in factory settings, which have limited space and unique characteristics, presents spatial, environmental, and energy-related challenges.

This paper endeavors to mitigate these challenges by devising a data collection system implemented through an edge-based lightweight platform. A comparison and evaluation of database operation on edge devices are conducted. To conduct the evaluation, a benchmarking tool called CDI Benchmark is developed, utilizing the characteristics of existing factories involved in practical applications. The evaluation results revealed that RDBMS systems like MySQL encountered errors in the database due to high data insertion loads, making them inoperable. On the other hand, InfluxDB, thanks to its highly efficient compression algorithm, demonstrated compression rates about 6 times higher than MyRocks according to the evaluation. However, it was observed that MyRocks outperformed InfluxDB by a significant margin, recording a maximum processing time approximately 80 times faster compared to InfluxDB.

**Keywords:** Database, FEMS, Time series data, Benchmark, Data Collection System, LSM tree, Evaluation

<sup>1</sup> 고등기술연구원 에너지환경 IT 융합그룹 연구원(woju\_1@iae.re.kr)

<sup>2</sup> 고등기술연구원 에너지환경 IT 융합그룹 선임연구원(cylim@iae.re.kr)

<sup>3\*</sup> 고등기술연구원 에너지환경 IT 융합그룹 연구위원, 교신저자(jaehoi@iae.re.kr)

## I. 서론

전 세계적으로 에너지 소비 절감을 위해 기존의 경험적 에너지 관리 방식이 아닌 새로운 에너지 관리 방식인 데이터 기반, 인공지능 기반의 에너지 관리 방식에 대한 연구와 수요가 증가하고 있다[1]. 이러한 전통적 에너지 관리의 한계를 개선하고 데이터의 고도화 및 관리, 최적 제어 등의 기술을 적용하여 에너지를 절감하고자 에너지 관리 방식과 다양한 기법들을 시스템화한 EMS(Energy Management System)라고 하는 에너지 관리 시스템을 통하여 에너지를 관리 및 절감하는 방안에 대해 연구되었다. 에너지 관리 시스템은 고도화된 ICT 기법을 활용하여 빌딩, 공장, 주택, 사회인프라 등을 대상으로 에너지의 흐름 분석 및 시각화, 최적화를 통한 에너지 관리 시스템이다. 이러한 에너지 관리 시스템은 사용자 별로 BEMS(Building Energy Management System), HEMS(Home Energy Management System)등으로 분류한다. 그 중 가장 규모가 크고 널리 사용되는 공장 에너지 관리 시스템(Factory Energy Management System)은 대한민국의 제 3 차 에너지 기본 계획을 통해 2025 년부터 10 만 TOE(Ton of Oil Equivalent)이상의 에너지 다소비 사업장의 경우 FEMS 의무화 정책이 수립되었으며, 10 만 TOE 이하의 공장에 대해서도 FEMS 보급에 대해 추진 중에 있다[2]. 정부의 의무화 정책과 사회적 필요에 의해 에너지 관리 시스템 고도화에 대한 연구로 에너지 관리 시스템의 다양한 기술이 개발되고 있다.

특히 에너지 관리 시스템 기술의 완성도를 높이기 위해 제어 최적화와 같은 인공지능 기법을 활용하기 위해서는 데이터 수집이 기반이 되어야 하기 때문에 데이터 수집 시스템 운용은 중요한 역할을 한다. 데이터 수집 시스템의 운용을 위해서는 해당 데이터를 수집할 센서, PLC(Programmable Logic Controller), OPC-UA(Open Platform Communications Unified Architecture) 서버 및 데이터베이스 서버 등을 활용해야 하기 때문에 대규모 시스템을 위한 시스템 구축이 필요하다. 하지만 그러한 시스템을 구축하는 경우 서버를 위해 특정 습도와 온도를 맞춰야 하고, 그로 인해 큰 공간을 차지해야 하는 서버 공간의 문제는 공간 및 환경의 제약이 많은 공장에서는 특히 부각된다. 또한 서버 시스템의 매우 높은 전력 사용량과 온, 습도와 같은 서버에 적합한 환경으로 맞추기 위한 부가 장비는 에너지 관리 시스템을 통해 에너지를 절약을 위해 에너지를 과소비해야 하는 역설적인 상황을 야기한다. 이와 같이 모든 서버를 현장에 구축하는 것은 인공지능 학습 시 사용 패턴에 대한 학습을 후에는 변화가 크지 않은 공장 에너지 관리 시스템의 경우 에너지적으로도 공간적으로도, 자원적으로도 큰 손해를 야기한다.

따라서 본 논문에서는 이러한 환경, 에너지 차원의 문제를 개선하기 위해 기존의 서버 시스템이 아닌 ARM 프로세스 기반의 엣지 디바이스에 데이터 수집 시스템을 구축하여 평가한다. 하지만 엣지 디바이스를 기존의 서버 시스템을 대체하여 운용하는 것에 있어서는 성능, 저장공간, 신뢰성 등에 문제가 발생할 수 있다. 데이터 수집 시스템의 경우 지속적으로 손실 없이 데이터를 수집해야 하는 조건에 따라 여러 문제가 발생할 여지가 있어 평가를 통해 실 사용 시 문제 발생여지를 파악한다. 또한 향후 데이터 수집 후 엣지 디바이스에서 데이터를 가공하고 정제하며, 실시간 분석 등을 하는 역할을 추가적으로 엣지 디바이스가 분담하여 처리할 수 있도록 하기 위해서 데이터 수집 시 성능에 부담이 적은 데이터베이스를 선택하는 것이 하나의 관건이 될 수 있기 때문에 개발한 벤치마크를 사용하여 엣지 디바이스에서 데이터베이스 신뢰도와 데이터베이스의 오류에 의한 문제 방지가 가능한 최적의 데이터베이스를 확인하고 발생할 수 있는 위험을 방지할 수 있는 연구를 진행한다.

본 논문에서는 시계열 데이터를 적재할 때 선택할 수 있는 대표적인 데이터베이스 MySQL, MyRocks, InfluxDB 세가지를 엣지 디바이스에서 구축하여 실제 세계를 기반으로 벤치마크를 진행하는 CDI Benchmark 를 개발하여 수행해 데이터베이스 신뢰도 및 성능 평가를 진행한다[3][4][5]. 평가를 진행한 결과로는 MySQL 의 경우 일정 이상의 데이터를 삽입할 시 I/O 병목으로 인해 테이블 문제가 발생하여 삽입이 불가능한 오류를 발생시켰다. MyRocks 는 가장 삽입 성능이 좋았으나 평균 CPU 사용률이 높게 유지되었다. InfluxDB 의 경우 가장 적은 데이터베이스 용량을 사용하였지만 삽입 시 발생하는 오버헤드에 의해 대량 적재(Bulk Insert)를 사용하여도 1000 만개의 데이터 삽입 시 긴 소요 시간을 지니는 것을 볼 수 있었다.

본 논문은 총 세가지를 주요하게 기여한다. 첫째로 엣지 디바이스의 데이터 수집 시스템 구축에 대한 평가를 통해 엣지 디바이스의 신뢰성에 대한 연구를 진행한다. 두번째로 이중

데이터베이스를 지원하는 벤치마크를 개발함에 따라 다양한 데이터베이스에서 동일한 평가 지표를 지닌 평가 도구를 사용한 평가를 가능하게 한다. 마지막으로 실제를 모사한 벤치마크를 통해 임베디드 디바이스를 활용한 데이터 수집 시스템 및 데이터베이스 구축에 대한 표준 평가로 활용 가능한 평가 도구로써 활용 가능하다.

## II. Background

### 2.1 데이터 수집 시스템

전통적인 데이터 수집 시스템은 DAQ(Data Acquisition)이라 불리며 변환기, 센서, 데이터 수집, PLC 만을 의미하는 협의의 데이터 수집 과정이자 시스템을 의미하였다[6].

하지만 협의의 데이터 수집 시스템 개념은 빅데이터 및 인공지능 시대와 고성능 컴퓨팅, 데이터의 개념의 확장에 의한 부분에 대해서는 처리할 수 없었다. 그에 따라 데이터 수집 시스템은 데이터베이스와 같은 프로그램을 활용한 수집 데이터의 저장, 인공지능과 같은 응용을 위한 데이터의 전 처리, 데이터 실시간 처리 등의 광의의 개념까지 포괄하여 사용하게 되었다. 이와 같이 확장된 개념의 데이터 수집 시스템은 데이터 수집을 사용하는 대부분의 시스템에서 중요한 역할을 담당한다.

그에 따라 데이터 수집 시스템을 보다 높은 효율로 운영하기 위하여 데이터베이스 및 데이터 전 처리 부분들을 엷지 디바이스 측으로 낮추어 말단에서 보다 많은 작업을 할 수 있도록 다양한 연구를 진행하고 있다. 본 논문에서도 엷지 디바이스에서 효율적인 데이터베이스를 사용하기 위한 평가를 통해 기여를 하고 엷지 디바이스의 신뢰성 연구 또한 함께 진행한다.

### 2.2 데이터베이스

데이터 수집 시스템에서는 센서, PLC 등을 통해 수집된 데이터를 최종적으로 저장하게 되는데, 이 때 향후 데이터 사용 및 데이터의 관리를 위해 파일 기반 저장에 비해 관리 및 활용에 용이한 데이터베이스를 주로 사용하게 된다.

따라서 데이터베이스는 데이터의 보존 및 관리에 있어 매우 중요하다. 하지만, 공장 에너지 관리 시스템이 구축되는 협소한 공간 및 작업 환경 등의 환경적 요인과 기존의 서버 시스템의 한 대당 1000W 가 넘어가는 전력 소비량과 온, 습도 조절을 위한 장비들의 전력 소비와 같은 문제에 있어 엷지 디바이스에 대한 요구가 커지고 있다. 하지만 그에 반해 엷지 디바이스에서 구동 가능한 데이터베이스 및 신뢰도에 대한 연구는 매우 부족한 실정이다.

기존 연구로 엷지 디바이스에서 데이터베이스의 스토리지 엔진을 평가한 논문이 존재하나, 해당 논문은 MySQL 에서 Storage engine 만 바꾸어 가며 테스트를 진행한 연구였으며 엷지 디바이스에 대한 평가보다는 Storage engine 에 대한 평가를 주로 다루었으며, 한 개의 데이터베이스만을 타겟으로 제작한 벤치마크를 수행한 것이기에 이중 데이터베이스에서는 구동할 수 없는 문제가 존재했다[7].

또한 엷지 디바이스에서 최적으로 구동 할 수 있는 데이터베이스를 연구를 진행한 경우도 존재하나 이중 데이터베이스에 대한 평가 및 실제 데이터를 기반으로 실험을 진행하지 않아 엷지 디바이스를 현실에 적용하는 것을 고려하기 어려웠다[8].

엷지 디바이스의 성능은 과거에 비해 크게 발전하였으나, 성능과 안정성은 다른 지표이기 때문에 엷지 디바이스에서 검증 없이 데이터 수집을 수행하는 것은 여러 문제점을 발생시킬 수 있다. 특히 데이터베이스는 수집한 데이터의 최종 종착점인 만큼 무결성과 높은 신뢰성이 보장되어야 한다. 따라서 본 논문에서는 특성이 다른 세가지 데이터베이스를 비교하고 각각의 데이터베이스를 평가한 후 데이터 수집 시스템에서의 엷지 디바이스 활용도 및 데이터베이스 구동에서 있어서 엷지 디바이스의 신뢰성에 대해 고찰한다.

#### 2.2.1 MySQL

전통적 데이터베이스의 일종인 관계형 데이터베이스이다. 오픈소스 데이터베이스 중 가장 많은 사용자를 지니고 있으며, MySQL 의 확장성 및 편의성, 이식성 등의 다양한 장점에 의해

선택된다. 또한 InnoDB 라는 B-tree 기반 스토리지 엔진을 사용하여 키를 인덱싱하여 사용해 높은 조회 성능을 지닌다.

### 2.2.2 MyRocks

Facebook 에서 개발한 MyRocks 라는 데이터베이스는 발전된 MySQL 의 생태계를 사용할 수 있으며 Storage engine 을 Facebook 에서 개발한 Key-value store 인 RocksDB 로 하여 제작된 데이터베이스이다. RocksDB 는 LSM tree 를 통해 쓰기를 순차 쓰기로 만들고, B tree 의 in-place update 를 제거하여 쓰기 성능을 향상시키는 알고리즘을 적용하였다[9].

### 2.2.3 InfluxDB

시계열 데이터를 처리하기 적합한 시계열 데이터베이스의 일종으로 수많은 시계열 데이터베이스 중 가장 많은 사용자를 지니고 있는 데이터베이스이다. 높은 시계열 쓰기 처리를 위해 개발된 데이터베이스이다.

본 논문에서는 각각 RDBMS 에서 대표적인 MySQL 과 Backend storage engine 이 Key-value store 인 MyRocks, 시계열 데이터베이스에서 가장 대표적인 InfluxDB, 총 세가지 데이터베이스를 엷지 디바이스에서 평가하여 낮은 사양의 디바이스에서 시계열 데이터 처리에 가장 적합한 데이터 베이스를 평가하고 수행 시의 데이터베이스 신뢰도 검증을 진행한다.

**Table 1.** Database characteristic

표 1. 데이터베이스 특징

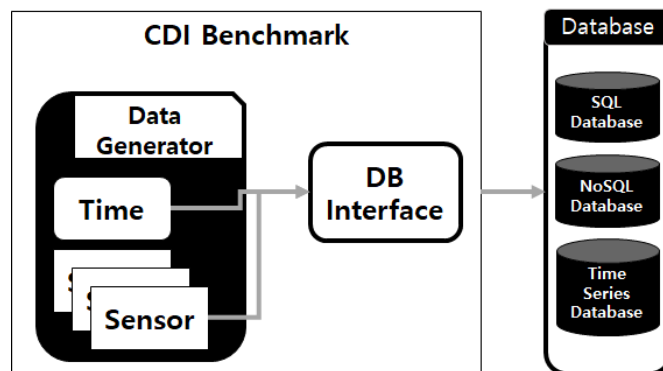
	MySQL	MyRocks	InfluxDB
SQL Support	Support	Support	partial support
Storage Engine	InnoDB(B tree)	RocksDB(LSM tree)	TSM tree

## III. CDI Benchmark

본 논문에서는 엷지 디바이스에서 구동 가능한 데이터베이스들을 평가하기 위해 Cross Database Integration Benchmark 라고 하는 평가 도구를 개발한다.

CDI Benchmark 의 경우 삽입 집약적 작업이 주로 이루어지는 공장 에너지 관리 시스템의 데이터 수집 시스템을 모사하기 위하여 초당 N 개의 데이터를 삽입할 수 있도록 제작하였다.

실증을 진행 중인 식품 공장에서 수집중인 데이터를 기반으로 매초마다 Bool, Double, String 타입의 센서 데이터를 수집하는 것을 가정하였다. 각각의 타입들의 값의 경우 랜덤하게 생성하여 대량 적재할 수 있도록 했다.



**Figure 1.** CDI Benchmark Architecture

그림 1. CDI Benchmark 의 구조

CDI Benchmark 의 경우 기존의 벤치마크와 평가하고자 하는 부분이 다르기 때문에 벤치마크 구성에 차이를 보인다.

첫번째로 CDI Benchmark 는 한가지 데이터베이스만을 지원하는 벤치마크가 아닌 다양한 이종 데이터베이스를 지원하는 벤치마크이다. 현재 MySQL, MyRocks, InfluxDB 를 지원하며 향후 MongoDB 와 같은 추가적인 데이터베이스에 대한 지원을 할 수 있도록 지원 여지를 생각하여 개발하였다.

두 번째로 실제 세계에서 모사를 위한 센서의 조정을 지원한다. 기존의 데이터베이스를 위한 벤치마크들은 고정 변수 만을 지원하는 경우가 대다수이다. 이는 센서 개수 조정과 변수 타입 설정을 통해 실증 전 적용할 엷지 디바이스 평가와 공장의 시뮬레이션으로 활용 가능하다.

마지막으로 기존 벤치마크와 같은 성능 테스트만을 진행하는 것이 아닌 센서 개수에 따른 현실 세계에서 적합도 검증에 위한 부하를 테스트를 진행한다. 타 벤치마크의 경우 대량 적재(Bulk Insert)만을 사용하여 데이터베이스의 부하를 검증하는 것에 집중하지만, CDI Benchmark 는 다양한 현실 기반 부하와 예상되는 센서 데이터의 타입들을 설정하여 벤치마크 하여 실증을 원하는 곳에 해당 디바이스를 적용해도 되는지에 대한 검증을 진행할 수 있다.

기존 벤치마크 대비 위와 같은 차이를 보이는 CDI Benchmark 를 각 데이터베이스의 특성에 맞도록 수정, 개선하여 개발을 진행하였다.

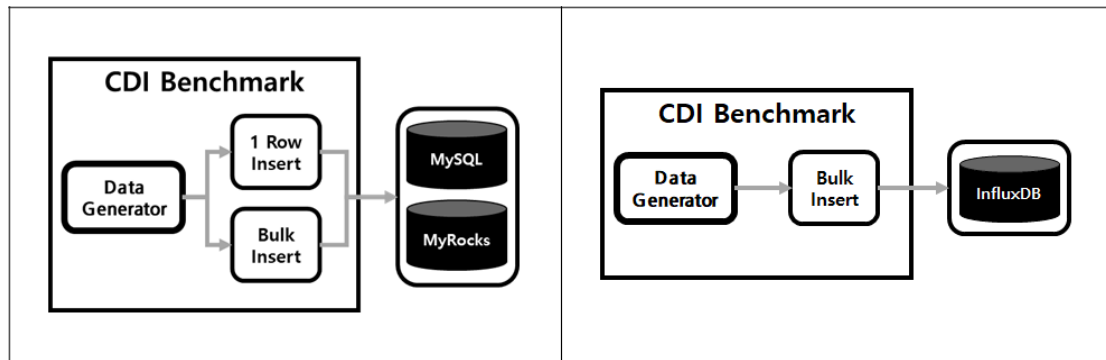


Figure 2. Each Database CDI Benchmark Architecture  
그림 2. 각 데이터베이스의 CDI Benchmark 세부 구조

CDI Benchmark 를 MySQL 과 MyRocks 에 적합하게 개발한 것은 그림 2 와 같다. 한번의 삽입 시 단일 행이 삽입되는 단일 행 삽입 방식과 부하를 테스트하기 위한 대량 적재 모드를 나누어 개발하였다. 실제 세계에서 보편적으로 샘플링한 시간당 한 행의 데이터가 도착하는 것을 토대로 기존의 부하테스트와 다른 삽입 방식으로 벤치마크를 제작하였다. 대량 적재에 대한 부분은 여러 PLC, OPC-UA 서버를 붙이는 등 병렬적으로 동작해야 하는 대형 시스템이거나 단 시간 높은 부하 발생을 확인하고자 개발하였다.

그림 2.의 우측 InfluxDB 에 대한 그림의 경우 InfluxDB 에 적합하도록 개발한 CDI Benchmark 의 구조이다. InfluxDB 는 기존에는 센서당 1 회 삽입을 발생시켰지만 데이터 전송에 대한 오버헤드를 최소화하기 위하여 한 번에 한 센서의 값만 들어오는 기존의 방식이 아닌 그림 2.와 같이 샘플링한 시간에 수집된 모든 센서 데이터를 한번 삽입하여 데이터 전송 오버헤드를 줄일 수 있도록 하였다.

본 논문에서는 세가지 데이터베이스에 최적화된 CDI Benchmark 를 사용하여 현재 실증 중인 식품 공장의 관제 포인트와 일치시키고, 각 관제 포인트의 변수 타입으로 일치시켜 실증 공장에서의 엷지 디바이스에서의 데이터베이스의 신뢰성 및 유효성을 평가하고자 한다.

## IV. Evaluation

### 4.1 평가 환경

#### 4.1.1 엣지 디바이스 사양

평가를 진행하기 위해 가장 대표적인 엣지 디바이스인 Raspberry Pi 4 model B 를 사용을 하여 평가를 진행하였다[10].

Raspberry Pi 4 의 운영체제는 Ubuntu 22.04 를 사용하였고, 저장 장치는 Samsung 의 Evo plus 256GB MicroSD 에 설치하여 구동하였다.

데이터베이스 관리 시스템은 각각 MariaDB 는 10.2 에서 스토리지 엔진을 InnoDB 와 RocksDB 로 교체하여 평가했으며 InfluxDB 의 경우 1.6.4 버전으로 평가를 진행하였다.

Table 2. Edge Device Spec.

표 2. 엣지 디바이스 사양

Model	CPU	RAM	Storage	Network
Raspberry Pi 4	BCM 2711 (ARM v8 1.8GHz)	4 GB	256GB mSD	100Mbps

데이터는 실증 중인 실제 식품 공장을 모사하기 위하여 초 단위로 데이터를 수집하였고, CDI Benchmark 를 통해 실증 공장에서 수집되고 있는 83 개의 관제점에서 1 초 주기 데이터를 모사하였다. 각각의 센서 데이터는 랜덤하게 생성된 numeric 데이터를 사용한다. 소량 데이터 삽입 평가와 단일 행 삽입 평가는 100 만개의 데이터로 평가를 진행하고, 대량 데이터 삽입 평가 시 1000 만개의 데이터, 7GB 가량의 데이터로 평가를 진행한다. 해당 데이터의 관제점은 유량, 압력, 온도이며 삽입되는 데이터는 해당 관제점에서 수집된 데이터는 실수 데이터, 정수 데이터와 시간 데이터로 이루어져 있다.

#### 4.1.3 데이터베이스 설정

각각의 데이터베이스의 기본 설정을 사용하는 것이 엣지 디바이스 자체의 성능 평가에 가장 적합할 것으로 판단하여 기본 설정을 사용하였다. 하지만 해당 데이터베이스들의 실사용을 가정하였기 때문에 압축 알고리즘은 표 3 과 같이 설정하여 평가를 진행하였다.

Table 3. Database Compression Algorithm

표 3. 데이터베이스 압축 알고리즘

	MySQL	MyRocks	InfluxDB
Compression	Zlib	LZ4	Gzip

#### 4.1.4 평가 방법

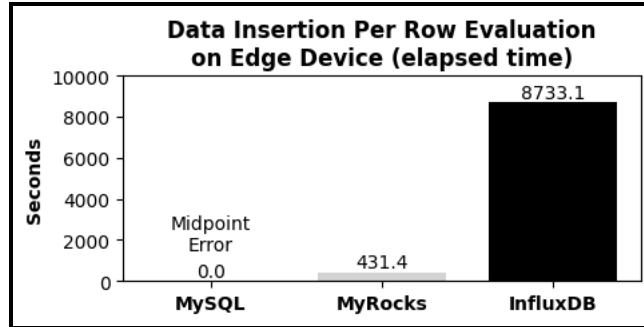
평가의 경우 MySQL, MyRocks 의 경우 단일 행 삽입 평가와 대량 적재(Bulk Insert) 평가 두 가지 평가를 모두 수행한다. 단일 행 평가의 경우 소량 데이터 삽입 평가로 평가를 진행하고, 대량 적재에 대한 평가는 소량 데이터 삽입과 대량 데이터 삽입 두 가지 평가를 진행한다. InfluxDB 는 대량 데이터 삽입 평가 한 가지로 진행한다.

## 4.2 평가 결과

#### 4.2.1 단일 행 삽입 평가

단일 행 삽입 평가를 위한 적은 부하 테스트에서는 700MB 가량의 100 만개의 실제 세계와 유사한 데이터를 적재하는 평가를 진행한다.

단일 행 삽입 평가에서는 대량 적재를 사용하지 않을 시 손해가 발생하는 삽입 소요 시간에 대한 평가를 진행한다.

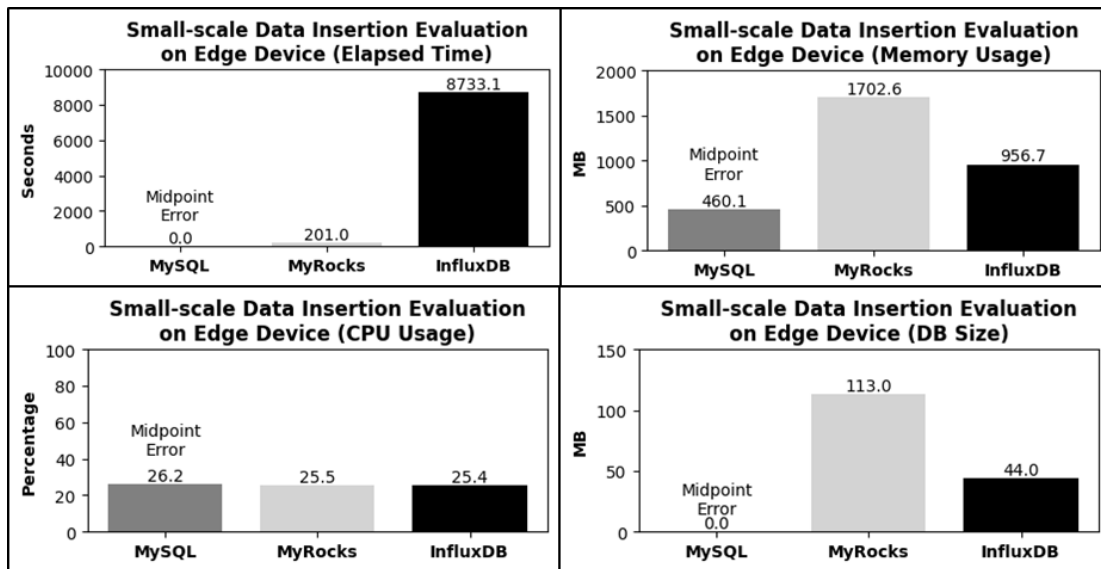


**Figure 3.** Data Insertion Per Row Evaluation Across Databases Using CDI Benchmark  
 그림 3. CDI Benchmark 를 통한 데이터베이스 별 단일 행 데이터 삽입 평가

그 결과 그림 3 과 같이 InfluxDB 대비 MyRocks 가 20 배 가량 빠른 삽입 시간을 보여주며, 쓰기 자체에 큰 이점이 있다는 것을 볼 수 있다. MySQL 은 30 만 행 전후에서 I/O 병목 현상에 의해 Table Repair 가 필요하다는 데이터베이스 오류가 발생하여 삽입이 불가능 하였다.

#### 4.2.2 소량 데이터 삽입 평가

소량 데이터 삽입의 대량 적재 평가를 위한 테스트에서는 700MB 가량의 100 만개의 실제 세계와 유사한 데이터를 적재하는 평가를 진행한다.



**Figure 4.** Small-scale Data Insertion Evaluation Across Databases Using CDI Benchmark  
 그림 4. CDI Benchmark 를 통한 데이터베이스 별 소량 데이터 삽입 평가

평가 결과는 그림 4 와 같이 소요 시간은 MySQL 에서 활용하는 대량적재를 활용하기 때문에 기존 보다 더 큰 차이를 보이는 것을 알 수 있었으며 전체적으로 MyRocks 가 가장 좋은 성능을 보이는 것을 볼 수 있었다.

첫번째로 MySQL 의 경우 단일 행 삽입 시 보다 적은 20 만 행 가량 삽입 시 Table Repair 오류로 인해 데이터를 저장하지 못하는 것을 볼 수 있었다. 대량적재를 활용하여 보다 빨라진 삽입으로 인해 InnoDB 내부에서 발생하는 In-place update 에 의해 발생하는 Small write 로 과한 I/O 트래픽이 발생하여 I/O 병목과 트랜잭션 관련 문제에 의해 테이블 문제가 발생하게 된다. 메모리와 평균 CPU 사용률은 다른 데이터베이스와 큰 차이를 지니지 않음을 볼 수 있다.

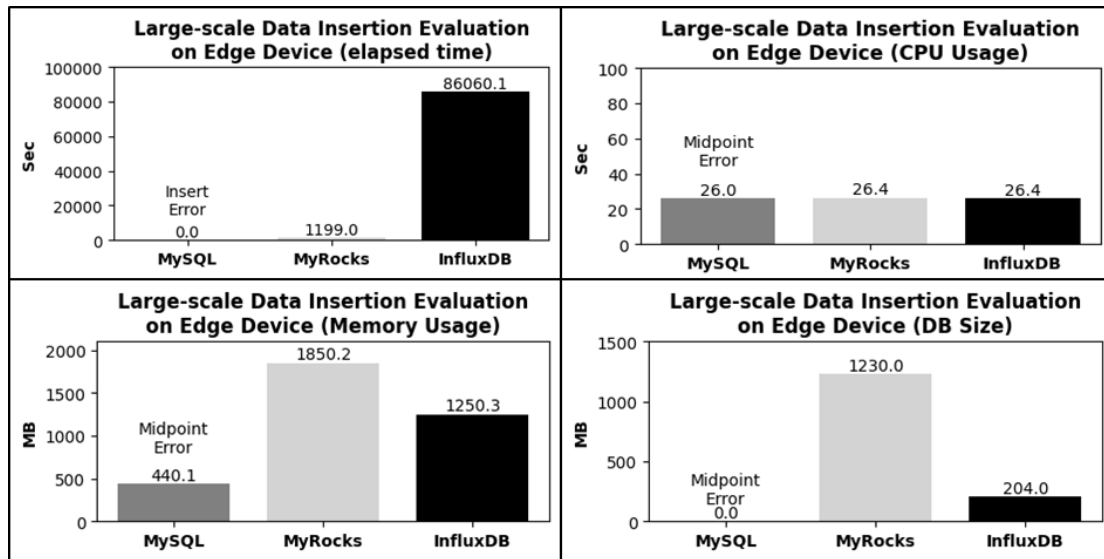
MyRocks 를 평가 시 삽입 수행시간이 매우 큰 차이가 나게 짧은 것을 볼 수 있었다. 이는 LSM tree 의 장점과 응용의 높은 자원 활용률, 낮은 요구 자원에 의한 것으로 해석된다. 평균 메모리

사용량이 InfluxDB 와 1.8 배 가량 차이나는 것을 볼 수 있는데, 이는 I/O 병목이 적고 효율적인 자원 활용이 가능하여 발생하는 것임을 알 수 있다.

InfluxDB 의 경우 데이터베이스 크기는 Gzip 압축 알고리즘에 의해 제일 작은 것을 볼 수 있다. 하지만 높은 요구 자원과 임베디드 디바이스라는 낮은 컴퓨팅 파워에 의한 압축 병목 및 I/O 병목 현상이 지속적으로 발생하였으며, 삽입 시 사용하는 API 부하 및 데이터 전송 부하에 의해서도 추가적인 부하 발생으로 매우 느린 속도를 지니는 것을 볼 수 있다.

#### 4.2.3 대량 데이터 삽입 평가

대량 데이터 삽입의 대량 적재 평가를 위한 테스트에서는 7GB 가량의 1000 만개의 실제 세계와 유사한 데이터를 적재하는 평가를 진행한다.



**Figure 5.** Large-scale Data Insertion Evaluation Across Databases Using CDI Benchmark  
 그림 5. CDI Benchmark 를 통한 데이터베이스 별 대량 데이터 삽입 평가

그림 5 와 같이 소량 데이터를 삽입하였을 때와 마찬가지로 여전히 MySQL 은 20 만행 전후에서 더 이상 구동을 하지 못하였다.

MyRocks 의 경우 80 배 가량 빠른 삽입 속도로 큰 장점이 있었으나, LZ4 라는 압축 알고리즘에 의해 InfluxDB 대비 6 배 가량 큰 데이터베이스 용량을 지니는 것을 볼 수 있었다. 이 부분은 MyRocks 의 경우 다양한 압축 알고리즘을 지원하므로 ZSTD 와 같은 알고리즘을 사용하면 성능에서 손해를 보지만, 적은 데이터베이스 용량을 얻을 수 있을 것으로 보인다.

평가 결과 관계형 데이터베이스는 임베디드 디바이스에서 데이터 수집 시스템을 구축하였을 때 채택하기 어려운 데이터베이스였으며, 소량 데이터 삽입과 대량 데이터 삽입, 단일 행 삽입 평가 1000 개 이하의 센서에서 초당 데이터 수집을 진행한다면 데이터베이스 용량을 보다 적게 차지하는 InfluxDB 을 선택할 수 있다. 하지만 대체로 성능 및 자원 활용 면에서 MyRocks 가 월등하였다. 따라서 백업 데이터베이스로 InfluxDB 를 선택하여 MyRocks 에 적재된 데이터를 InfluxDB 로 백업하여 저장하는 방식을 고려할 수 있다.

## V. 결론

이 논문은 관계형 데이터베이스인 MySQL 과 MySQL 의 Storage engine 을 RocksDB 로 대체하여 Key-value store 의 장점을 지닌 MyRocks, 시계열 데이터베이스인 InfluxDB 를 CDI



Benchmark 를 통해 엣지 디바이스에서의 안정성 및 신뢰도를 각각 비교하여 엣지 디바이스에서 안정적으로 운영할 수 있는 데이터베이스에 대해 알아보았다.

이 평가를 통해 MySQL 과 같은 RDBMS 기반의 데이터베이스는 낮은 성능의 디바이스에서 대량의 데이터가 전달될 시 문제가 발생하여 데이터베이스 자체의 동작을 보장할 수 없다는 결론을 얻게 되었다.

MyRocks 의 경우 같은 SQL 데이터베이스의 일종이지만 Backend Storage engine 이 RocksDB 라는 Key-value store 를 사용함으로 쓰기가 최적화되어 속도 면에서 가장 높은 성능을 보였다. 압축 알고리즘을 LZ4 로 선택하여 데이터베이스의 크기는 InfluxDB 보다 컸으나, 삽입 속도는 단일 행 삽입 시 20 배, 소량 삽입에서는 40 배, 대량 삽입에서는 80 배가량 빠른 것을 볼 수 있었다.

그리고 InfluxDB 의 경우 낮은 성능의 임베디드 디바이스에서 낮은 성능을 보여주지만, 데이터베이스 크기에 있어서는 압축 알고리즘을 통해 이득을 얻을 수 있는 것을 볼 수 있었다.

평가 결과 데이터 수집 시스템의 데이터베이스를 엣지 디바이스에서 구동하는 것은 삽입 연산에 장점이 없는 RDBMS 에선 신뢰성을 보장할 수 없다. 하지만 MyRocks 와 InfluxDB 는 높은 신뢰성을 보여주었으며 메인 데이터베이스는 MyRocks 와 같은 NoSQL 을 기반으로 데이터베이스로 하는 것이 성능 면에서 이득을 얻을 수 있으며, 백업 데이터베이스는 InfluxDB 와 같은 높은 압축률을 지니는 데이터베이스를 선택하는 것이 이득임을 알 수 있다.

향후 연구로 Storage engine 의 변경이 아닌 Key-value store, Document store 데이터베이스 자체의 비교 및 압축 알고리즘, Parameter 등의 최적화를 통한 성능의 손익 등에 대해 연구하며, 엣지 디바이스에 데이터 수집 시스템만이 아닌 에너지 관리 시스템의 전 주기를 사용할 수 있는지에 대한 연구를 진행하고자 한다.

## VI. 감사의 글

이 논문은 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구입니다. (No. 20202020800290)

## VII. 참고문헌

- [1] Chae-Young Lim, Chae-Eun Yeo, Hyung-Ah Lee, Sung-Yul An, Jae-Hoi Gu.(2022).Design and Implementation of IoE System for Energy Consumption Platforms.Proceedings of Symposium of the Korean Institute of communications and Information Sciences,( ),449-450.
- [2] Kim, C. W., Kim, J., Kim, S. M., & Hwang, H. T. (2015). Technological trends and case studies of factory energy management systems (FEMS) for energy saving in manufacturing industries. Equipment Journal, 44(1), 22-27.
- [3] MariaDB, “InnoDB”, <https://mariadb.com/kb/en/innodb/>
- [4] Matsunobu, Yoshinori, Siying Dong, and Herman Lee. "MyRocks: LSM-tree database storage engine serving facebook's social graph." Proceedings of the VLDB Endowment 13.12 (2020): 3217-3230.
- [5] InfluxDB, “InfluxDB”, <https://www.influxdata.com/>
- [6] Innopolis, “Data Acquisition System Market”, 2021.04.
- [7] Woojin Cho, Chea-eun Yeo, Jae-Hoi Gu, Chae-Young Lim. (2023). Evaluation of Storage Engine on Edge-Based Lightweight Platform using Sensor-OPC-UA Simulator. The Journal of the Convergence on Culture Technology (JCCT), 9(3), 803-809.
- [8] Yang, Yang, Qiang Cao, and Hong Jiang. "EdgeDB: An efficient time-series database for edge computing." IEEE Access 7 (2019): 142295-142307.
- [9] RocksDB, “RocksDB”, <https://rocksdb.org/>
- [10] Raspberry PI 4 model B, “Raspberry PI 4 model B”, <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/?variant=raspberry-pi-4-model-b-2gb>

## 저자소개

---



**조우진(Woojin Cho)**

2022년 2월 대구경북과학기술원 석사  
2022년 11월 ~ 현재 고등기술연구원 연구원

관심분야: 데이터베이스, 컴퓨터 시스템, 임베디드 디바이스, AI



**임채영(Chae-young Lim)**

2019년 02월 단국대학교 전자전기공학과 박사 수료  
2020년 11월 ~ 현재 고등기술연구원 선임연구원  
2021년 05월 ~ 현재 민군협력진흥원 S/W 분과 기획위원

관심분야 : 지능형 임베디드 S/W, 헬스케어, 디지털영상처리, 강화학습



**구재희(Jae-hoi Gu)**

1994년 2월 부경대학교 대학원 환경공학과 석사  
2008년 2월 아주대학교 대학원 환경공학과 박사

관심분야 : 데이터사이언스, AI, 네트워크 프로세서, 보안

---