

# 마이크로서비스 기반 대용량 시계열 데이터 수집 처리 시스템 연구

권동우, 지영민  
한국전자기술연구원

{dwkwon, ym.ji}@keti.re.kr

## A Study on the Microservice-based Data Acquisition System for Large Amounts of Time Series Data

Dongwoo Kwon, Youngmin Ji  
Korea Electronics Technology Institute (KETI)

### 요 약

본 논문은 다양한 데이터 특성을 가진 대용량의 실내외 환경 데이터, 설비 운용 데이터 및 에너지 사용량 등 데이터 수집을 위한 실시간 시계열 데이터 수집/처리 시스템 구조를 제안한다. 제안하는 시스템은 마이크로서비스 형태로 동작하며, 유연한 모듈형 데이터 처리 파이프라인을 통해 실시간 전처리, 가공, 감시, 다중 목적지 서버 라우팅, 프로토콜 변환, 전송 등의 기능을 제공한다. 제안하는 시스템의 성능평가 결과, 대용량 시계열 데이터에 대해 높은 읽기/쓰기 처리 성능을 나타낸다.

### I. 서 론

주거 및 상업, 공업구역 등의 실내외 환경 정보와 해당 구역에서 운용 중인 각종 설비들의 상태 정보, 에너지 사용량 정보 등을 나타내는 다양한 시계열 데이터를 수집/분석하기 위해서는 각 현장에 설치된 엣지 게이트웨이(edge gateway, EGW)로부터 전송된 계측 및 센싱 데이터를 실시간 수집 처리하는 시스템(data acquisition system, DAS)[1]이 필요하다. 이러한 DAS는 인증, 데이터 전처리, 가공, 감시, 목적지 결정, 프로토콜 변환, 전송 등의 기능을 제공해야 하며, 유연한 확장 구조를 지원해야 한다. 또한 대규모 관제점에서의 실시간 데이터 수집 처리를 위해서 대용량의 시계열 데이터 처리를 지원[2]해야 한다.

본 논문에서는 실시간으로 데이터를 수집, 처리, 변환하여 사전에 정의된 다중 목적지 서버로 처리 완료된 데이터를 전송하는 Ingress-Router (IGR) 시계열 데이터 처리 마이크로서비스를 제안한다.

IGR 마이크로서비스는 Preprocessor (PRP), Converter (CVT), Forwarder (FWD), Postprocessor (PSP) 모듈로 구성되며, 라우팅 테이블 구성에 따라 EGW로부터 전송된 데이터를 처리, 변환하여 영구 시계열 데이터베이스(TSDB)에 저장한다. 또한 데이터 분석 및 다른 응용 서비스와의 연동을 위해 메시지 브로커에 데이터를 실시간으로 발행(publish)한다. 이와 같은 서비스를 제공하기 위해서 IGR은 내부 모듈 간 순차적으로 데이터를 전달하는 데이터 처리 파이프라인을 구성한다.

마지막으로 본 논문에서는 개발된 실시간 시계열 데이터 수집 처리 시스템의 성능 평가를 수행하고 그 결과를 제시한다.

### II. 본론

EGW로부터 전송되는 센싱 데이터, 계측 데이터 등은 IGR 마이크로서비스의 수집 처리 과정을 통해 영구 TSDB에 저장된다. IGR은 TSDB로의 진입점으로서, 실시간 스트리밍 데이터 처리를 수행한다.

IGR 마이크로서비스는 EGW로부터 수신한 데이터를 처리하고 변환하여 라우팅 테이블에 기술된 목적지 서버로 데이터를 전송하며, 그림 1과 같이 크게 PRP 모듈, CVT 모듈, FWD 모듈, PSP 모듈로 구성되어 있다. PRP 모듈은 데이터의 전체 형식을 변환하거나 전송된 시계열 데이터를 기반으로 새로운 데이터를 생성하는 역할을 수행한다. CVT 모듈과 FWD 모듈은 결정된 라우팅 목적지의 프로토콜과 데이터 형식에 맞도록 데이터를 변환하고 전송하는 기능을 수행하고, PSP 모듈은 응답 메시지를 생성하고 최종 처리된 데이터들을 추적하는 기능을 제공한다.

먼저 PRP 모듈은 Data preprocessor 세부 모듈과 Data Enricher 세부 모듈로 구성된다. Data Preprocessor 모듈은 IGR 마이크로서비스로 전송된 전체 데이터 형식의 변환이 필요하거나 데이터 항목의 수정이 필요할 때 사용되는 모듈이다. 예를 들어, CSV 형식의 데이터를 JSON 형식으로 변환하거나, 센서 데이터들의 전반적인 값을 IGR 수준에서 보정할 필요가 있을 때 사용된다.

Data Enricher 모듈은 전송된 데이터를 기반으로 새로운 데이터 필드를 생성하여 추가하는 모듈이다. 예를 들어, 실내에 설치된 환경 센서 장치에서 전송된 이산화탄소 농도 데이터와 인체 감지 센서 데이터 등을 기반으로 실내 재실자 여부 및 재실자 수를 산출하는데 사용된다.

Data Preprocessor 모듈은 인라인(inline) 형식으로 IGR 내부에서만 데이터를 처리하는 것을 설계 원칙으로 한다. Data Enricher 모듈은 인라인 뿐만 아니라, 새로운 데이터 항목을 가공해내기 위해 사용되는 자원(CPU/GPU 연산, 네트워크, IO)의 소모가 커서 데이터 처리의 지연 가능성이 클 때 외부 Data Enricher 모듈로

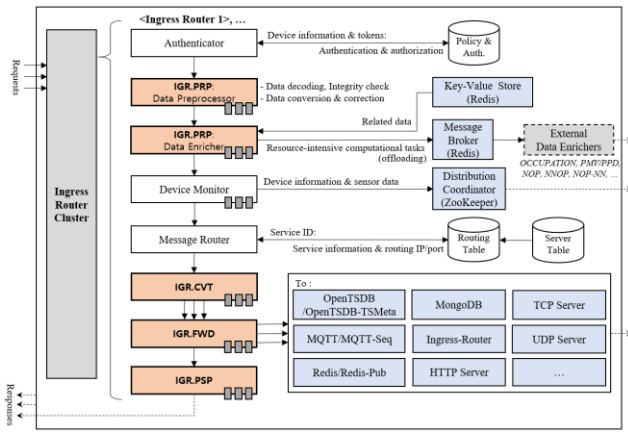


그림 1. 실시간 데이터 수집 처리 마이크로서비스 구조

데이터 처리를 오프로딩 (offloading)하는 체계를 제공한다.

이러한 자원 집약적 작업의 경우에는 IGR 마이크로서비스 내부에 특정 Enricher 모듈을 구현하여 데이터를 처리하기보다는 Redis 브로커와 같은 Pub/Sub 시스템을 활용하여 반응형 합성(reactive composition) 방식으로 외부 Data Enricher 모듈을 통하여 데이터를 처리한다.

CVT 모듈은 라우팅 테이블에서 Service ID 와 목적지 서비스 유형에 따라 파이프라인 중간 데이터의 형식을 변환한다. EGW 들은 데이터 처리 방법과 데이터 전달 목적지 대상을 식별하는 Service ID 를 가지고 있으며, Service ID 이름은 게이트웨이가 설치된 장소 또는 제공하는 서비스의 종류를 내포하도록 명명되어 있다. IGR 의 메시지 라우터는 Service ID 에 따라 라우팅 테이블을 참조하여 라우팅 목적지 서버를 결정하고 해당 목적지에 데이터를 전송한다.

FWD 모듈은 각 유입 데이터마다 부여된 Service ID 를 기반으로 결정된 라우팅 목적지 서버들의 프로토콜에 따라 데이터를 전송하는 기능을 수행한다. 전송 시에는 데이터 전송의 효율을 위해서 목적지 응용 서버별 연결 관리 및 라이브러리를 제공하는 컴포넌트인 Connection Manager 를 사용한다.

PSP 모듈은 IGR 마이크로서비스의 내부 데이터 처리 파이프라인 끝에 위치하는 모듈로 데이터 처리 및 전송 완료 후 EGW 에 데이터 처리 결과 응답 메시지를 생성하여 전송한다. 응답 메시지 생성 외에도 데이터 파이프라인 최종 처리 결과 데이터를 메시지 브로커의

QPS : 180.50	DPS : 24949.84227	(Send Data : 23.0)	completed : 575/575 (100.00%)	failed : 0/575 (0.00%)
QPS : 147.62	DPS : 21742.56799	(Send Data : 24.0)	completed : 600/600 (100.00%)	failed : 0/600 (0.00%)
QPS : 185.01	DPS : 28650.80743	(Send Data : 25.0)	completed : 625/625 (100.00%)	failed : 0/625 (0.00%)
QPS : 120.44	DPS : 22497.71508	(Send Data : 26.0)	completed : 650/650 (100.00%)	failed : 0/650 (0.00%)
QPS : 200.56	DPS : 24151.52521	(Send Data : 27.0)	completed : 675/675 (100.00%)	failed : 0/675 (0.00%)
QPS : 410.54	DPS : 28205.62889	(Send Data : 28.0)	completed : 700/700 (100.00%)	failed : 0/700 (0.00%)
QPS : 100.32	DPS : 28421.16142	(Send Data : 29.0)	completed : 725/725 (100.00%)	failed : 0/725 (0.00%)
QPS : 195.39	DPS : 25506.51838	(Send Data : 30.0)	completed : 750/750 (100.00%)	failed : 0/750 (0.00%)
QPS : 145.56	DPS : 23582.98189	(Send Data : 31.0)	completed : 775/775 (100.00%)	failed : 0/775 (0.00%)
QPS : 164.15	DPS : 27422.38145	(Send Data : 32.0)	completed : 800/800 (100.00%)	failed : 0/800 (0.00%)
QPS : 145.66	DPS : 23769.13821	(Send Data : 33.0)	completed : 825/825 (100.00%)	failed : 0/825 (0.00%)
QPS : 120.44	DPS : 22497.71508	(Send Data : 34.0)	completed : 850/850 (100.00%)	failed : 0/850 (0.00%)
QPS : 130.17	DPS : 23777.45521	(Send Data : 35.0)	completed : 875/875 (100.00%)	failed : 0/875 (0.00%)
QPS : 150.22	DPS : 25135.32452	(Send Data : 36.0)	completed : 900/900 (100.00%)	failed : 0/900 (0.00%)
QPS : 145.29	DPS : 25083.08951	(Send Data : 37.0)	completed : 925/925 (100.00%)	failed : 0/925 (0.00%)
QPS : 164.07	DPS : 31443.74428	(Send Data : 38.0)	completed : 950/950 (100.00%)	failed : 0/950 (0.00%)
QPS : 202.82	DPS : 24935.43772	(Send Data : 39.0)	completed : 975/975 (100.00%)	failed : 0/975 (0.00%)
QPS : 236.47	DPS : 29407.22170	(Send Data : 40.0)	completed : 1000/1000 (100.00%)	failed : 0/1000 (0.00%)
QPS : 191.09	DPS : 25083.08951	(Send Data : 41.0)	completed : 1025/1025 (100.00%)	failed : 0/1025 (0.00%)

Start time : 2022-10-25 14:36:18.060515 | End time : 2022-10-25 14:36:22.302354 | Running time : 0:00:04.241839  
Data Points : 101475 | Average DPS : 26967.24 | Real DPS : 23922.407238935757

그림 3. 시계열 데이터 쓰기 처리 성능 측정

특정 채널에 발행하여 데이터 처리 과정을 추적하는 기능도 함께 제공한다.

본 논문에서 제안한 IGR 마이크로서비스를 개발하고, 실제 운영 중인 공장 에너지 관리 시스템에 배포하여 대용량 시계열 데이터에 대한 읽기 및 쓰기 성능 평가를 수행하였다. 읽기 성능 시험은 실증 공장의 10 개 관제점을 대상으로 관제점별 TSDB 데이터 질의 후, 처리된 데이터 건수를 처리 소요 시간으로 나누어 초당 처리 건수를 산출하였다. 그 결과, 데이터 초당 조회 건수가 369,003 건으로 우수한 성능을 보였다. 그림 2 는 평가대상 관제점 중 두 개 관제점에 대한 읽기 성능 평가 결과를 나타내고 있다.

쓰기 성능 시험은 관제점 데이터 생성기에서 25 개 관제점에 대해 10 만 개의 모사 데이터를 생성 후 IGR 마이크로서비스로 전송하여 진행되었다. 전송된 데이터는 표준 공장 관리 태그 형태로 변환된 뒤, TSDB 에 저장된다. 이 때, 처리된 데이터 건수를 실제 소요 시간으로 나누어 데이터 쓰기 처리 성능을 산출하였다. 이와 같은 방법으로, 쓰기 시험 수행 결과, 그림 3 과 같이 101,475 건의 시험 데이터에 대해 초당 23,922 건의 높은 쓰기 성능 결과가 산출되었다.

### III. 결론

본 논문에서는 각종 환경 데이터, 설비 운용 데이터 등 대용량 시계열 데이터를 수집 처리하기 위해서 실시간 데이터 수집 처리 시스템 구조를 제안하고 마이크로서비스 형태로 개발하여 실제 실증 공장 환경에 배포한 후 성능 평가를 수행하였다. 평가 결과, 제안하는 시스템은 대용량 시계열 데이터에 대해 높은 읽기/쓰기 처리 성능을 보였으며, 모듈형 확장 구조를 통해 프로토콜 및 데이터 형식에 제약이 없는 유연한 체계를 제공하고 있다.

### ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20202020900290).

### 참 고 문 헌

- [1] L. Angrisani, M. D'Arco, and D. Grillo, "Using Data Acquisition Systems for Time Period Measurements Based on Averaging Approaches," Proc. IEEE International Forum on Research and Technology for Society and Industry (RTSI), Sep. 2021.
- [2] D. Kwon, K. Ok, and Y. Ji., "IBFRAME: IoT Data Processing Framework for Intelligent Building Management," Proc. IEEE Big Data, Dec. 2019.

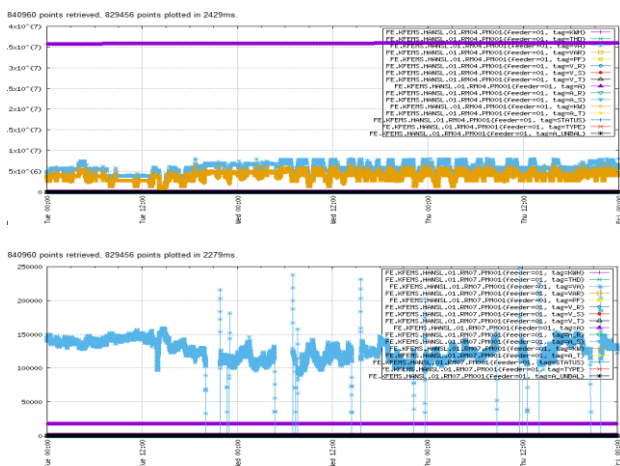


그림 2. 시계열 데이터 읽기 처리 성능 측정