

효율적인 IoT 데이터 수집을 위한 분산 저장 에지 노드 설계 및 데이터 파이프라인에 관한 연구

오흥기, 박상현, 김예원, 김종익, 최훈
충남대학교

ohk1119@o.cnu.ac.kr, sanghyun.park@cnu.ac.kr, yewonez@cnu.ac.kr, jongik@cnu.ac.kr, hc@cnu.ac.kr

A Study on Distributed Data Storing Edge Node and Data Pipeline for Efficient IoT Data Acquisition

Honggi Oh, Sanghyun Park, Yewon Kim, Jongik Kim, Hoon Choi

Chungnam National University

요약

최근 다양한 IoT 장비들이 발달함에 따라 장비들로부터 수집되는 데이터를 처리하기 위해서 확장된 에지 노드 구성이 개발 중이다. 본 논문에서는 에지 노드에 오픈 소프트웨어인 MinIO를 탑재한 분산 시스템으로 구성된 에지 노드를 구성함으로써 빠르게 데이터를 수집하고, 각 에지 노드에서 수집한 데이터를 처리하여 유의미한 데이터가 발생할 경우 선별하여 클라우드 서버까지 전송할 수 있는 파이프라인이 설계된 IoT 시스템을 기술한다.

I. 서론

IoT (Internet of Things) 또는 사물 인터넷은 네트워크를 통해 연결된 장치들 및 장치와 장치, 장치와 서버 간의 통신을 용이하게 하는 기술을 의미한다 [1]. 다양한 센서 장비들의 등장과 고대역폭 통신 기술의 발전으로 인해 스마트 홈, 스마트 팩토리, 스마트 팜 등 IoT를 이용하는 다양한 서비스가 등장하고 이를 효율적으로 사용하기 위한 연구가 활발히 진행되고 있다.

IoT 장치들 사이의 직접 통신을 이용하여 사용자에게 유용한 서비스를 제공하기도 하지만 대부분 서비스에서 중앙 서버 역할을 수행하는 컴퓨팅 장비를 통해 IoT 장치들로부터 전송되는 데이터를 저장하고 처리하는 것이 일반적이다. IoT 장치의 데이터를 수집 처리하기 위해 클라우드 서비스 등을 활용할 수도 있지만, IoT를 사용하는 응용은 실시간 응답이 매우 중요하여 클라우드 서비스를 이용하여 데이터를 수집하는 경우 실시간 응답을 저하시킬 수 있는 문제를 가지고 있다 [2].

이러한 문제를 해결하기 위해 에지 컴퓨팅 기술[3]을 사용한다. 에지 컴퓨팅 기술은 실시간 응용을 지원하기 위해 IoT 장비 내부 또는 에지 노드라고 불리는 IoT 장비와 근접한 위치의 지역 서버에서 IoT 데이터를 수집 처리하는 기술이다. 하지만, 대규모의 IoT 장비들로부터 수집되는 데이터를 처리하기 위해서는 에지 노드의 규모가 커져야 하며 확장성을 보장해야 하는 문제가 존재한다. 에지 노드의 초기 구축 비용을 예측하고 향후 에지 노드를 어떠한 규모로 확장해야 하는지를 결정하는 것은 매우 어려운 문제이다. 또한, 기계학습 등 큰 규모의 컴퓨팅 파워가 있어야 하는 작업을 에지 노드에서 수행하는 것은 한계가 있다.

이러한 문제는 최소 규모의 에지 노드를 하나의 에지 레이어로 묶어 구축하여 실시간 응답을 요구하는 응용을 처리하고 더 많은 컴퓨팅 자원을 활용하기 위해 에지 노드가 클라우드 서버에 연결되는 구조로 확장하여 해결할 수 있다. 그리고 그 과정에서 에지 레이어에 데이터를 효율적으로 저장하는 분산 시스템을 도입함으로써 데이터 저장 속도를 개선할 수 있다.

본 논문은 IoT 장비 - 에지 노드 - 클라우드 서버의 구조를 가지는 시스템에서 IoT 장비로부터 전달되는 데이터를 효과적으로 수집하기 위한 에

지 노드의 구조를 제안한다. 특히, 분산 시스템을 지원하는 오픈 소프트웨어인 MinIO 오브젝트 스토리지[4]를 에지 노드들에 도입하여 데이터를 수집하고, 에지 컴퓨팅 과정을 통해 클라우드에서 필요로 하는 유의미한 데이터를 선별하여 전송하는 과정을 포함하는 에지 노드의 데이터 파이프라인을 고안하였다.

본 논문은 다음과 같이 구성된다. II 장에서는 본 논문에서 제안하는 에지 노드의 구조를 설명한다. III 장에서는 에지 노드에서의 효율적인 데이터 수집 및 전송을 위한 기법을 설명한다. IV 장에서는 본 논문에서 제안한 기법을 활용하는 경우 데이터 수집 속도가 얼마나 향상되는지 실험을 통해 보여준다. V 장에서는 결론 및 향후 연구를 기술한다.

II. IoT 서비스를 위한 에지 노드의 구성

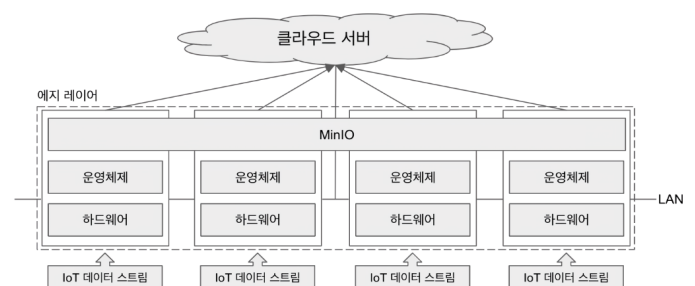


그림 1. 에지 노드 구성

그림 1은 본 논문에서 제안하는 시스템의 전체 구조를 보여준다. 각종 센서와 에지 노드를 연결하여 각 센서에서 나오는 IoT 데이터들이 스트림으로 각 에지 노드에 흘러 들어간다. 에지 노드들은 대규모의 IoT 장비들로부터 데이터를 수신받기 위해 각각 데이터를 수신 및 저장하는 MinIO를 탑재한다. 다량의 IoT 데이터를 분산 저장한 이후 각 에지 노드들은 그 데이터를 분석 혹은 처리하는 과정을 거치며, 필요에 따라 클라우드 서버로 선별된 데이터를 전송한다.

III. 데이터 파이프라인

클라우드 서비스의 실시간 응답이 저하될 수 있는 문제를 방지하기 위해서는 데이터의 수신 및 저장 과정이 개선되어야 한다. 본 논문은 MinIO를 데이터 파이프라인의 핵심으로 사용함으로써 대용량의 데이터를 수집하는 과정에서 에지 노드들의 분산 저장을 통해 저장 과정의 속도를 개선하는 것에 초점을 두고 있다. 이를 통해 특정 에지 노드에 과부하를 초래할 수 있는 대용량 데이터를 효과적으로 분산 저장할 수 있게 된다.

그림 2는 본 논문에서 제안한 IoT 시스템의 각 에지 노드에서 진행되는 데이터 파이프라인의 전체 구조를 보여준다. 제안하는 에지 노드의 파이프라인 구조에서 그림 2의 ①은 IoT 기기에서 발생한 데이터들을 MinIO를 통해 에지 레이어의 각 에지 노드에 분산 저장되는 과정을 말한다. MinIO를 이용하여 다량의 데이터를 분산 저장함으로써 에지 노드들은 빠르게 데이터를 수집 및 저장할 수 있음은 물론, 효과적으로 데이터를 관리하는 것을 가능하게 한다. 그중에서도, 데이터 중복 (replication) 기능을 통해 디스크 손상으로 인한 데이터의 유실을 방지할 수 있게 되며, 암호화 (encryption) 기능을 통해 데이터의 전송 과정에서의 데이터의 무결성을 지켜줄 수 있게 된다. 그림 2의 ②는 에지 노드에 저장된 데이터 중 유의미한 데이터만을 클라우드 서버에 전송하기 위해 에지 컴퓨팅을 진행하는 과정이다. 마지막으로 그림 2의 ③은 에지 컴퓨팅을 통해 얻은 데이터가 클라우드 서버의 처리를 받을 만큼 의미 있는 경우나 혹은 클라우드 서버에서 데이터를 원하는 경우, 이를 클라우드 서버로 전송하는 과정이다.

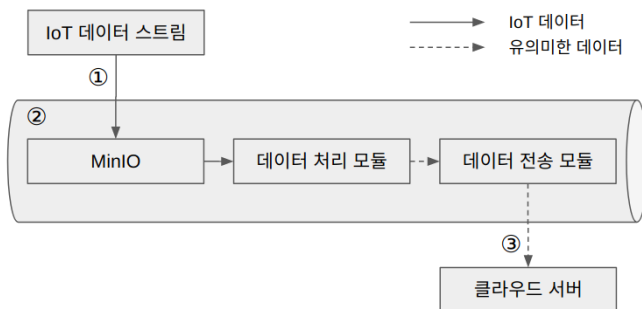


그림 2. 데이터 파이프라인 개요

IV. 실험

표 1은 에지 노드에 대한 실험 환경이다. 동일한 에지 노드 4개를 사용하였으며, MinIO(release.2022-10-02T19-29-29Z)를 이용한 에지 노드에 대한 성능 실험을 위해 벤치마킹 툴인 WARP(v0.6.3 - d691c32)를 사용하여 진행하였다.

표 1. 실험 환경

에지 노드	운영체제	Ubuntu LTS 20.0.4
	RAM	Samsung DDR4 16 GB
	CPU	12th Gen Intel(R) Core(TM) i7-12700K
	파일 시스템	XFS
	저장 장치	TOSHIBA HDWD120 2TB
	평균 네트워크 속도	1 GiB/s

실험을 위해서 단일 모드와 분산 모드의 경우 하나의 노드에 MinIO 서버를 작동시키고, 분산 모드의 경우 4대의 에지 노드에 MinIO 서버를 작동시킨다. 그리고 MinIO 서버에게 데이터를 전달할 WARP 클라이언트를 각 에지 노드에서 실행하여 WARP 서버의 명령을 기다린다. 이후 한 에지 노드에서 WARP 서버의 명령을 통해 실험을 진행하였다.

실험 데이터는 WARP의 실험 과정에 따라 16 MiB 크기의 랜덤한 데이

터 스트림 객체를 1분 동안 MinIO 서버가 얼마나 수집할 수 있는지에 대해서 실험하였으며, 데이터를 에지 노드로 전송하는 PUT과 에지 노드로부터 데이터를 가져오는 GET 과정으로 나누어 진행하였다.

표 2. 실험 결과

	GET/PUT	WARP 클라이언트	평균 속도	데이터 스트림 객체 처리량
단일 모드	GET	1	66.45 MiB/s	4.35 obj/s
	PUT	1	72.48 MiB/s	4.75 obj/s
분산 모드	GET	1	111.98 MiB/s	7.34 obj/s
	PUT	1	86.22 MiB/s	5.65 obj/s
	GET	4	231.07 MiB/s	15.14 obj/s
	PUT	4	112.61 MiB/s	11.26 obj/s

표 2는 단일 모드와 분산 모드에 대한 실험 결과를 나타낸다. 단일 모드와 비교하였을 때 분산 모드의 속도가 GET의 경우 약 3.47 배, PUT의 경우 약 1.55 배 빠른 성능을 보였다. 하드웨어 자원의 제약으로 WARP 클라이언트를 에지 노드에서 구동하였음에도 이와 같은 좋은 성능이 나온 것은 본 논문이 제안하는 에지 노드의 효율성을 증명한다. 특히, 주로 클라우드 서버에서 사용하는 분산 시스템을 에지 레이어에서 사용함으로써 다량의 데이터를 각 에지 노드에서 빠르게 수집할 수 있다는 점이 주목할 만하다.

실험 결과는 분산 시스템에 데이터를 생성하여 전송하는 WARP 클라이언트의 개수에 따라서도 달라지는데, WARP 클라이언트가 한 개인 경우에 비해 에지 노드 개수와 같은 4개인 경우 속도가 더 빠른 것을 볼 수 있다.

V. 결론

본 논문에서는 대규모의 IoT 장비들로부터 생성되는 데이터를 효율적으로 저장할 수 있는 에지 노드 및 데이터 파이프라인을 제시하였다. 각 에지 노드들에 오픈 소프트웨어인 MinIO를 탑재함으로써 대용량의 데이터를 고속 분산 저장하였다.

향후 IoT 서비스가 발달함에 따라 다량의 데이터에 대한 응답을 실시간으로 받기 위해서 에지 노드의 효율적인 데이터 저장 및 처리 과정의 발전이 지속해서 필요할 것으로 보인다. 더불어, 분산 저장 시스템을 사용하면 발생할 수 있는 통신 오버헤드 등으로 인한 성능 저하를 해결하는 것 역시 중요한 향후 연구라고 볼 수 있다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2022-0-00817)

참 고 문 헌

- [1] Atzori, Luigi, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey." *Computer networks* 54.15 (2010): 2787-2805.
- [2] Schulz, Philipp, et al. "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture." *IEEE Communications Magazine* 55.2 (2017): 70-78.
- [3] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," in *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637-646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.
- [4] "Open source object storage system MinIO", [online] Available: <https://min.io/>.