

<https://doi.org/10.7236/IIBC.2018.18.4.219>

IIBC 2018-4-30

IIoT 미들웨어 플랫폼을 활용한 연속 제조공정의 환경센서 빅데이터 정제시스템

Big Data Refining System for Environmental Sensor of Continuous Manufacturing Process using IIoT Middleware Platform

윤여진*, 김태형**, 이준희***, 김영곤****

Yeo-Jin Yoon*, Tea-Hyung Kim**, Jun-Hee Lee***, Young-Gon Kim****

요약 산업용 사물인터넷(IIoT:Industrial Internet of Thing)은 기존의 공정의 자동화란 범주를 넘어 모든 제조공정을 정보화 하는 것을 의미한다. 또한 각 공정에 설치된 센서로 부터 수집되는 데이터를 토대로 정보화 시스템을 구축하여 각 공정을 실시간으로 관리하고 자동화하여 최적의 생산성을 유지하는데 그 목적을 두고 있다. 각 공정의 센서로 부터 수집되는 데이터는 비정형성을 띄고 있으며 이러한 비정형데이터를 효과적으로 수집하고 처리하기 위해 많은 연구가 이루어지고 있다. 본 논문에서는 효과적인 빅데이터 수집 및 처리를 위하여 미들웨어로 Node-RED를 사용한 시스템을 제안하였다.

Abstract IIoT(Industrial Internet of Thing) means that all manufacturing processes are informed beyond the conventional automation of process automation. The objective of the system is to build an information system based on the data collected from the sensors installed in each process and to maintain optimal productivity by managing and automating each process in real time. Data collected from sensors in each process is unstructured and many studies have been conducted to collect and process such unstructured data effectively. In this paper, we propose a system using Node-RED as middleware for effective big data collection and processing.

Key Words : Smart Factory, Industrial Internet of Things, Big Data, Node-RED

1. 서 론

산업용 사물 인터넷(IIoT:Industrial Internet of Thing)은 기존의 공정의 자동화란 범주를 넘어 모든 제조공정을 정보화 하는 것을 의미한다. 이는 일련의 제조공정에 설치된 센서로 부터 수집되는 데이터를 토대로 정보화

시스템을 구축하여 각 공정을 실시간으로 관리, 모니터링하고 나아가 환경, 사람, 자재, 재고, 생산관리 시스템과의 통합을 통해 최적화된 생산성을 유지하는 것에 목적을 두고 있다. 이를 위해 공장 내 모든 센서를 통합 연결하고 센서로 부터 데이터를 수집하고 효율적인 빅데이터 분석을 실시해 최종적으로 분석 결과를 사용자에게

*준회원, 한국산업기술대학교 컴퓨터공학과

**준회원, 한국산업기술대학교 컴퓨터공학과

***준회원, 한국산업기술대학교 컴퓨터공학과

****정회원, 한국산업기술대학교 컴퓨터공학과

접수일자 2018년 5월 10일, 수정완료 2018년 7월 10일

게재확정일자 2018년 8월 10일

Received: 10 May, 2018 / Revised: 10 July, 2018 /

Accepted: 10 August, 2018

*Corresponding Author: ykkim@kpu.ac.kr

Dept of Computer Engineering, Korea Polytechnic University,
Korea

시각화하여 보여주거나 기계학습 기반 데이터로써 사용하기 위한 많은 연구가 진행되고 있다.

각 공정의 센서로 부터 수집되는 데이터는 비정형성을 띄고 있다. 하지만 물리 저장소의 용량 한계, 비용, 속도 등의 시스템의 제약으로 인해, 방대한 입력 받는 비정형 센서 데이터를 모두 수집하기에는 어려움이 따른다. 또한 수집된 비정형데이터를 분석하기 용이한 정형데이터로 변환하는데도 많은 시간과 자원이 소모된다.

본 논문에서는 효과적인 빅데이터 수집 및 처리를 위해 공장의 연속된 제조공정의 설비 및 라인별 구획을 구분하여, 공정 환경 모니터링 센서를 연결하고 효과적인 데이터 수집 및 처리를 위해 미들웨어로 Node-RED를 사용한 시스템을 제안한다.

II. 관련 연구

1. MQTT

MQTT 프로토콜은 보편성, 유연성, 경량성, 신속성을 가지며 메시지 전달의 신뢰도와 보안성이 있는 양방향 메시징 프로토콜이다^[1]. 원격 검침기는 대부분 소형이며 통신 대역폭과 전원이 한정적인 환경에서 동작한다. 이는 배터리 용량이 제한적이고 통신 품질을 일정 수준으로 유지하기 어렵다는 점에서 스마트폰 환경과 매우 유사하다. MQTT 데이터 송수신 구조는 그림 1과 같다.

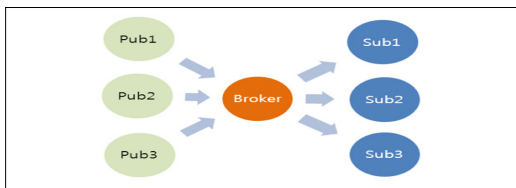


그림 1. MQTT 데이터 송수신 구조
Fig. 1. MQTT data transmission/reception structure

발행자는(Pub) 토픽을 발행하고, 구독자(Sub)는 관심 있는 토픽을 구독하는 구조이다.

MQTT 개발 시 고려한 원격 검침 환경에서는 온도, 정보 등의 검침 정보가 주기적으로 발생된다. 발생되는 검침 환경에서 오류가 발생할 시 오류 전달의 우선순위를 두어 전달하는 것이 효과적이다. 이러한 구조를 구현하기 위해 MQTT에서는 QoS(Quality of Service) 수준을 0, 1, 2의 세 단계로 정의한다. QoS 수준은 표 1과 같다.

표 1. QoS 수준

Table 1. QoS level

QoS 수준	정의
QoS 0	한 번만 전달하고 전달 여부는 확인하지 않음
QoS 1	적어도 한 번 이상 전달하고 전달여부 확인
QoS 2	4단계의 핸드셰이킹(Handshaking)을 통해 정확히 한 번만 전달

2. Node-RED

Node-RED는 IBM에서 개발한 대표적인 툴로서 제안 플랫폼의 기능과 플로우 매쉬업을 시각화하여 사용자에게 보여준다^[2]. 또한 현재 로컬영역의 사물인터넷 플랫폼인 AllJoyn을 통합하여 로컬에 연결된 디바이스 간 서비스 플로우 설정이 가능하다.

Node-RED는 사물인터넷을 위한 Front-End 비주얼 툴이다. Node-RED에서 각 디바이스 또는 프로토콜이 입력과 출력을 가지는 노드로 추상화되어 있다. 사용자는 웹을 이용하여 이러한 노드를 서로 연결하여 새로운 서비스 또는 데이터 처리 루틴을 만들 수 있다. 또한 Node.js 기반으로 개발된 Node-RED는 Node.js가 제공하는 많은 라이브러리를 개발자가 사용 가능하고 쉽게 새로운 노드를 추가 할 수도 있다^[3]. Node-RED의 특징은 표 2와 같다.

표 2. Node-RED 특징

Table 2. Node-RED Features

No.	특징
1	사물인터넷 응용을 제작하는 비주얼 도구
2	간단한 런타임 배포, 시작품 제작에 적합
3	간단한 자동 실행 런타임을 쉽게 작성
4	다양한 연동을 간단하게 그려서 확장 가능
5	낮은 진입 장벽으로 누구나 쉽게 배우고 사용 가능
6	개방형 표준, 유연성, 공유

3. 빅데이터 분석

빅데이터는 데이터의 양(Volume)에서는 대용량(High Volume)을, 데이터의 종류(Variety)에 관해서는 비정형성(High Variety)을, 데이터의 속도(Velocity)와 관련해서는 빠른 속도(High Velocity)를 구비하는 것을 그 기술적 특징으로 한다^[4]. 최근 이러한 빅데이터를 분석하는 기술들이 각광받고 있다.

빅데이터 처리를 위한 과정은 기본 5단계로 진행된다^[5]. 빅데이터 처리 과정은 그림 2와 같다.

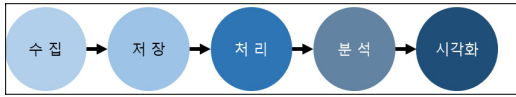


그림 2. 빅데이터 처리 과정
 Fig. 2. Big Data Processing

빅데이터를 수집하고 처리하여 분석하기 위해 많은 미들웨어들이 사용되고 있다. 하지만 빅데이터를 수집하고 저장하기에 물리적으로 큰 용량의 저장소가 필요할 뿐만 아니라 수집한 데이터에서 분석을 위한 데이터를 추출하는데 많은 시간이 소요된다. 즉 빅데이터를 효율적으로 수집하고 처리하는 것이 현재 빅데이터 분석의 최우선 과제이다.

III. 시스템 설계 및 시뮬레이션

1. 전체 시스템 구성도

각 공정의 데이터를 수집하기 위해 본 논문에서는 온도센서, 습도센서, 소음센서를 모듈화 하여 공정의 주요 설비에 부착하였고, 라즈베리파이에 이를 수집하는 서버를 구축하고 Node-RED를 통해 각각의 컴포넌트들을 연결하였다. 전체 시스템 구성도는 그림 3과 같으며 센서모듈 하드웨어는 그림 4와 같다.

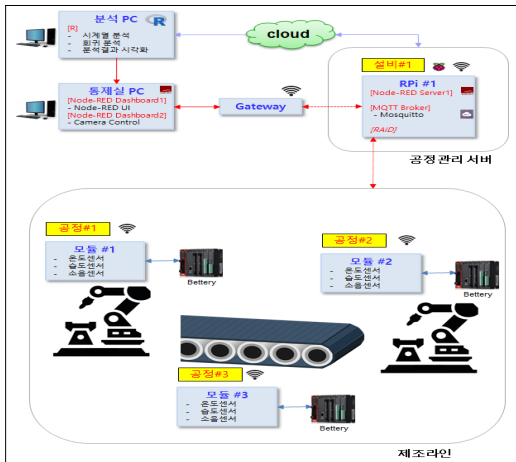


그림 3. 전체 시스템 구성도
 Fig. 3. Overall System Configuration Diagram

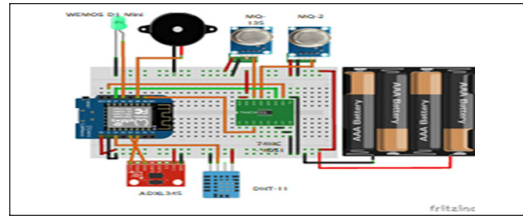


그림 4. 센서모듈 하드웨어
 Fig. 4. Sensor Module Hardware

센서모듈은 74HC4051 멀티플렉서 IC를 사용하여 입력된 아날로그 센서 값을 디지털 값으로 변환하고, 수집된 센서의 데이터를 와이파이를 통해 서버로 송신하는 NodeMCU는 아두이노 기반의 WEMOS D1 MINI를 사용하였다. 또한 제조공정의 각종 설비에 탈부착하여 데이터수집이 용이하고 추가적인 배선작업이 불필요 하도록 전원은 배터리만으로 동작가능하게 구성하였다.

2. 시스템 네트워크 구성도

센서에서 수집되는 데이터는 기본적으로 스트림 형태의 비정형데이터일 뿐만 아니라 시간이 지남에 따라 수집할 데이터를 저장할 대규모 저장소가 필요하다. 또한 수집된 데이터는 분석하기 용이한 정형데이터 형태로 변환한 뒤 분석에 필요한 데이터를 정제하기 위한 작업들이 필요하다. 본 논문에서는 이러한 데이터 수집, 처리 단계를 간소화하기 위해 미들웨어인 Node-RED를 사용하여 Json 포맷으로 변환하였고, 데이터를 저장할 공간을 확보하기 위해 Json 파일을 Cloud에 구성된 NoSQL DB에 저장하였다. 또한 연결된 IoT의 데이터를 받아 실시간으로 그래프 및 시각화 할 수 있도록 구현하였다. 제안 시스템 네트워크 구성도는 그림 5와 같다.

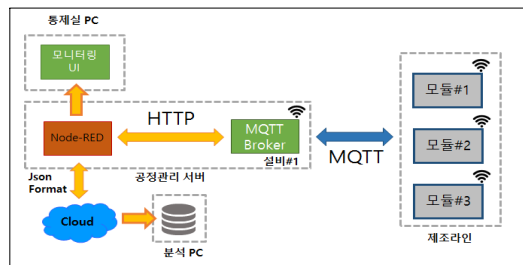


그림 5. 제안 시스템 네트워크 구성도
 Fig. 5. Suggested system network diagram

센서모듈에서 측정된 데이터 값은 MQTT 브로커를 통해 서버로 전송되고, 이는 미들웨어인 Node-Red를 통해 Cloud NoSQL DB인 MongoDB로 전송, 저장되며 동시에 사용자가 직관적으로 파악 가능하도록 HTTP를 통해 웹페이지로 시각화된다. 전체 시스템 구현 화면은 그림 6과 같다.

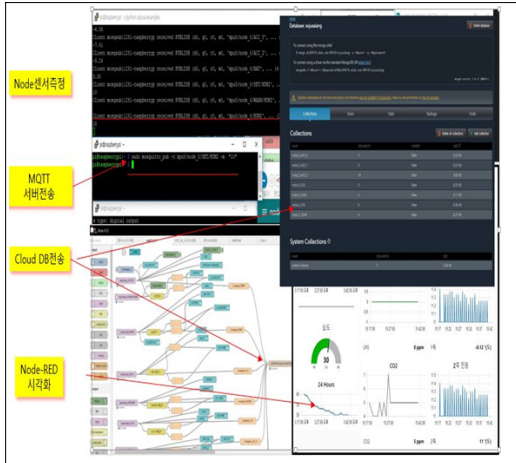


그림 6. 전체 시스템 구현 화면
Fig. 6. Overall system implementation screen

3. 빅데이터 수집 및 처리

본 논문에서는 빅데이터 수집 및 처리를 위해 Node-RED를 사용하였다. 본 시스템은 MQTT포맷을 사용하여 각 센서의 데이터를 수신하였는데, 이는 필요한 데이터를 설정한 시간에만 수집하기 때문에 매 초마다 지속적으로 데이터를 수집하는 기존 방법에 비해 분석에 필요한 데이터만을 수집한다는 장점이 있다. 이러한 장점은 곧 저장하는 전체 데이터의 양이 줄어들 뿐만 아니라 분석에 필요한 데이터만을 선별적으로 수집하기에 빅데이터 분석 과정을 수행하기 전 단계들에서 소요되는 시간을 대폭 줄일 수 있고 저장되는 데이터양이 현저하게 적다는 장점이 있다.

Node-RED를 사용한 본 논문의 시스템의 데이터 전체량을 비교하기 위하여, Wireshark를 사용하여 실시간 스트림 데이터를 수집하고, Hadoop HDFS에 데이터를 분산 저장, 저장된 데이터에서 분석에 필요한 데이터를 정제하기 위해 Hive와 Pig를 사용한 시스템을 사용하였다. 비교 대상 미들웨어 수집, 처리단계 구성도는 그림 7과 같다.

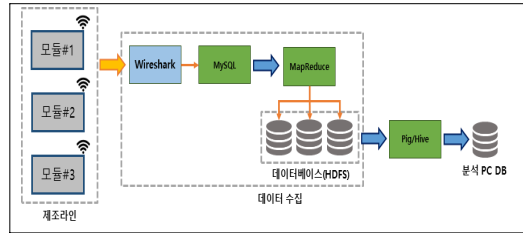


그림 7. 비교 대상 미들웨어 수집, 처리 단계 구성도
Fig. 7. Comparison target middleware collection and processing step configuration diagram

4. 빅데이터 수집 비교 모델

비교 모델에서 각 모듈의 센서 데이터는 TCP 포맷으로 수집되며, 연속적인 스트림 데이터 형태로 수집된다. 이러한 비정형 센서 데이터를 수집하기 위하여 Wireshark를 사용하여 스트림 데이터를 수집하고 분석에 용이한 형태인 정형데이터로 변환하기 위해 MySQL을 사용하여 데이터를 1차 처리하였다. Wireshark를 통한 비정형 센서 데이터 수집은 그림 8과 같고, MySQL을 사용하여 변환된 정형 데이터는 그림 9와 같다.

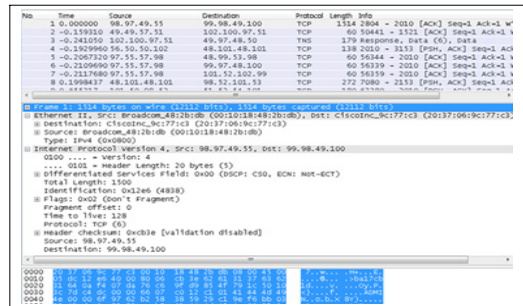


그림 8. Wireshark를 통한 비정형 센서 데이터 수집
Fig. 8. Unstructured sensor data collection via Wireshark

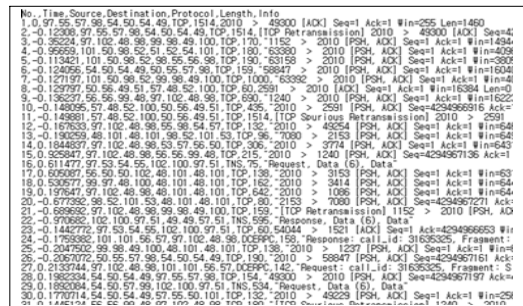


그림 9. MySQL을 사용하여 변환된 정형 데이터
Fig. 9. Transformed structured data using MySQL

실시간으로 수집되는 데이터 특성상 하나의 저장소에 데이터를 저장하기에 저장소의 크기가 커야한다. 이러한 문제를 해결하기 위해 MySQL을 통해 변환된 정형 데이터는 분산 데이터 저장을 지원하는 Hadoop ECO System의 HDFS를 저장장치로 사용하였고, 분산저장을 관리하는 MapReduce를 통해 수집된 데이터를 HDFS에 분산 저장하였다.

저장된 정형 데이터 중 분석에 필요한 데이터만을 추출하기 위하여 Hadoop의 Pig와 Hive를 통해 데이터를 2차 처리하였다. Pig를 사용한 데이터 처리는 그림 10과 같고, Hive를 사용한 데이터 처리는 그림 11과 같다.

```

grunt> tt
48136,56.102.102.57,54.58.54.49,TCP,68,65491,2818,74,34,15)
32784,56.102.102.57,54.58.54.49,TCP,68,65491,2818,1,9,84)
73354,56.102.102.57,54.58.54.49,TCP,68,65491,2818,64,34,92)
72426,56.102.102.57,54.58.54.49,TCP,68,65491,2818,16,65,13)
37843,56.102.102.57,54.58.54.49,TCP,68,65491,2818,92,17,66)
66162,56.102.102.57,54.58.54.49,TCP,68,65491,2818,1,69,16)
82488,56.102.102.57,54.58.54.49,TCP,68,65491,2818,88,15,69)
82787,56.102.102.57,54.58.54.49,TCP,68,65491,2818,59,63,58)
57148,54.58.54.49,51.49.108.101,TCP,68,NA,NA,97,41,74)
53393,54.58.54.49,51.49.108.101,TCP,68,NA,NA,67,98,54)
64433,54.58.54.49,51.49.108.101,TCP,68,NA,NA,87,27,93)
24688,54.58.54.49,51.49.108.101,TCP,68,NA,NA,43,92,78)
45882,54.58.54.49,51.49.108.101,TCP,68,NA,NA,74,45,77)
167327,54.58.54.49,51.49.108.101,TCP,68,NA,NA,5,36,42)
26223,54.58.54.49,51.49.108.101,TCP,68,NA,NA,3,7,82)
62558,54.58.54.49,51.49.108.101,TCP,68,NA,NA,93,3,82)
78718,54.58.54.49,51.49.108.101,TCP,68,NA,NA,91,6,82)
21885,54.58.54.49,51.49.108.101,TCP,68,NA,NA,18,5,82)
84217,54.58.54.49,51.49.108.101,TCP,68,NA,NA,11,4,82)
77948,54.58.54.49,51.49.108.101,TCP,68,NA,NA,81,7,82)
14821,54.58.54.49,51.49.108.101,TCP,68,NA,NA,71,7,82)
43857,54.58.54.49,51.49.108.101,TCP,68,NA,NA,51,2,82)
9388,54.58.54.49,51.49.108.101,TCP,68,NA,NA,52,2,82)
73763,54.58.54.49,51.49.108.101,TCP,68,NA,NA,74,7,82)
Time taken: 19.387 seconds, Fetched: 123778 row(s)
    
```

그림 10. Pig를 사용한 데이터 추출
 Fig. 10. Data extraction using Pig

```

hive> select * from log where protocol = 'TCP';
5
57327 54.58.54.49 51.49.108.101 TCP 68 NA NA 5
6
42
26223 54.58.54.49 51.49.108.101 TCP 68 NA NA 3
82
62558 54.58.54.49 51.49.108.101 TCP 68 NA NA 93
43
78718 54.58.54.49 51.49.108.101 TCP 68 NA NA 91
45
21885 54.58.54.49 51.49.108.101 TCP 68 NA NA 18
5
84217 54.58.54.49 51.49.108.101 TCP 68 NA NA 11
5
77948 54.58.54.49 51.49.108.101 TCP 68 NA NA 81
29
14821 54.58.54.49 51.49.108.101 TCP 68 NA NA 71
2
13
43857 54.58.54.49 51.49.108.101 TCP 68 NA NA 51
17
9388 54.58.54.49 51.49.108.101 TCP 68 NA NA 52
56
73763 54.58.54.49 51.49.108.101 TCP 68 NA NA 74
23
Time taken: 19.387 seconds, Fetched: 123778 row(s)
    
```

그림 11. Hive를 사용한 데이터 추출
 Fig. 11. Data extraction using Hive

Pig와 Hive를 통해 2차 처리가 완료된 정형 데이터는 csv 포맷으로 Hadoop HDFS에 저장된다. 2차 처리까지 수행하여 분석에 필요한 데이터를 추출한 뒤 수집된 데이터에서 이상 징후를 감지하기 위해 빅데이터 분석을 수행한다.

빅데이터 분석을 위해 Hadoop HDFS에 저장된 2차 정제 데이터를 빅데이터 분석을 위해 분석용 PC의 저장소에 로드한다. 로드된 데이터는 csv 포맷의 정형 데이터 형태를 띄고 있어 분석에 용이할 뿐만 아니라 분석에 필요한 데이터만을 2차 처리하였기 때문에 보다 효율적으로 분석할 수 있다. 2차 처리 완료 데이터는 그림 12와 같다.

```

time,source_ip,destination_ip,protocol,length,source_port,destination_port,x,y,temperature,lat,lon
-21.266884,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,42,35,31,-20.42,181.62
-22.1810858,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,97,79,20,-20.52,181.03
-22.816249,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,47,39,23,-25,184.1
-22.859053,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,73,48,31,-20.42,181.96
-22.859053,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,29,20,21,-19.58,184.31
-22.178256,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,48,92,29,-11.7,186.1
-22.542709,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,2,88,27,-29,11,181.03
-22.350471,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,54,39,31,-29.74,181.74
-22.941155,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,95,45,31,-17.47,178.32
-22.1162004,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,19,70,22,-21.44,180.69
-22.961871,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,64,4,23,-12.26,187
-21.2381,50.98.98.48,97.101.98.51,TCP,1514,1060,14000,31,29,31,-18.54,182.11
-19.204833,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,26,79,23,-21,181.66
-18.1125896,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,82,11,26,-20.7,169.56
-19.1002636,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,33,45,31,-15.94,184.95
-18.679679,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,20,36,20,-19.64,166.96
-18.765895,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,95,10,27,-19.54,181.5
-18.635679,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,50,26,20,-23.5,179.78
-18.806189,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,35,97,30,-20.64,181.16
-18.765295,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,96,10,22,-20.64,181.16
-18.806189,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,35,97,30,-20.64,181.16
-19.031225,97.102.48.98,57.48.52,100,TCP,1514,1098,2010,14,24,24,-23.3,180.16
-19.94259,97.102.48.98,57.48.52,100,TCP,1514,1103,2010,32,31,-20.3,182
-19.965158,97.102.48.98,57.48.52,100,TCP,1514,1103,2010,16,24,23,-19.66,180.28
-19.108808,97.102.48.98,57.48.52,100,TCP,1514,1103,2010,76,79,23,-21,181.69
-19.1211308,97.102.48.98,57.48.52,100,TCP,1514,1103,2010,36,74,21,-14.72,167.51
-19.108808,97.102.48.98,57.48.52,100,TCP,1514,1103,2010,76,79,23,-21,181.69
-19.1457378,97.102.48.98,57.48.52,100,TCP,1514,1103,2010,6,71,31,-20.97,181.47
-17.20629,97.102.48.98,99.98,49,100,TCP,1514,1152,2010,95,89,28,-16.92,185.74
-17.41329,97.102.48.98,99.98,49,100,TCP,1514,1152,2010,57,14,31,-22.58,179.24
-17.53845,97.102.48.98,99.98,49,100,TCP,1514,1152,2010,95,89,28,-16.92,185.74
-17.65947,97.102.48.98,99.98,49,100,TCP,1514,1152,2010,39,55,27,-15.95,185.05
-17.79251,97.102.48.98,99.98,49,100,TCP,1514,1152,2010,15,17,24,-22.95,180.8
-17.95881,97.102.48.98,99.98,49,100,TCP,1514,1152,2010,77,78,23,-16.3,186
0.68,56.98.49,97.102.48.98,19.2,22,-25,82,178.98
-0.605873,56.98.49,97.102.48.98,TCP,1514,1240,2010,39,51,23,-18.73,169.23
-0.229893,56.98.49,97.102.48.98,TCP,1514,1240,2010,100,97,26,-17.84,181.29
-0.685853,56.98.49,97.102.48.98,TCP,1514,1240,2010,50,31,26,-17.66,181.4
0.181673,56.98.49,97.102.48.98,TCP,1514,1240,2010,53,25,27,-18.52,169.38
0.169563,56.98.49,97.102.48.98,TCP,1514,1240,2010,39,87,21,-37.37,176.78
    
```

그림 12. 2차 처리 완료 데이터
 Fig. 12. Second processed data

5. 빅데이터 분석

본 논문에서는 빅데이터 분석을 위해 R을 사용하였고, Node-RED를 통해 수집한 데이터의 최종형태인 Json 포맷의 파일을 R에 로드하였다. R에 로드한 Json 파일은 그림 13과 같다.

```

~/R/time_test - RStudio Source Edi...
log
Filter
time temperature humidity dB
1 100002 29 47 42
2 100102 22 46 44
3 100202 23 48 44
4 100302 24 49 43
5 100402 22 49 44
6 100502 29 48 43
7 100602 25 48 43
8 100702 23 49 43
9 100802 31 48 43
10 100902 21 48 43
11 101002 27 47 43
12 101102 26 47 43
13 101202 26 47 42
14 101302 29 46 42
15 101402 31 46 42
16 101502 27 47 44
17 101602 22 47 43
18 101702 27 47 43
19 101802 27 47 44
20 101902 25 48 44
21 102002 28 47 44
22 102102 21 48 45
23 102202 21 48 44
24 102302 25 49 45
25 102402 27 44 44
26 102502 31 47 43
27 102602 29 46 42
28 102702 30 46 42
Showing 1 to 29 of 200 entries
    
```

그림 13. R에 로드한 Json 파일
 Fig. 13. Json file loaded into R

R에 로드한 Json 파일은 Table 형태로 저장되며, 본 논문에서 분석에 사용하기 위해 추출한 시간, 온도, 습도, 소음 4가지 데이터를 1분 간격으로 수집하였다. 이렇게

수집된 데이터는 각 시간별로 수집되었다는 특징이 있기 때문에 시간을 기본 변수로 두고 분석하기 용이하다는 장점이 있다.

일정 시간동안 수집된 데이터를 분석하기 적합한 분석 모델은 시계열 분석이다. 일정 시간 수집된 데이터를 시계열 데이터라고 부르는데 시간 추이에 따라 평균, 분산이 불변할 경우 안정적인 시계열로 볼 수 있고 그렇지 않은 경우 불안정한 시계열로 볼 수 있다. 불안정한 시계열의 경우 로그, 차분 등을 통해 시계열을 안정적으로 변환하여 분석을 진행하여야 한다. 시계열 분석 시각화 그래프는 그림 14와 같다.

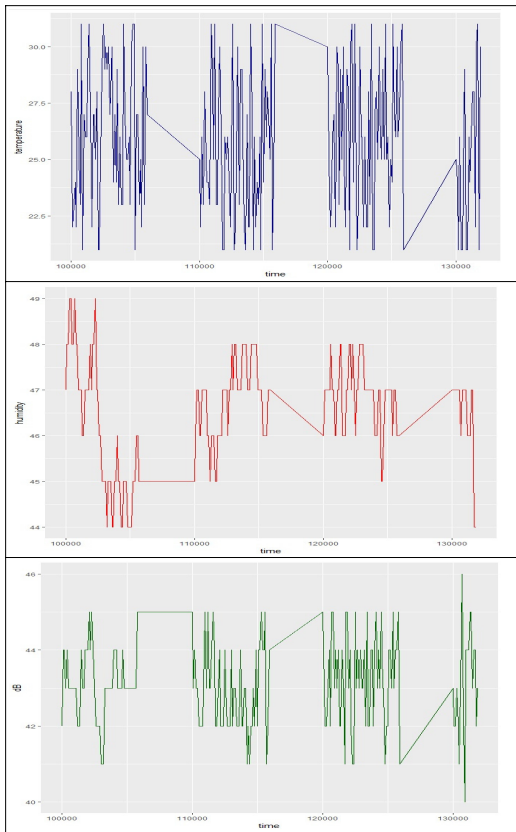


그림 14. 시계열 분석 시각화 그래프
Fig. 14. Time Series Analysis Visualization Graph

시계열분석 결과 공정 가동시간인 11시부터 13시에 1분 간격으로 수집된 데이터의 온도, 습도, 소음은 일정한 분산과 분포를 가지고 있으며, 증가하거나 감소하는 추세가 아닌 평균값을 중심으로 일정한 값이 유지되는 것을 확인할 수 있다. 이는 안정적인 시계열임을 의미하고,

공정에서 수집된 센서가 정상적으로 동작하고 있음을 확인할 수 있다. 만일 데이터가 큰 폭으로 변하게 된다면 데이터의 이상으로 정의할 수 있으며, 각 공정의 이상 진단을 수행할 수 있다.

IV. 시뮬레이션 결과 분석

본 논문에서는 빅데이터 수집, 처리를 위해 미들웨어로 Node-RED를 사용하였고 Node-RED의 성능 평가를 위해 Hadoop ECO System을 사용해 데이터를 수집, 처리하는 과정을 동일 공정에 적용하여 비교하였다. 데이터 수집 단계 저장 용량 비교는 그림 15와 같다.

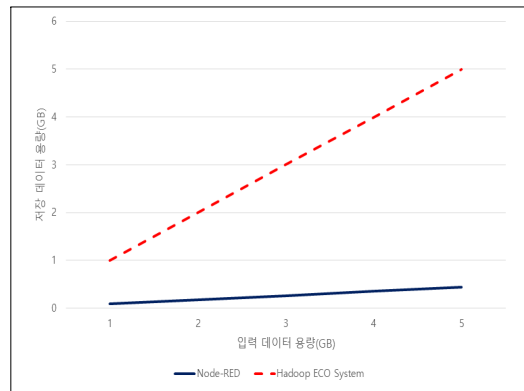


그림 15. 데이터 수집 단계 저장 용량 비교
Fig. 15. Data collection phase Storage capacity comparison

Hadoop ECO System을 사용한 데이터 수집의 경우 Wireshark를 통해 데이터 스트림을 입력받아 MySQL을 통해 데이터를 저장하기 때문에 수집단계에서는 저장 데이터 용량이 입력 데이터의 양에 비례한다. Node-RED의 경우 데이터 수집 시 사전에 입력한 필요한 데이터만을 선별적으로 수집하여 Cloud에 저장하기 때문에 수집 단계에서 저장되는 데이터 용량이 Hadoop에 비해 현저하게 적다. 또한 Hadoop ECO System의 경우 수집한 데이터를 물리적 저장장치인 HDFS에 저장하기 때문에 시스템 구성 시 물리적 저장장치의 용량에 영향을 많이 받는다. Node-RED의 경우 물리적 저장소를 거치지 않고 Cloud에 직접 저장하기 때문에 비교적 물리 저장장치의 용량에 영향을 받지 않는다.

V. 결 론

산업용 사물 인터넷(IIoT:Industrial Internet of Thing)의 발전에 따라 공장의 각 공정에서 생성되는 데이터의 중요성이 커지고 있다. 공장의 데이터는 실시간으로 수집되며 비정형성을 띄는 빅데이터이며 이를 분석하기 위해 많은 연구가 이루어지고 있다.

본 논문은 스마트팩토리 환경에서 효율적인 빅데이터 데이터 수집 및 처리과정에서 발생하는 저장소의 물리적 한계를 해결함과 동시에 데이터 처리에 소요되는 시간을 줄이고자 Node-RED를 미들웨어로 사용한 시스템을 제안하였다.

제안 시스템은 Node-RED를 사용해 입력 데이터를 비정형데이터 상태 그대로 저장하는 것이 아닌 정형데이터로 변환함과 동시에 필요한 데이터만을 선별적으로 저장한다는 장점이 있다. 제안 시스템과 Hadoop ECO System을 비교한 결과 약 10배 가깝게 전체 데이터양을 줄일 수 있었고, 수집과 동시에 데이터의 정형화, 처리를 수행하기 때문에 Hadoop의 Pig나 Hive를 사용하는 것보다 손쉽고 빠르게 분석 데이터를 추출할 수 있었다.

향후 실제 스마트팩토리 제조공정 전체를 모니터링하기 위해 더 많은 모듈과 다양한 센서를 추가하게 될 경우 본 시스템의 특성상 기존의 시스템에 관련 컴포넌트를 추가하는 것만으로 구현가능하기에 큰 어려움 없이 개발이 가능할 것이라 기대한다. 또한 수집된 센서들의 빅데이터 상관관계를 분석하고, 학습을 통해 자동화된 설비 제어까지 가능하게 된다면 스마트팩토리의 최종목표인 생산성 높은 무인화 공장을 위한 유용한 연구가 될 것이라 기대한다.

References

- [1] Hun Jung, Chong-Won Park, "Design and Implementation of MQTT Based Real-time HVAC Control Systems", Journal of the Korea Institute of Information and Communication Engineering, Volo. 19, No. 5, pp. 1163~1172, May 2015.
DOI : <http://dx.doi.org/10.6109/jkiice.2015.19.5.1163>
- [2] Node-RED, A visual tool for wiring the Internet of Things, <http://nodered.org/> accessed on 22 Aug. 2015.
- [3] Kideok-Kwon, Young-Hwan Yoo, "IoT Platform for Network Service Self-Configuration Based on Data Flow," The Journal of Korean Institute of Communications and Information Sciences, Vol.40, No.10, pp.2047-2053, 2015.
DOI : <http://dx.doi.org/10.7840/kics.2015.10.10.2047>
- [4] Kyoung-woo Cho, Min-ho Jeon, Chang-heon Oh, "Development of Equipment Control System based on DB Access Method for Industrial IoT", Journal of the Korea Institute of Information and Communication Engineering, Vol. 20, No. 6, pp.1142-1147, Jun 2016.
DOI : <http://dx.doi.org/10.6109/jkiice.2016.20.6.1142>
- [5] Sang-Yook Cha, "A Study on Big Data Circumstance and Privacy Protection", IT & Law REVIEW, Vol. 8, pp.193-259, Feb 2014.
- [6] Michael Blackstock, Rodger Lea, "Toward a Distributed Data Flow Platform for the Web of Things (Distributed Node-RED)", 2014 WoT '14 Proceedings of the 5th International Workshop on Web of Things, Vol. 8, pp.34-39, Oct 2014.
DOI : <http://dx.doi.org/10.1145/2684432.2684439>
- [7] Hyun-Je Jo, Phill-Kyu Lee, "Distributed Recommendation System Using Clustering-based Collaborative Filtering Algorithm", The Journal of IIBC, Vol. 14, No. 1, pp.101-107, Feb 2014.
DOI : <http://dx.doi.org/10.7236/JIIBC.2014.14.1.101>
- [8] Jae-Young Chang, "An Experimental Evaluation of Box office Revenue Prediction through Social BigdataAnalysis and Machine Learning", The Journal of IIBC, Vol. 17, No. 3, pp. 153-158, Jun 2017.
DOI : <http://dx.doi.org/10.7236/JIIBC.2017.17.3.167>
- [9] Jeong-Joon Kim, Kwang-Jin Kwak, Don-Hee Lee, Yong-Soo Lee, "Study of Trust Bigdata Platform", The Journal of IIBC, Vol. 16, No. 6, pp. 255-230, Dec 2016.
DOI : <http://dx.doi.org/10.7236/JIIBC.2016.16.6.225>
- [10] Julie Kim, Hyokyung Bahn, "An Efficient Log Data Management Architecture for Big Data Processing in CloudComputing Environments",

The Journal of IIBC, Vol. 13, No. 2, pp. 1-7, Apr 2013.

DOI : <http://dx.doi.org/10.7236/JIIBC.2013.13.2.1>

- [11] Han-chun Song, Hyuk-jong Ahn, "Development of Efficient Data Distribution Storage Algorithm for High Speed Data Backup in DRAM based SSD", vol. 15, No. 6, pp. 11-15, Dec 2015.

DOI : <http://dx.doi.org/10.7236/JIIBC.2015.15.6.11>

- [12] Hoyoung Hwang, Seung-Cheon Kim, Kwanghyun Ro "Edge Security System for Factory Automation Devices" The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 12 no.2, pp.251 - 258, 2012.

DOI : <http://dx.doi.org/10.7236/JIWIIT.2012.12.2.251>

- [13] Sang-Yule Choi, "Hand-Held Power Quality Monitoring System for Factory Electrical Installation" The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 15 no.3, pp.113 - 118 2015.

DOI : <http://dx.doi.org/10.7236/JIIBC.2015.15.3.113>

이 준 희(준회원)



- 2016.2 : 고려대학교 컴퓨터정보학과 (공학사)
 - 2018.2 : 한국산업기술대학교 컴퓨터공학과(공학석사)
- <관심분야 : 소프트웨어공학, 정보통신시스템, 객체지향 분석 및 설계>

김 영 곤(정회원)



- 1983.2 : 경북대학교 전자공학과(공학사)
- 1985.2 : 연세대학교 본대학원 전자공학과(공학석사)
- 2000.2 : 한국과학기술원 전산학과(공학박사)
- 1985 ~ 2007 : KT 수석연구원

- 2007 ~ 한국산업기술대학교 컴퓨터공학과 교수
- <관심분야 : 소프트웨어공학, 정보통신시스템, 객체지향 분석 및 설계>

저자 소개

윤 여 진(준회원)



- 2011.2 : 우송대학교 관광컨벤션학과 (학사)
 - 2014.2 : 한국산업기술대학교 컴퓨터융합학과(공학석사)
 - 2017.2 ~ 현재 : 한국산업기술대학교 컴퓨터공학과 박사수료
- <관심분야 : 소프트웨어공학, 정보통신시스템, 객체지향 분석 및 설계>

김 태 형(준회원)



- 2012.2 : 한국산업기술대학교 컴퓨터공학과(공학사)
 - 2017.3 ~ 현재 : 한국산업기술대학교 컴퓨터공학과 석사과정
- <관심분야 : 소프트웨어공학, IoT, 스마트팩토리>