

# 임베디드 환경에서의 빅데이터 성능 비교연구

신동빈, 윤태복  
서일대학교 소프트웨어공학과  
e-mail:popoasd123@naver.com

## A comparative study of big data performance in embedded-computing environment

Dong-Bin Shin and Taebok Yoon  
Dept. of Software Engineering, Seoil University

### 요약

최근 라즈베리파이와 같은 싱글 보드 컴퓨터의 성능이 점차 향상되고 있으며 싱글 보드 컴퓨터의 사용 분야도 드론, 임베디드 개발, 네트워크 공유 저장장치 등 여러 분야에서 사용되고 있다. 본 연구는 ARM 계열 CPU를 사용하는 싱글 보드 컴퓨터보다 범용성이 좋다고 알려진 Intel x86 계열 CPU를 탑재한 싱글 보드 컴퓨터를 이용하여 빅데이터 구현 및 처리 성능을 측정하였다. 일반적인 데스크톱 컴퓨터보다 처리 속도는 느리지만 선행연구의 싱글 보드 컴퓨터인 라즈베리파이를 다수 연결해서 분석을 진행한 것과 비교하여 가격, 공간, 전력 소모, 속도 부분에서 강점을 보여주었으며 임베디드 환경에서도 사용이 가능할 것으로 예상된다. 이를 통해 가정집이나 연구실에서 서버 컴퓨터의 활용 목적 또는 컴퓨터가 많이 보급되지 않은 지역이나 열악한 환경에서도 싱글 보드 컴퓨터를 이용하여 빅데이터 분석이 가능할 것으로 기대한다.

### 1. 서론

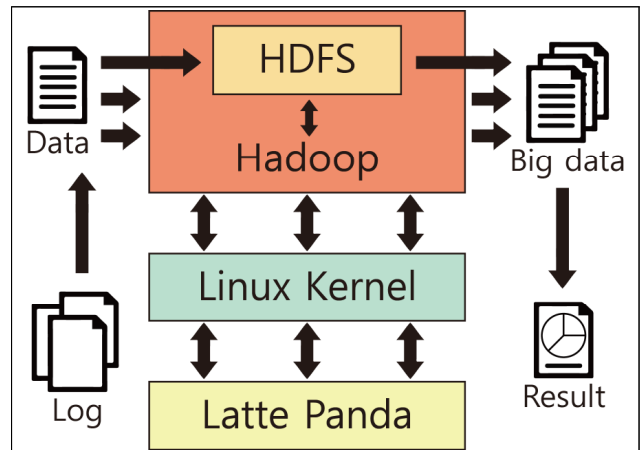
모바일 기기와 무선 네트워크의 발달로 가정에 많은 IoT 기기들이 보급되어 집밖에서도 사용자의 설정에 맞게 집안의 상황을 컨트롤할 수 있게 되었다. 또 기존에 판매되고 있는 IoT 제품뿐 아니라 아두이노, 라즈베리 파이 같은 싱글 보드 컴퓨터를 이용하여 개인이 직접 IoT 환경을 구축하여 사용하는 경우도 증가하고 있다[1].

위의 사례와 같이 개인의 싱글 보드 컴퓨터의 수요가 점점 늘어가고 있으며 기기의 성능도 더욱 상승되고 있다. 싱글 보드 컴퓨터도 특정 IT 분야에서만 사용되는 것이 아니라 교육용 컴퓨터 제작, 임베디드 개발, 네트워크 공유 저장 장치, 소형 로봇 제작 등 점차 넓어지고 있다.

본 연구는 다양한 분야에서 사용되고 있는 싱글 보드 컴퓨터를 이용하여 빅데이터 분석 및 처리 성능을 측정해보고자 하며 기존에 많이 사용되고 있는 ARM 계열 CPU를 탑재한 싱글 보드 컴퓨터가 아닌 Intel x86 계열 CPU를 탑재한 싱글 보드 컴퓨터를 사용하고 리눅스 OS와 하둡을 이용하여 싱글 보드 컴퓨터에서의 빅데이터 분석 처리를 진행해보자 한다. 그 결과 데스크톱 컴퓨터와 비교하여 사용할 때 성능 부분에서 문제가 없는지 연구를 제안하고자 한다.

### 2. 라떼판다에서 빅데이터 분석

본 연구의 진행에 있어 데이터 분석을 위한 프로그램으로는 하둡을 사용하였고 보드 컴퓨터는 상용으로 판매 중인 라떼판다를 사용하였다. 저전력 x86 CPU인 Intel Atom을 탑재하여 일반적으로 사용하는 데스크톱 프로그램들과 전부 호환이 가능하고 나중에 진행할 데스크톱과 연구와의 비교 분석에 용이하다는 점에 주목하였다.



[그림 1] 라떼판다 기반 빅데이터 처리 과정

본 연구를 진행하기 위해 라떼판다에 맞게 커스텀이 되어 있는 우분투를 설치하였으며 설치 후 연구 진행을 위해 JDK와 하둡을 설치하고 홈 경로를 잡아주었다.

연구에 사용할 데이터는 2020년 한국어 위키백과 6월 백업본을 준비하였으며 다양한 분야와 여러 단어들 존재하고 큰 용량을 가지고 있어 연구에 이용하기 적합하다고 판단하였다. 받은 데이터는 11GB 텍스트 파일로 구성되어 있으며 텍스트 파일을 연구에 사용하기 위해 100MB, 200MB, 300MB, 500MB, 1GB, 1.5GB, 2GB, 3GB, 5GB로 파일을 나누어서 작업을 진행하였다.

연구에 진행할 파일을 하둡 HDFS에 fs -put 명령어를 통해 /input폴더에 위치시켰으며 워드카운트로 파일 분석을 진행한 결과물은 /output폴더에 출력되도록 설정하였다. 데이터 처리 방식으로는 워드카운트 기능을 사용하여 YARN과 맵리듀스를 통해 분석이 진행되었으며 입력한 데이터에 중복되는 단어가 몇 개 있는지 카운트해 주는 결과물을 받았다. 이를 통해 사용자들의 관심사나 버스 이용 시간 기록, 최근 판매가 많이 되고 있는 물품 분석 등 빅데이터 분석에 있어서 기초가 되는 자료를 얻어낼 수 있었다.

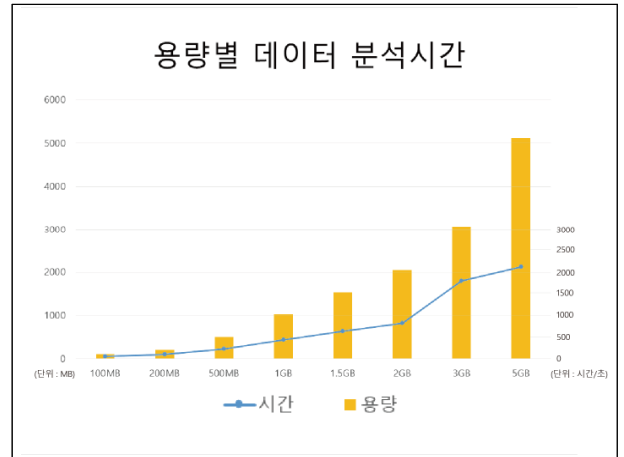
### 3. 실험

[표 1] 워드카운트 분석시간

	분석시간
100MB	53초
200MB	1분 34초
500MB	3분 47초
1GB	7분 7초
1.5GB	10분 36초
2GB	13분 43초
3GB	30분 3초
5GB	35분 26초

위의 표는 싱글 보드 컴퓨터에서 하둡 워드카운트 실행결과이며 저용량 고용량 파일을 가리지 않고 특정 용량을 분석할 때는 비슷한 시간을 소요하였다. 이는 김영근 외 2명이 쓴 논문에서 사용한 싱글 보드 컴퓨터를 5대 연결한 것보다 빠른 속도를 보여주고 있으며 또 김영선 외 4명이 사용한 싱글 보드 컴퓨터를 4대 연결한 것과 가격과 기기의 크기, 전력 소모 등과 비교하면 성능이 크게 차이 나지 않았다[2][3].

결과를 통해 적게 소비하는 전력과 공간, 낮은 가격을 장점을 삼아 가정집, 연구실에서의 서버 컴퓨터 또는 열악한 환경에서의 빅데이터 연구가 가능하다는 것을 확인하였으며 위의 두 논문과 같은 조건으로 공유기를 통해 4~5대를 연결하여 라떼판다를 네트워크로 연결하여 빅데이터 분석을 진행한다면 더욱 빠른 속도를 기대할 수 있을 것으로 예상된다.



[그림 2] 싱글 보드 컴퓨터에서의 데이터 분석소요시간

### 4. 결론 및 향후연구

현재 싱글 보드 컴퓨터는 저렴한 가격과 더욱 높아진 성능으로 판매되고 있으며 드론, IoT, 임베디드 시스템 등 IT 산업뿐 아니라 다양한 분야에서 사용되고 있다. 저렴한 가격, 높아지는 성능, 저전력 시스템, 작은 크기로 싱글 보드 컴퓨터는 다양한 분야에서의 진출 가능성은 무궁무진하다. 이번 실험으로 인해 싱글 보드 컴퓨터에서 빅데이터 분야를 적용시키도 무리가 없음을 알았고 싱글 보드 컴퓨터의 사용되는 분야를 더욱 넓혀갈 수 있었다.

과거에 출시한 싱글 보드 컴퓨터보다 저렴하지만 더욱 높아진 성능으로 여러 대를 연결하여 빅데이터 분석에 사용하면 더욱 빠르게 결과를 도출할 수 있을 것으로 예상된다.

싱글 보드 컴퓨터의 장점인 가격과 작은 크기, 저전력 시스템으로 인해 특정 장소에서의 서버 컴퓨터 설치, PC가 많이 보급되지 않은 지역/나라, 전기가 부족한 지역 등 어느 곳에서나 간편하게 사용할 수 있을 것이다. 향후 제시할 과제로는 싱글 보드 컴퓨터를 이용하여 직접 데이터 수집을 진행하고 얻은 정보를 바탕으로 빅데이터를 분석하는 것과 자료를 더욱 빠르게 여러 대를 연결하여 분석하는 것이다.

#### 참고문헌

- [1] 이수진, 유현주, 정민수, "안드로이드와 아두이노를 이용한 홈케어 시스템 설계 및 구현", 한국산학기술학회 학술대회논문집, pp. 388-390, 2014.
- [2] 김영근, 조민희, 김원중, "라즈베리파이 보드 기반의 빅데이터 분석을 위한 학습 시스템", 한국전자통신학회 논문지, 제11권 4호, pp. 433-439, 2016.
- [3] 김영선, 박지영, 윤보람, 이정현, 옹환승, "라즈베리 파이 클러스터와 아파치 스파크를 활용한 빅데이터 분석 플랫폼 연구", 추계학술발표대회 논문집, 제22권 2호, pp. 1272-1275, 2015.