

SmartX-mini Center 를 통한 NUC 클러스터의 Big Data 처리 가능성 검증

송지원, 이준기, 김승룡, 김종원
 광주과학기술원 정보통신공학부
 e-mail : {jwsong, jglee, srkim, jongwon}@nm.gist.ac.kr

Feasibility Verification of Big Data Processing employing SmartX-mini Center with NUC Cluster

Jiwon Song, Jungi Lee, Seungryong Kim, JongWon Kim
 School of Information and Communications, Gwangju Institute of Science and Technology

요 약

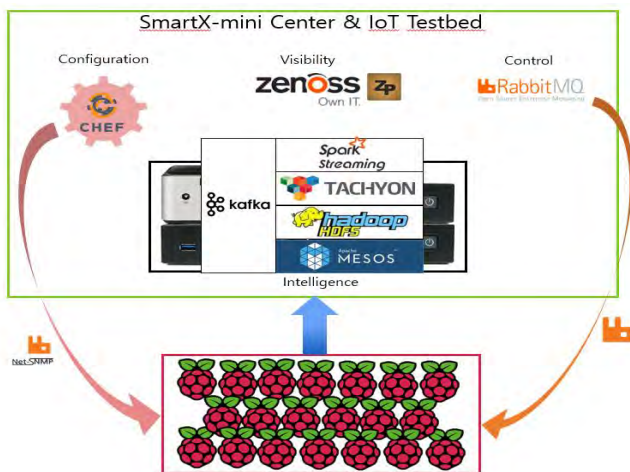
IoT 의 발달로 인해 새롭게 빅데이터와 그의 실시간 처리의 중요성이 증대되고 있다. 본 논문에서는 사물인터넷의 관제 및 데이터 처리 기능을 갖춘 SmartX-mini 센터를 통하여 NUC 클러스터의 빅데이터 처리 가능성을 제시하고, 이를 검증하기 위하여 SmartX-mini 테스트베드를 활용한다. SmartX-mini Center 의 Spark 프레임워크를 이용한 실험을 통해 IoT 환경에서의 NUC 클러스터의 빅데이터 처리 기능에 대한 가능성을 검증하였다.

1. 서론

최근 사물인터넷(Internet of Things, 이하 IoT) 관련 기술의 발달로 인해 다양한 장비들로부터 기존에 비해 수집할 수 있는 데이터의 양이 급속도로 증가하였다[1]. 이러한 빅데이터의 발달은 이를 처리하는 방법과 그것들이 작동할 수 있는 충분한 자원을 제공하는 인프라의 중요성 증대로 이어지고 있다.

mini-Box 및 μ -Box 를 이용하여 SmartX-mini 테스트베드(그림 1 참조)를 구성하였다. 그리고 이 테스트베드의 자원 인프라를 중앙에서 관제하는 소프트웨어를 SmartX-mini 센터라 정의한다.

μ -Box 의 하드웨어로는 영국 Raspberry Pi 재단에서 만든 초소형 PC 인 Raspberry Pi 2 가 사용되며, mini-Box 의 하드웨어로는 Intel 사의 소형 PC 인 NUC 이 사용된다. NUC 은 상용 서버에 비해 차지하는 공간과 전력소모가 적어 경제적이고 실제 IoT 환경에 보다 가까우면서도 IoT 실험을 위해 충분한 성능을 제공할 것으로 예상된다. 본 논문에서는 이러한 특성을 가진 NUC 을 복수로 이용해 클러스터를 구성하여 데이터 처리 기능을 시험함으로써 SmartX-mini 센터를 통한 NUC 클러스터의 빅데이터 처리 기능 가능성을 검증한다.

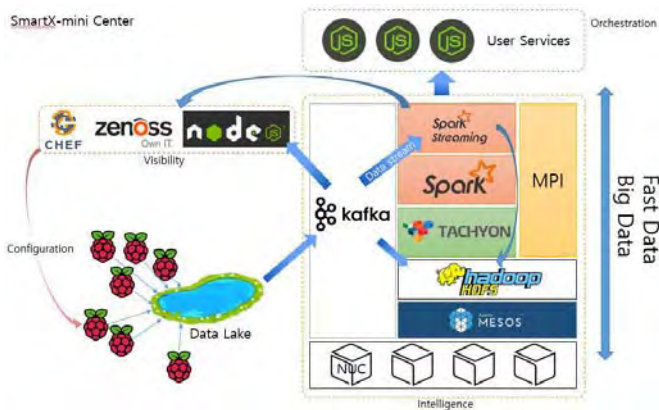


(그림 1) SmartX-mini 테스트베드

클라우드 컴퓨팅 환경에 있어 요구되는 조건은 컴퓨팅, 네트워킹, 스토리지 자원들을 동적으로 할당하고 지속적으로 필요한 서비스를 제공하는 것이다. 이 논문에서는 이러한 조건을 만족하는 융합형 자원 장치를 SmartX Box 라 칭하며[2], 이 중 IoT 환경에 대한 실험을 위해 IoT 에지(edge) 환경에 적합한 장치인

2. NUC 클러스터를 활용한 SmartX-mini Center 소프트웨어 및 하드웨어 구성

SmartX-mini 센터를 구성하는 mini-Box 들은 빅데이터 처리를 위하여 Apache Mesos[4] 클러스터로 구성되어 Apache Hadoop HDFS, Apache Tachyon 및 Apache Spark 가 Mesos 프레임워크 형태로 분산 컴퓨팅 방식으로 지원된다. 다수의 μ -Box 들과의 통신을 위해 분산 메시징 시스템인 Apache Kafka[5], 모니터링을 위해 Zenoss Server[6]가 설치되며, RabbitMQ 를 통해 μ -Box 로 명령을 전달하여 실행한다. 여기에 사용자에게 UI 를 통한 μ -Box 관제 기능을 제공하기 위해 node.js 기반의 웹 서버가 이용되며, Chef 를 이용하여 각각의 μ -Box 에 필요한 소프트웨어를 자동으로 설치한다.



(그림 2) SmartX-mini Center 의 소프트웨어

SmartX-mini Center 는 여러 대의 Intel NUC 로 구성된 클러스터에서 실행된다. Intel NUC D54250WYKH 1 대를 Mesos Master 및 Hadoop NameNode 로 설정하였고, 복수의 Intel NUC DC53427HYE 를 Mesos Slave 및 Hadoop DataNode 로 설정하였다. Mesos Slave 로 사용할 수 있는 NUC 의 수는 필요에 따라 가변적으로 조절할 수 있으며, 본 논문에서는 3 대가 사용되었다. 모든 NUC 은 한 대의 유무선 공유기에 100base-T 이더넷을 통하여 100Mbps 의 속도로 네트워크에 연결하였다.

<표 2> NUC 사양

모델명	Intel NUC D54250WYKH	Intel NUC DC53427HYE
CPU	i5-4250u	i5-3427u
RAM	16GB	16GB
저장 공간	SSD 480GB	SSD 240GB
네트워크	기가비트 이더넷	
운영체제	Ubuntu Server 14.04	

Apache Spark 는 Hadoop MapReduce 를 대체할 목적으로 개발된 분산 컴퓨팅 프레임워크로, 기계학습, 양방향 쿼리, 스트리밍 처리에 강점이 있는 빅데이터에 특화된 기술이다. 메모리에서 작업 시 Hadoop 에 비해 최대 100 배의 성능 향상이 있는 등 MapReduce 에 비해 성능상의 이점을 가지고 있다[7]. SmartX-mini Center 에서 Spark 는 Mesos 프레임워크로 실행되며, HDFS 를 스토리지로 사용한다. Spark API 가 제공되는 프로그래밍 언어로는 현재 Scala, Python 및 Java 가 있다.

3. NUC 클러스터를 활용한 Big Data 처리 가능성 검증

Apache Spark 를 이용하여 NUC 클러스터의 Big/Fast Data 처리 가능성 검증을 위해 Scala 코드를 Spark 를 이용하여 실행하는 실험을 진행하였다. 실험은 NUC 클러스터 상에서 실행되는 Spark Shell 에서 일련의 Scala 스크립트 명령을 실행하여 최초 작업 할당부터 데이터 처리 및 결과 출력까지의 총 소요 시간을 측정하는 방식으로 진행되었다. Spark Shell 에서 실행된 작업은 Mesos 프레임워크의 task 형태로 실행된다. 실험

내용은 10000 개의 난수를 생성한 후 그 중 특정 범위에 있는 수를 필터링하는 Scala 코드를 실행하는 것으로, 난수들은 다수의 μ -Box (Raspberry Pi 2)로부터 수집된 데이터베이스라고 가정할 수 있다.

실험 결과, 코드 실행이 완료될 때까지 걸린 시간은 10.35 초로, 현재 SmartX-mini 테스트베드에서 사용되는 20 대의 μ -Box 들로부터 수집되는 데이터를 처리하기에 충분한 수준인 것을 확인할 수 있었다.

<표 2> 실험 결과

과정	시간
총 소요 시간	10.35 초

4. 결론 및 향후 계획

본 논문에서는 SmartX-mini Center 의 Big/Fast Data 처리 모델을 제시하고, Spark 를 이용한 데이터 처리 실험을 통하여 이를 검증해보았다. 향후 SmartX-mini 센터에 실제로 Raspberry Pi 2 를 이용한 다수의 μ -Box 들로부터 Kafka 를 통해 수집된 데이터들을 와 Spark Streaming 을 통하여 실시간 처리하는 환경을 구현하여 보다 실증적인 IoT 환경에 대한 실험을 진행할 계획이다.

감사의 글

본 논문은 2014 년 미래창조과학부의 재원으로 SW 융합기술고도화 사업의 지원을 받아 수행된 연구임 (S1004-14-1045).

참고문헌

- [1] K. S. Yeo, M. C. Chian, T. C. W. Ng and A. T. Do, "Internet of Things: Trends, challenges and applications," in *Proc. IEEE ISIC*, 2014.
- [2] 김남곤, 김종원, "효율적인 자원 활용을 위한 SmartX 노드 가상화 적용", 2013 년 한국컴퓨터종합학술대회 논문집, pp. 1296, 2013 년
- [3] 미니 PC: 인텔® NUC, <http://www.intel.co.kr/content/www/kr/ko/nuc/overview.html>
- [4] B. Hindman, A. Konwinski, M. Zaharia, A. Ghodsi, A. D. Joseph, R. H. Katz, ... and I. Stoica, "Mesos: A Platform for Fine-Grained Resource Sharing in the Data Center," *NSDI*, Vol. 11, 2011.
- [5] J. Kreps, N. Narkhede, and J. Rao, "Kafka: A distributed messaging system for log processing," *Proceedings of 6th International Workshop on Networking Meets Databases (NetDB)*, Athens, Greece, 2011.
- [6] Zenoss, <http://www.zenoss.com/>
- [7] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, & I. Stoica, "Spark: cluster computing with working sets." *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, 2010.