

# Hadoop을 이용한 대용량 데이터 변환

이강은\*, 정민진\*, 정다빈\*, 김성석\*\*, 양순옥\*\*\*  
 \*서경대학교 컴퓨터과학과  
 \*\*서경대학교 컴퓨터과학과, 교신저자  
 \*\*\*가천대학교

e-mail:{gang3039, alsuls4390, zzzx1227}@naver.com, sskim03@skuniv.ac.kr,  
 soyang@gachon.ac.kr

## High volumes of data conversion based on Hadoop

Kang Eun Lee\*, Min Jin Jeong\*, Dabin Jeong\*, Sung Suk Kim\*\*

\*Dept of Computer Science, Seokyeong University

\*\*Dept of Computer Science, Seokyeong University, corresponding author

\*\*\*Gachon University

### 요 약

Hadoop은 대용량 데이터의 분산 처리 응용을 지원하는 프레임워크이다. 이는 마스터 노드와 데이터 노드간에 Map-Reduce 과정을 거쳐 분산 처리를 지원한다. 이에 본 연구에서는 3D 프린팅을 위해 생성한 3D 모델을 프린터가 인식할 수 있는 G-code로 변환하는 작업을 Hadoop에서 수행하였다. 3D 모델은 대개 2차원 개체(페이셋)를 이용하여 표면을 표현하는데, 이 개체를 높이(Z 축)에 따라 슬라이싱한 후 각 레이어별로 G-code를 생성하여야 한다. 우선 5대의 컴퓨터에 Hadoop 클러스터를 설치한 후, 대상 3D 모델에 다양한 속성값을 변경하면서 변환작업을 진행하여 Hadoop 프로그래밍의 장점을 확인할 수 있었다.

### 1. 서론

최근 데이터의 중요성이 커지면서, 데이터의 획득과 이를 활용하기 위한 연구가 활발하게 진행되고 있다. [1]에서는 2020년 이후 발생하는 데이터의 양이 폭발적으로 증가할 것으로 예상하고 있다. 데이터의 양이 커진다는 것은 대용량 데이터의 효율적 처리 알고리즘 개발뿐만 아니라 실제 처리할 수 있는 기반 기술도 중요해진다. 본 논문에서는 활용 분야가 넓어지고 있는 3D 프린팅 분야에서 대용량 데이터 변환 작업을 이용하여 대용량 데이터 처리를 수행한 결과를 담고 있다.

3D 프린팅 혹은 적층형(additive) 제조 기법은 디지털 파일에 따라 재료를 쌓아서 3차원 제품을 만드는 기술이다[2]. 대상 제품은 대개 제일 하단에서부터 각 층마다 재료를 뿌려 제조되며, 그 하나의 층은 매우 얇은 두께를 가진다. 이러한 3D 프린팅은 전통적인 제조 방식에 비해 복잡한 형상을 쉽게 제조할 수 있게 한다.

프린팅을 하기 위해서는 우선 3D 모델을 생성해야 한다. 모델은 CAD와 같은 일반적인 전용 소프트웨어나 3D 스캐너와 같은 장비를 이용하여 제작할 수 있다. 일단 제작하고자 하는 3D 모델이 생성되면 이를 슬라이싱하여 수백~수천개의 레이어로 나누어야 한다. 각 레이어에는 그 레이어에서 재료를 뿌림으로써 생성되어야 할 제품의 윤곽 좌표가 있으며, 이는 3D 프린터가 인식할 수 있는 명령어(G-code)로 변환되게 된다.

아파치 하둡(Apach Hadoop)은 대용량 데이터의 분산

처리 응용을 지원하는 프리웨어 자바 소프트웨어 프레임워크이다[3, 4]. 분산처리 시스템인 구글 파일 시스템을 대체할 수 있는 하둡 분산 파일 시스템(HDFS: Hadoop Distributed File System)과 맵리듀스를 구현한 것이다. 프리웨어여서 여러 대용량 데이터 처리를 위한 응용이나 알고리즘 개발에 널리 활용되고 있다[5].

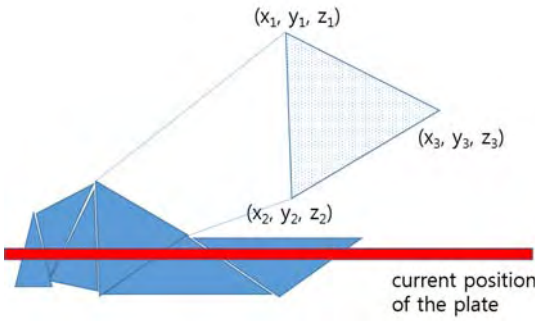
만약 3D 모델의 크기가 커지거나 대상의 정밀도가 높아진다면 제품의 표면을 구성하는 단위가 매우 작아지게 되고, 이를 G-code로 변환하는 작업은 매우 많은 시간을 소모하게 된다. 이를 위해 본 논문에서는 이를 Hadoop을 이용하여 이 변환작업을 수행하고자 한다.

### 2. Hadoop에서 3D 모델 데이터를 G-code로 변환

#### 2.1 G-code 생성 방법

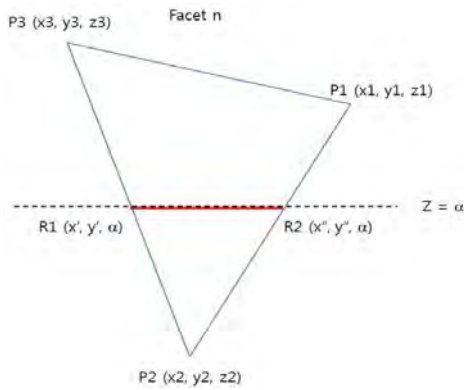
일반적으로 3D 모델은 오른쪽 그림과 같이 작은 크기의 2차원 도형을 연결하여 표현할 수 있다. 그림에서는 삼각형으로 표면을 나타냈으며(이러한 삼각형을 대개 페이셋(Facet)이라고 한다)[6], 따라서 3D 모델은 이 삼각형의 3차원 공간상의 좌표들을 담고 있어야 한다. 이 모델에 대하여 Z축으로 일정 크기만큼 이동하면서 슬라이싱한 후 그 레이어의 표면 데이터로부터 프린팅 정보를 추출하여 G-code를 생성할 수 있다((그림 1)참고).





(그림 1) 슬라이싱한 후 레이어와 겹치는 부분

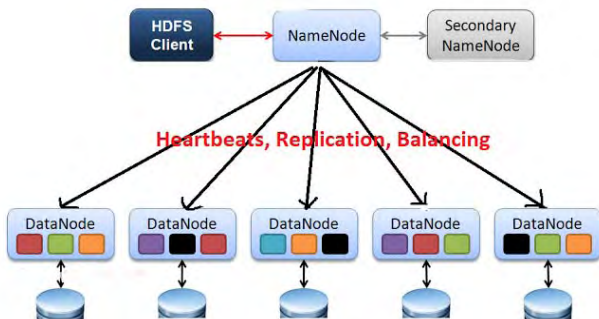
각 페이스트와 현재의 레이어와 겹치는 부분을 계산하기 위해서는 3D 모델 내에 있는 다수의 페이스트 중에서 현재 높이에 해당하는 레이어( $Z$ 축 =  $\alpha$ )와 겹치는 페이스트를 찾은 후, 각각의 페이스트에 대하여 (그림 2)에서 보여지듯이 겹치는 좌표를 계산해야 한다. 이 계산방법은 일반적인 교점 계산 수식을 활용한다.



(그림 2) 각 페이스트와 현재의 레이어(붉은 색)와 겹치는 부분 계산

만약 3D 모델의 크기가 커지거나 정밀도가 높아진다면 표면을 구성하는 페이스트의 개수는 제품의 비율로 증가하게 되므로 이러한 상당한 변환시간을 소요하게 된다.

## 2.2 Hadoop을 이용한 변환 작업



(그림 3) 하둡을 이용한 변환작업

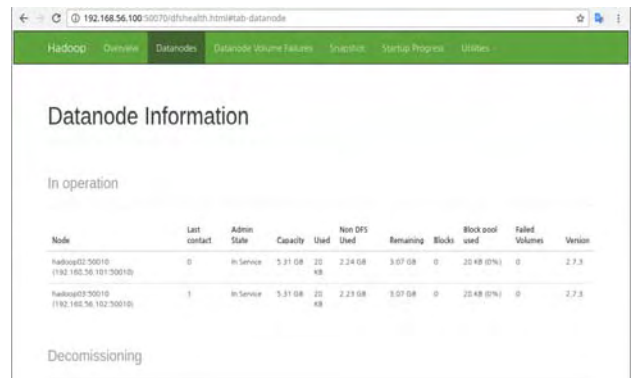
작은 하둡 클러스터에는 하나의 마스터와 여러 워커 노드들로 구성 되어 있다. 마스터 노드들은 JobTracker,

TaskTracker, NameNode, DataNode로 구성 된다. 슬레이브 또는 워커 노드는 DataNode와 TaskTracker로서 동작을 한다((그림 3) 참고).

**Hadoop**을 이용한 변환을 위해서 입력 파일에 대한 전처리 - Map - Shuffling - Reduce 과정을 진행하였다. 1) 전처리는 3D 모델에 포함된 페이스트들에 대하여 필요한 좌표값 정보만 추출한 입력 파일을 만든다. 2) 만들어진 입력 파일에 대하여 Map 단계에서는, 각 데이터 노드에게  $N$ 등분된 전처리 파일을 전달하여 직접 G-code를 생성하도록 한다. 읽어들인 각 ( $Key, Value$ )에서  $Key$ 는 레이어의 높이이며,  $Value$ 는 이전 전처리 단계에서 저장한 좌표값이다. 3) Shuffling 단계에서는 Map 단계에서 생성한 ( $Key, Value$ ) 데이터 중  $Key$  데이터 값을 이용하여 정렬을 수행한다. 이는 각 레이어( $Z = a$ )에 대하여 여러 변환되어야 할 좌표 정보(그림 2)들에 해당하며, ( $key, [value_1, value_2, \dots, value_n]$ ) 형태로 표현된다. 즉 하나의 높이( $key$ )에 생성된 여러 좌표는 추후 프린터가 재료를 뿌려야 할 위치에 해당된다. 4) Reduce 단계에서는 각 레이어별로 생성된 좌표에 대하여 프린팅 과정에서 헤드의 이동거리를 최소화 하도록 좌표에 해당하는  $value$  쌍들의 위치를 조정하는 작업이다.

## 3. 변환 결과

Hadoop은 5대의 컴퓨터 위에 설치되었으며(*hadoop1 ~ hadoop5*), 그 중 *hadoop1*이 마스터 노드이고 나머지는 데이터 노드로 동작한다((그림 4) 참고).



(그림 4) 웹 브라우저를 통해 확인한 Hadoop 클러스터

변환 작업을 위해 하나의 3D 파일에 대해 정밀도와 크기를 변환함으로써 변환시간을 살펴보았다. 즉, 기본 3D 파일(정밀도 = 3, 객체 높이 = 9cm)에 대해 높이를 변화시키거나 정밀도를 변화시켰으며, 이는 모델 내에 포함된 페이스트의 개수를 변화시키게 된다.

(그림 5)의 변환시간에서 X축은 데이터 노드의 개수이고, Y축은 변환 시간을 의미한다. 당연히 데이터 노드의 개수가 증가하면 분산도가 커져 변환시간이 향상되고 있음을 알 수 있다. 변환에서 단순히 높이만 변화시켰지만, 현실적으로는 높이가 증가하면 부피도 커져 페이스트의 개수는 더 많아질 것이며, 이 경우에는 변환시간의 단축 효

과가 더 커질 것으로 판단된다.

**참고문헌**

[1] The Exponential Growth of Data, (2017)  
<https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data>

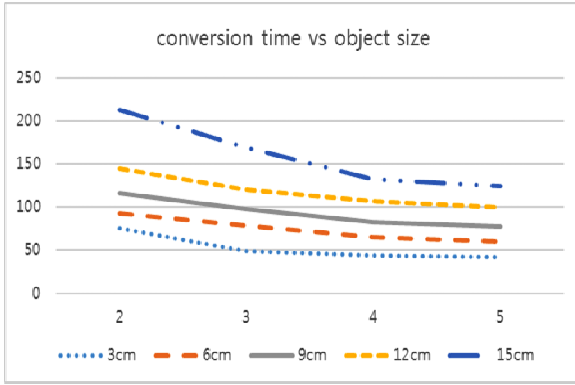
[2] J. Teixeira, G. Barros, V. Teichrieb, W. Correia, “3D Printing as a Means for Augmenting Existing Surfaces”, Symposium on Virtual and Augmented Reality (SVR), pp. 24-28, 2016

[3] [https://ko.wikipedia.org/wiki/아파치\\_하둡](https://ko.wikipedia.org/wiki/아파치_하둡)

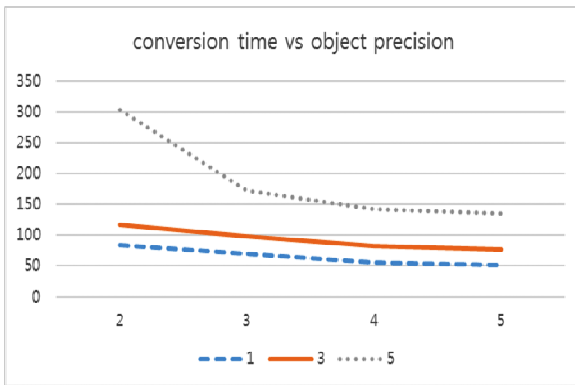
[4] G. Turkington, “Hadoop - Beginner’s Guide“, PACKT, 2013

[5] S. Vemula, and C. Crick, “Hadoop image processing framework”, IEEE Int’l Congress on Big Data, pp. 506-513, 2015

[6] Faul F. Jacobs, “Rapid prototyping & Manufacturing: Fundamentals of Stereolithography“, Society of Manufacturing Engineers, 1992



(a) 정밀도가 3일 때 높이의 변화



(b) 높이가 9cm일 때 정밀도의 변화  
 (그림 5) 변환시간

**4. 결론**

본 논문에서는 3D 프린팅 과정에서 대용량의 페이셋 데이터를 G-code로 변환하는 작업을 Hadoop 플랫폼을 이용하여 진행하였다. Hadoop 플랫폼의 Map-Reduce 기능을 활용하면 분산 작업 과정의 복잡도를 줄이면서도 효율적인 분산처리가 가능함을 알 수 있었다. 추후 더 복잡한 3D 모델과 더 많은 데이터 노드를 이용하여 대용량의 데이터 변환에도 적용하고자 하며, 이 작업이 원활하게 이루어질 경우 다양한 응용(빅데이터 처리나 기계학습을 위한 전처리 등)에도 적용할 수 있는 기반 기술로 활용하고자 한다.