

# 빅데이터 분석 서비스 운영 관리를 위한 빅데이터 서비스 브로커 설계 및 개발

김바울\*, 김상규\*, 김수빈\*, 구원본\*

\*이노그리드 클라우드컴퓨팅연구센터

baul@innogrid.com, souvenir@innogrid.com, cafe43@innogrid.com, wbkoo@innogrid.com

## Development of bigdata service brokers for bigdata analysis service operation and management

Baul Kim\*, Subin Kim\*, Sanggyu Kim\*, Wonbon Koo\*

\*Cloud Computing R&D Center, Innogrid

### 요 약

본 논문에서는 기존의 산업 및 서비스 변화에 따라 발생하는 빅데이터 분석 서비스 처리를 위한 빅데이터 분석 서비스 브로커 시스템을 제안한다. 기존의 빅데이터 분석 시스템은 분석하는 시간 동안 지속적으로 자원을 점유하고 있어야 하며, 이러한 서비스를 이용하기 위해 내부에 대규모의 시스템을 구축하고 지속적으로 운영해야하는 단점이 존재한다. 본 논문에서는 빅데이터 분석에 필요한 자원을 효과적으로 사용하기 위해 클라우드 기반의 자원 관리와 연계하고 서비스 이용을 용이하게 하기 위해 단일 엔드포인트 기반의 빅데이터 분석 서비스 호출 구조를 설계하였다. 이를 통해 빅데이터 서비스 분석에 소요되는 자원 점유에 따라 동적으로 자원을 생성 관리하여 자원을 보다 효과적으로 이용할 수 있는지 테스트베드를 구축하여 서비스 이용 및 자원 사용을 효과적으로 하는지 확인하였다. 또한, 이를 통해 대규모 자원을 지속적으로 점유해야하는 빅데이터 분석 플랫폼의 자원사용에 대한 한계를 일부 해소하여 자원을 효과적으로 이용할 수 있는 것을 확인하였다.

### 1. 서론

2020년 초부터 이어지고 있는 팬데믹(pandemic) 상황이 장기화되면서, 디지털 기술의 수요 증가와 맞물려 디지털 전환(Digital Transformation)이 빠르게 진행되고 있다. 디지털 전환의 중요 기술로는 모빌리티, 클라우드 컴퓨팅, 사물인터넷(IoT), 인공지능(AI)가 강조되며 각 기술에서 나오는 다양하고 방대한 데이터를 분석하고 처리해야 하는 빅데이터 분석에 대한 수요가 증가하였다.

빅데이터 분석은 다양한 소스에서 다양한 크기의 정형, 반정형 및 비정형 데이터를 포함하는 방대한 데이터에 대해 분석하는 프로세스이다. 빅데이터의 주요 요소로는 기본적으로 크게 정형 데이터와 비정형 데이터로 구분할 수 있는 데이터가 필요하고, 데이터 수집, 처리, 저장 기술인 NoSQL이나 Hadoop 등의 프레임워크가 필요하다. 다음으로 회귀분석, 군집화 등 데이터를 분석하여 정보를 도출하는 데이터 분석 및 지식 추출 기술이 필요하다. 하지만 모든 기업이 빅데이터 분석을 위해서 위에서 제시한 요소들을 구성하고 활용하려면 내부에서 보유하고 있는

자원적 한계가 있거나 기술적 개발 및 환경 구성을 위한 비용이 소요된다.

본 논문에서는 빅데이터 서비스 운영 및 관리의 어려움을 해결하기 위해 다중 호출 사용자 환경에서의 빅데이터 서비스 브로커 운영 관리 플랫폼을 제안한다. 다중 호출 사용에 대한 서비스 점유 상태 관리 및 빅데이터 분석 서비스에 대한 편의성 향상을 보완하기 위해 인터페이스 기반에 빅데이터 서비스 환경 구성 및 제공하는 시스템을 쿠버네티스를 기반 컨테이너로 구성하여 신속하고 자원 활용성이 향상된 시스템을 제안 한다. 또한 서비스 이용의 편의성을 위해 단일 엔드포인트로 여러 빅데이터 서비스와 연계하기 위한 연계 방식을 제안한다.

본 논문의 2장에서는 빅데이터 환경과 다양한 사용자 수용을 위한 멀티테넌트 기반 서비스와 관련된 연구를 살펴보고, 3장에서는 다중 호출 사용자 환경에서의 빅데이터 서비스 브로커 운영 관리 플랫폼을 제안하며, 4장에서는 구현 결과를 평가하고, 5장에서 결론을 내린다.

## 2. 관련 연구

### 2.1 빅데이터 분석 플랫폼 관련 기술 동향

빅데이터 분석 플랫폼 관련 기술은 데이터의 전 처리부터 분석과 운영을 위해 Data Science의 전 영역을 지원하기 위한 것으로, 다양한 형태의 데이터를 수집, 분석, 활용하기까지 주요 기술은 크게 이기종 데이터 융합 기술, 빅데이터 저장 관리 기술, 빅데이터 처리 기술, 빅데이터 분석 기술 및 지능가시화 기술과 관련된 연구가 이루어지고 있다.[1, 2]

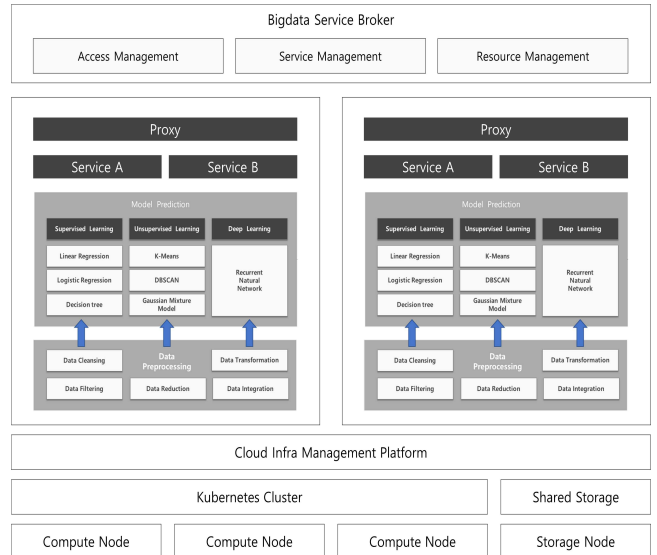
또한 빅데이터 분석 및 시각화 기술은 대부분 통계학, 데이터 마이닝, OLAP(OnLine Analytical Processing) 등 대용량의 데이터 속에서 상관관계, 패턴 등을 발견하며 예측 기술도 네트워크 이론과 신경망 이론을 이용하여 다양한 분석 및 시각화 기술들이 연구되고 있다.[3]

### 2.2 클라우드 멀티테넌트 관련 기술 동향

클라우드 멀티테넌트 관련 기술은 클라우드 자원을 사용자마다 제공하기 위해 자원 격리, 보안 등을 지원하는 기술로 다중 호출 사용자에게 따라 대규모 멀티 테넌트 기술을 적용하기 위해 서로 다른 멀티 테넌트 모델이 서로 다른 환경에 적용할 수 있도록 멀티테넌트 및 트레이드오프 등 다양한 형태의 공유를 통해 추상화 수준에서 리소스를 공유하고 격리 보장하도록 다중 멀티 테넌트에 대한 다양한 접근 방식을 연구하고 있다[4].

## 3. 시스템 구조

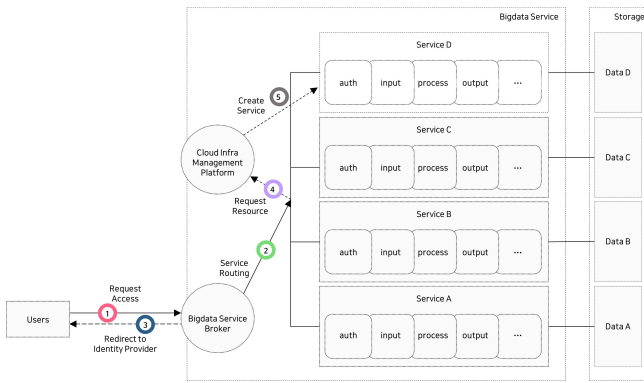
본 논문에서 제안하는 시스템은 그림 1과 같이 빅데이터 서비스를 제공하기 위해 클라우드 인프라를 기반의 자원과 이를 통해 운영되는 빅데이터 서비스와 이를 중개해주기 위한 빅데이터 서비스 구조로 구분되어 있다. 클라우드 인프라는 자원을 효율적으로 제공하고 신속한 서비스 운영 관리를 위한 컨테이너 기반의 클러스터를 구축하여 제공한다. 또한 빅데이터 분석에 필요한 자원을 저장하고 관리할 수 있도록 스토리지를 구성하여 데이터 업로드 및 분석 서비스 브로커가 접근하여 이용할 수 있도록 구성하였다. 빅데이터 서비스는 빅데이터 분석 환경을 구성하여 이를 기반으로 분석 서비스를 인터페이스 형식으로 제공하는 환경으로 구성되어 있다. 서비스 브로커는 구성된 빅데이터 서비스 및 신규 서비스 생성 관리 등의 기능을 통해 사용자와 빅데이터 분석 서비스와의 연결을 관리한다.



(그림 1) 빅데이터 서비스 브로커 구조

본 논문의 빅데이터 서비스 브로커 기능은 일반적인 API 게이트웨이와 같이 인터페이스를 기반으로 서비스를 중개하는 목적을 가지고 있지만 빅데이터 분석 서비스는 대용량의 데이터를 처리하는데 시간이 소요되며 이를 처리하기 위한 자원과 처리하는 동안에 세션유지 및 분석 플랫폼에 대한 자원이 중단되지 않고 분석 완료시까지 지속적으로 관리되어야 한다. 또한 사용자가 서비스를 편리하게 이용하기 위해 단일 엔드포인트 기반으로 서비스를 이용할 수 있는 브로커 주소와 서비스 명칭을 제공하며 백엔드에서는 클라우드 자원을 기반으로 여러 빅데이터 서비스들을 동적으로 생성하여 서비스 처리에 필요한 자원을 서비스 이용량에 따라 유동적으로 관리하는 기능을 제공해야 한다.

빅데이터 서비스 흐름은 그림 2와 같이 구성된다. 사용자가 필요한 빅데이터 서비스를 호출할 경우 빅데이터 서비스 브로커는 사용자가 요청한 서비스의 이름과 분석에 필요한 자원 정보를 제공하는지 확인하여 빅데이터 분석 서비스에게 요청 정보를 전달해준다. 요청받은 빅데이터 서비스는 분석에 필요한 데이터를 사용자가 업로드한 저장소에서 가져와 분석을 진행하고 분석 결과를 빅데이터 서비스 브로커를 통해 사용자에게 반환한다. 만일 이용할 수 있는 빅데이터 서비스가 없는 경우 혹은 다른 사용자가 점유하고 있는 경우 클라우드 관리 플랫폼에 요청하여 빅데이터 서비스 자원을 생성하고 생성된 자원을 활용하여 서비스 분석을 수행하여 결과를 반환한다. 서비스 분석이 완료된 자원은 클라우드 관리 플랫폼을 통해 자원을 반환한다.



(그림 2) 빅데이터 서비스 흐름도

4. 구현 결과

본 논문에서 구현 및 검증에 활용된 시스템의 테스트 환경은 표 1와 같다. 빅데이터 분석을 위해 고성능 테스트 환경을 구축하고 쿠버네티스를 구성하여 빅데이터 서비스 브로커의 동작을 확인하였다.

<표 1> 실험 환경

테스트 서버				OS/SW
컴퓨터 서버		스토리지 서버		
CPU	Memory	Disk	Disk	
AMD 3990X 64-Core	256GB	1TB SSD	8TB HDD	Ubuntu 20.04 Kubernetes 1.21

실험 데이터는 종관기상관측(ASOS) 데이터를 활용하여 RNN 모델을 이용한 분석 결과 정보를 제공하는 기능을 빅데이터 서비스를 만들고 사용자의 호출이 있을 때 해당 서비스 존재 여부에 따라 기존자원을 이용하거나 신규 자원을 생성하여 제공하는 것을 테스트하였다. 표 4는 사용자가 동일한 데이터를 가지고 2가지 분석 서비스를 각각 호출하고 출력된 결과이다. 빅데이터 분석 서비스 호출 주소는 동일하며 분석 모델에 따라 2가지 버전으로 분석 서비스 호출을 진행하였다. 동일한 데이터임에도 각각의 분석 기능에 따라 다른 결과가 호출된 것을 확인할 수 있다.

<표 2> 호출 결과

주소	192.168.100.150:9090/service/1/test	192.168.100.150:9090/service/2/test
분석 데이터	종관기상관측(ASOS) 데이터	종관기상관측(ASOS) 데이터
결과		

그림 3은 클라우드 관리 플랫폼을 통해 자원 생성이 이루어진 것을 확인하기 위한 결과 화면이다. 첫 번째 분석 서비스를 호출하고 나서 분석 단계일 때 두 번째 분석 서비스를 호출하고 해당되는 분석 서비스를 처리하기 위해 신규 빅데이터 분석 환경이 생성된 것을 확인할 수 있다. 이후 분석이 완료되고 결과를 반환하면 해당 자원은 중지상태로 변환되며 추가적인 이용이 없을 경우 자원을 반환하게 된다.

```

root@kubetest:~# kubectl get pods -n singleproject
NAME          READY   STATUS    RESTARTS   AGE
master-v1-8b7664dd8-plpkf  1/1     Running   0           111m
worker1-v1-bd9dbc79-lwm9n  1/1     Running   0           111m
worker2-v1-9dd787668-wmm6w  1/1     Running   0           111m
root@kubetest:~# kubectl get pods -n singleproject2
NAME          READY   STATUS    RESTARTS   AGE
master-v1-7664f897fc-bkdx9  1/1     Running   0           102m
nginx-vjrfsh-7d846dc598-nrsbk  1/1     Running   0           101m
worker1-v1-d5495fdb7-rf7kf  1/1     Running   0           102m
worker2-v1-594b4fd94b-591th  1/1     Running   0           102m
    
```

(그림 3) 빅데이터 분석 서비스 호출 시 자원 생성 결과

5. 결론

본 논문에서는 빅데이터 분석 서비스 이용의 편의성 향상과 자원 사용의 효율성 향상을 위한 빅데이터 분석 서비스 브로커를 제안하였다. 빅데이터 분석 서비스가 가진 특성에 따라 자원을 효과적으로 생성하고 관리하기 위해 클라우드 플랫폼과 연계하여 시스템을 설계하였다. 서비스 이용자로서는 단일 엔드포인트와 서비스 명칭을 가지고 서비스를 이용함으로써 서비스 이용의 편의성이 향상되었으며 시스템 관리자도 자원을 효과적으로 사용하여 자원 활용성에 유용함을 확인할 수 있었다.

Acknowledgment

본 연구는 산업통상자원부/한국산업기술진흥원의 국가혁신융복합단지지원사업의 지원으로 수행되었음 (과제번호 P0015306).

참고문헌

[1] 이미영, et al. "빅데이터 분석을 위한 빅데이터 처리 기술 동향." 정보처리학회지(2012): 22-23  
 [2] 손기준, et al. "하둡 기반 빅데이터 수집 및 처리를 위한 플랫폼 설계 및 구현" Sympathy Contents Session(2015): 1-2  
 [3] 김상락, et al. "빅데이터 분석 기술의 오늘과 미래" 정보과학회지(2014): 9-10  
 [4] Divyakant, A., et al. "Big Data and Cloud Computing: Current State and Future Opportunities." ICPS(2011): 1-4