

컨테이너 기반 분산 컴퓨팅 환경 구축을 위한 컨트롤러 개발

김기현, 문정훈, 석우진

한국과학기술정보연구원

{kkh1258, jhmoon, wjseok}@kisti.re.kr

Controller Development for Distributed Computing Environment Contracture based on Container

Ki-Hyeon Kim, Junghoon Moon, Woojin Seok

Korea Institute of Science and Technology Information

요약

최근 과학기술의 다양한 분야에서 빅데이터와 인공지능을 이용한 연구들이 활발하게 이루어지고 있다. 농업, 생명, 화학, 환경, 네트워크, 자동차 등 다양한 분야에서 인공지능 연구를 진행하고 있지만 연구 협업을 위한 빅데이터 고속 전송 및 인공지능 연구 환경을 제공하기 위한 시스템이 부재하다. 이에 따라 국가과학기술연구망에서는 빅데이터 고속전송을 위한 분산 ScienceDMZ 환경을 구성하고 이 환경에서 중앙집중적으로 데이터를 처리할 수 있도록 분산 컴퓨팅 시스템을 구성하여 이를 사용자들이 손쉽게 사용할 수 있도록 컨트롤러를 개발하였으며, 이 시스템에서 병렬 컴퓨팅과 분산컴퓨팅의 성능 차이를 비교하고 이를 통해 시스템의 성능 검증하고자 한다.

I. 서론

과학기술 분야에서는 거대한 장비에서 생성되는 데이터들과 실제 국가에서 관리하고 있는 데이터의 크기는 GB단위를 넘어 TB단위로 넘어가고 있고, 이를 이용한 다양한 계산 과학 및 인공지능 연구들이 이루어지고 있다. 사실 과학기술 분야에서는 이와 같이 거대한 데이터를 다른 연구기관들과 협업을 위해 데이터를 주고받고 있지만, 많은 기관들이 빅데이터를 위한 네트워크 시스템과 플랫폼을 활용하지 못하고 있다. 빅데이터를 전송하기 위한 네트워크 구조인 ScienceDMZ, 데이터 전송을 위한 서버인 DTN, 안전한 데이터 전송 플랫폼인 Globus Online을 이용하여 데이터를 안전하고 고속으로 전송할 수 있는 환경을 구성할 수 있다.

빅데이터 고속 전송 환경은 ScienceDMZ[1], DTN, Globus Online을 이용하여 데이터 전송의 문제를 해결했다면, 연구자들 간에 협업을 위한 인공지능 연구 플랫폼은 존재하지 않는다. 실제로 과학기술 연구 분야에서 인공지능 연구자들은 개별적으로 소규모의 컴퓨팅 환경을 구성하여 연구를 진행하는 것이 대부분이다. 이와 같은 문제를 해결하기 위해 인공지능 연구자들 간의 협업을 위한 인공지능 연구 플랫폼의 필요성이 대두되고 있다. 본 논문에서는 빅데이터 고속 전송을 위한 DTN 시스템에 GPU 컴퓨팅 환경을 구성하여 기존의 데이터 전송만을 위해 사용하던 방식을 전송을 활용하지 못하는 유휴시간에 데이터를 처리하기 위한 컴퓨팅 환경 두 가지 환경을 모두 제공할 수 있도록 개발하였다.

국가과학기술연구망(KREONET)[2]에서는 이러한 연구자들 간 협업을 위한 빅데이터 고속 전송 환경의 네트워크를 구축하고 그 위에 인공지능 플랫폼 개발을 진행하고 있고, 이 플랫폼을 이용하여 다양한 분야에 플랫폼을 접목 시키는 사업들을 진행하고 있다. 최근에는 농업, 환경 분야에서 필요로 하는 스마트 농업 분야의 빅데이터 관리 및 공유를 위한 플랫폼을 개발하고 있고, 이를 통해 농업 분야의 인공지능 연구 분야 활성화를 목표로 삼고 있다.

인공지능 연구 플랫폼은 기존의 과학 빅데이터를 빠르게 전송하고 이를

전송한 사용자는 바로 인공지능 연구를 수행하기 위해 개발된 플랫폼이다. 이 플랫폼은 인공지능 연구자들 간에 데이터를 교환하고 서로의 연구 환경을 공유할 수 있는 다양한 환경을 개발할 수 있는 협업을 위한 기능들을 제공한다. 본 논문에서는 인공지능 연구 플랫폼의 GPU 컴퓨팅을 이용한 인공지능 처리 성능을 병렬 컴퓨팅과 분산 컴퓨팅을 통한 성능 결과를 보이고, 이를 통한 시스템의 활용성 측면에서 대해 서술하고자 한다.

II. 본론

인공지능 연구 플랫폼은 다양한 오픈소스들을 이용하여 각 기능을 위한 시스템으로 구성하고, 이를 하나의 서비스로 구성하기 위해 컨트롤러를 개발하였다. 그림 1을 보면 인공지능 연구 플랫폼에서 활용하고 있는 솔루션들에서 사용하고 있는 오픈소스를 보여준다. 인공지능 연구 플랫폼은 컨테이너 관리 모듈, 로그인 관리 모듈, 스토리지 관리모듈, 컨테이너 이미지 관리 모듈로 구성되어 있다. 컨테이너 관리 시스템은 Kubernetes, Docker를 사용하고, 로그인 관리를 위해서는 KeyCloak을 활용하며, 스토리지 관리를 위해 Ceph을 사용하고, 이미지를 관리하기 위해 Harbor를 사용하고, 모니터링을 위해 Prometheus, Grafana, Elasticsearch, Kibana를 사용한다.



그림 1. 인공지능 연구 플랫폼 사용 오픈소스 리스트

인공지능 연구 플랫폼에서는 다양한 기능들이 존재하지만, 실제로 사용자들이 사용할 수 있는 컴퓨팅의 수는 한정적이라는 문제가 존재한다. 이를 위해 서버에 분산되어 존재하는 컴퓨팅을 하나로 통합할 수 있는 방법이 필요하다. 이를 위해 오픈소스 Kubeflow[3]를 이용하여 분산 컴퓨팅 환경을 구성한다.

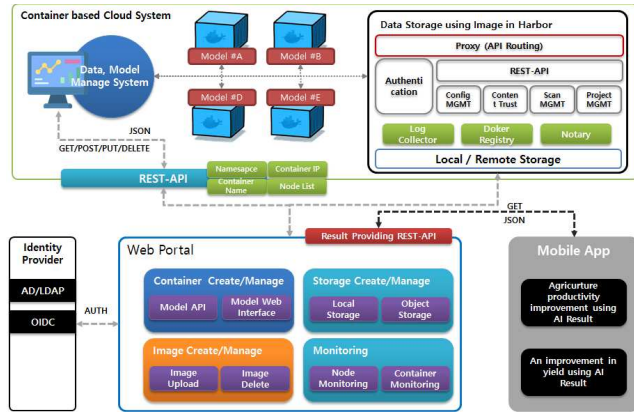


그림 2. 인공지능 연구 플랫폼 구성도

오픈소스들은 개별적으로 대쉬보드를 생성하고 각 시스템을 통합적으로 관리할 수 있는 시스템은 존재하지 않는다. 다만, Kubernetes[4]와 연계하여 구성하였기 때문에 CLI(Command Line Interface)를 이용한 활용은 부분적으로 활용 가능하다. 하지만 이 경우에는 사용자들이 접속하여 CLI를 사용하게 되면 너무 복잡하고 활용도가 떨어질 수 있다. 이를 위해 본 논문에서는 사용자들의 활용도를 확대할 수 있는 Web 기반의 컨트롤러를 개발한다. 그림 2는 이와 같은 인공지능 연구 플랫폼의 구성도이다.

Web 기반의 컨트롤러는 모든 오픈소스시스템들을 통합시키기 위해 활용되지만, 이러한 시스템들의 사용자 측면에서 자동화시켜 쉽게 사용할 수 있도록 구성하고 있다. 사용자가 하나의 연구 환경을 구성하기 위해 필요한 REST-API를 개발하였다.

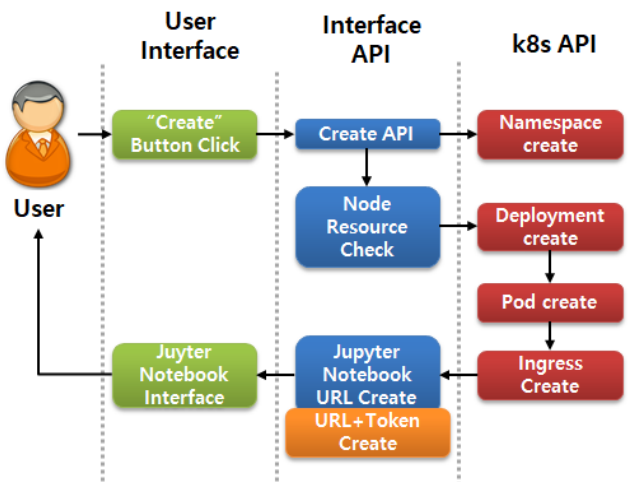


그림 3. 연구환경 생성 시 수행 알고리즘

본 논문에서는 이 플랫폼의 성능을 확인하기 위해 간단한 실험을 수행하였다. 그림 3은 인공지능 연구 플랫폼의 성능을 나타내는 그림이다. 실험 환경은 GPU 개수를 1개부터 4개까지 사용하여 실험을 수행하였으며, GPU의 사용 모델은 TAITAN, 2080Ti를 이용하였다. 실험에 사용된 데이터 셋은 CIFAR-10이며, CNN 딥러닝 모델을 사용하여 트레이닝 시간의 변화를 확인하였다. 수행한 에포치의 수는 50번 수행하였으며, 배치 사

이즈는 600개씩 실험을 수행하였다. 실험의 결과 GPU의 개수가 증가할수록 시간의 감소가 나타나는 것을 확인할 수 있었다. 이 실험 결과를 보면 GPU의 개수가 작을 때는 병렬컴퓨팅의 처리 시간이 짧았으며, 분산컴퓨팅은 GPU의 수가 증가 할수록 처리 시간이 줄어드는 것을 확인할 수 있었다.

이 실험을 통해 GPU를 사용하는 사용자들이 자신의 데이터의 크기에 따라 병렬컴퓨팅을 수행할지 아니면 분산 컴퓨팅을 사용할지 정의 할 수 있다. 대체적으로 데이터의 양이 적은 GPU를 3개 이하로 사용하는 사용자들의 경우 병렬컴퓨팅을 사용하여 컴퓨팅을 사용할 때 더 빠른 결과를 도출할 수 있고, 데이터의 수가 많은 GPU의 수가 4개 이상으로 사용하는 사용자들의 경우 분산 컴퓨팅을 사용하는 경우 더 빠른 결과물을 도출할 수 있다.

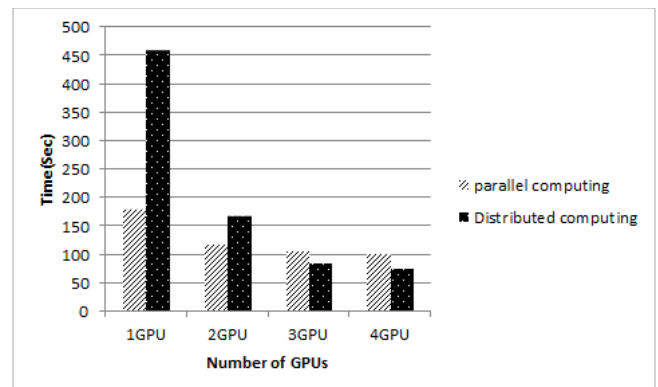


그림 4. 인공지능 연구 플랫폼 성능 결과

III. 결론

본 논문에서는 인공지능 연구자들의 협업을 위한 인공지능 연구 플랫폼을 개발하였으며, 이 플랫폼의 활용도를 높이기 위해 Web 기반의 컨트롤러를 개발하였다. 컨트롤러는 각 모듈의 기능을 담당하는 오픈소스들을 하나의 서비스로 통합하고 사용자 측면에서 사용하기 쉽도록 구성하기 위해 개발되었다. 인공지능 연구 플랫폼의 성능을 검증하고자 병렬컴퓨팅과 분산컴퓨팅을 비교하였다. 병렬컴퓨팅의 경우 GPU의 수가 3개 이하인 경우 활용할 때 높은 성능을 가져오며, 분산컴퓨팅의 경우 GPU의 수가 4개 이상의 경우 거대한 컴퓨팅을 수행해야 하는 경우 활용할 때 빠른 결과를 가져올 수 있다는 것을 확인할 수 있다. 또한 이 두 컴퓨팅 환경 모두 활용성 측면에서 높다는 것을 확인할 수 있다.

ACKNOWLEDGMENT

본 연구는 2023년도 한국과학기술정보연구원(KISTI) 주요사업 과제 로 수행한 것입니다.

참 고 문 헌

- [1] E. Dart, L. Rotman, B. Tierney, M. Hester, and J. Zurawski, "The ScienceDMZ: A network design pattern for data-intensive science," Scientific Programming, vol. 22, no. 2, pp. 173-185, 2014.
- [2] KREONET, <https://www.kreonet.net/>, 01, 2023.
- [3] Kuibeflow, <https://www.kubeflow.org/>, 01, 2023.
- [4] Kubernetes, <https://kubernetes.io/>, 01, 2023.