

API 네트워킹 기반 과학분야 워크플로우 구현방안 연구

석우진¹, 김기현¹, 문정훈¹,

¹한국과학기술정보연구원 과학기술디지털융합본부
wjseok@kisti.re.kr, khkim@kisti.re.kr, jhmoon@kisti.re.kr

Researches on API Networking based Scientific Workflow Implementation

Woojin seok¹, Kihyeon Kim¹, Jeonghoon Moon¹,

¹Div. of Science Digital Convergence, KISTI

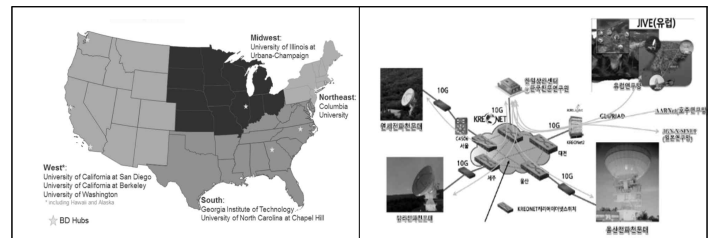
요 약

본 논문은 ICT 기술을 활용한 과학적 연구 적용사례를 분석하고, 데이터의 전송, 저장, 처리를 위한 ICT 기술의 적용모델을 소개하고자 한다. 과학을 다루는 많은 분야에서 플랫폼을 기반으로 분석SW들의 API 네트워킹을 통하여 분석모델들을 연계하여 결과치를 활용하는 방식으로 진화하고 있다. 이러한 API 네트워킹을 위한 ICT 기술들의 구축 모델을 분석하고자 한다.

1. 서론

과학분야는 실험방식을 벗어나서 실험장비에서 생성된 데이터를 중심으로 분석하는 방식으로 이미 발전한 상태이다. 다양한 과학분야에서 실험장비 데이터를 위한 관측 장비, 분석 장비들의 기술발전이 비약적으로 이루어 지고 있으며, 이렇게 생성된 과학 데이터는 별도로 데이터를 센터를 구축하여 관리되고 있다. 미국은 NSF(국가연구재단)의 재원으로 미 전역에 분야별 지역별 빅데이터혁신허브(Big Data Innovation Hub)를 구축하여 (그림1 왼쪽) 과학분석을 촉진하고 있으며, 이를 위하여 고속전송인프라, 대용량저장데이터센터 기술을 적용하고 있다. 또한, 저장된 데이터는 분석을 위하여 고성능 컴퓨팅 자원들과 연계하여 데이터의 저장, 전송, 처리가 SW 적으로 연계될 수 있도록 진화하고 있다. 또한, 천문분야에서는 분산된 천과망원경을 통하여 생성된 과학 빅데이터 전송을 위하여 전용의 고속 네트워크를 구축하여(그림1 오른쪽) 대용량의 관측데이터를 분석 장비에 전송하고 있다.

최근에는, 데이터 기반의 컴퓨팅 분석 방식과 더불어 분석SW들의 상호연계를 통하여 모델을 연계하여 복잡한 결과를 도출하고자 하는 방식으로 진화하고 있다. 본 논문에서는 과학분야에서의 분석SW들의 연계를 위한 ICT 기술의 적용방식을 소개하고자 한다.

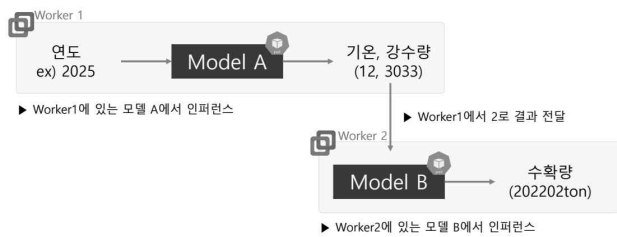


[그림1] 데이터 기반의 과학분석 방식의 진화 사례:
(좌) 과학빅데이터 지역적 관리 (우)천과망원경의 과학빅데이터전송 사례

2. API 네트워킹 기술 필요성

농업 작물모델 분석 분야에서는, 환경 및 관리 조건 하에서 작물의 성장과 발달을 시뮬레이션하기 위한 모델 개발을 중심으로 발전하였다 [1]. 이는 작물과 관련된 다양한 분야에서 의사결정을 지원하는 데 사용될 수 있으며, 기상 조건에 대응하여 단기 운영 관리 관행이나 작물 시스템을 최적화하는 데 필요하며, 또한 기후 변화에 따른 작물 성장 시뮬레이션 결과를 활용하여 식량 안보 정책에 설계된다 [2]. 이러한 모델 분석하고 적용하기 위하여 컴퓨팅 환경을 준비하고 작물 모델별 투입 및 산출 데이터를 관리하기 위한 데이터 저장 환경이 더욱 요구되고 있다.

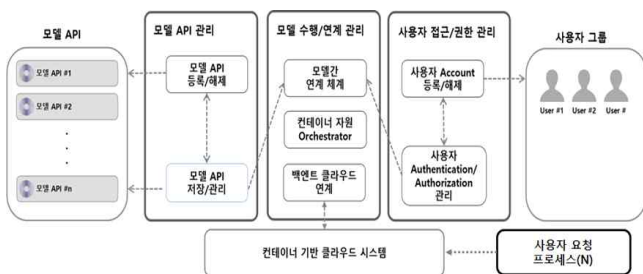
환경분야에서는 예측모델을 구현하기 위하여 다양한 분야의 모델의 연계 결과를 도출하고자 한다. 물 분석 모델, 농업작물 모델, 기상 모델 등 다양한 분야의 다양한 모델의 분석 결과가 상호 연계된 결과치가 필요하다. 이를 위하여 모델의 결과치가 다른 모델의 입력치가 되는 파이프라이닝(그림2) 혹은 몇 개 모델의 결과치가 다른 모델의 입력치로 모두 사용되는 형태의 앙상블 모델이 필요하며, 이를 위한 SW 들이 API 네트워킹이 구현되어야 한다.



[그림2] 파이프라이닝 API 네트워킹 모델(예시)

3. API 네트워킹을 위한 ICT모델 제시

농업분야 작물모델 분석의 경우, 컨테이너 기반 클라우드 중심의 데이터 처리 환경을 제시함으로써 농업연구환경을 별도 구성할 필요없이 클라우드 접근으로 기존 연구환경을 접근하고 재사용하는 편리성을 제공하며, 데이터 처리를 위한 순차적 클라우드 접근 방식을 one 클릭으로 데이터 접근과 분석이 필요하다. 특히 API 네트워킹을 위하여 몇가지 모델들의 결과치를 최종 모델의 입력치로 제시받는 앙상블 형식의 API 네트워킹 요구사항을 반영하였다.



[그림 3] 작물모델 분석을 위한 API 네트워킹 모델 제시

환경분야 모델의 경우, 다양한 분야 모델들의 상호 연계가 요구됨에 따라 파이프라이닝 방식과 앙상블 방식이 모두 요구되고 있다. 복잡한 API 네트워킹을 반영하기 위하여, 모델들간의 API 연계를 포함하여 하나의 파드(POD: 컨테이너의 모음)에 구현되는 방식을 제안한다.



[그림 4] 다양한 모델 연계 일괄 포함 방안

4. 결론

과학분야에서의 ICT 역할은 이미 많은 진전이 진행되고 있으며, 농업이나 환경 분야와 같이 다양한 모델들 간의 연계가 필요하게 되었다. ICT 기술인 API 네트워킹을 통한 연계모델이 이러한 모델들 간의 연계에 좋은 적용모델이 될 것이다. 또한 중요 데이터의 기반으로 컴퓨팅 자원의 동작으로 분석되는 플랫폼의 구현이라는 점에서, 플랫폼의 보안은 중요사항으로 추가적으로 고려되어야 한다. 국가적으로 중요한 농업 및 환경 데이터 등의 보안은 API 네트워킹을 구현하는 과정에서 OAuth 등의 구현 방안등에서 고려되어야 할 것이다.

디지털 농업에서 정부가 추진중인 스마트팜의 농업빅데이터의 일관성 있는 저장을 기반으로 모델들 간의 연계를 제공하는 플랫폼에 좋은 방안으로 제시되기를 기대한다.

Acknowledge

이 논문은 2023년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다. (양자암호통신 기반 공동활용 네트워크 기반 구축, K-23-L04-C02-S01)

참고문헌

- [1] Jame, Y. W. and H. W. Cutforth, "Crop growth models for decision support systems", Canadian Journal of Plant Science. 76(1): 9-19, 1996
- [2] Corbeels, M et al., "Can we use crop modelling for identifying climate change adaptation options ?", Agricultural and Forest Meteorology 256 - 257, 46 - 52. 2018