

CANVAS: A Cloud-based Research Data Analytics Environment and System

Seongchan Kim*, Sa-kwang Song**

*Senior Researcher, Dept. of Machine Learning Data Research, KISTI, Daejeon, Korea

*Adjunct Professor, Dept. of Data & High Performance Computing Science, UST-KISTI, Daejeon, Korea

**Principal Researcher, Research Data Commons Team, KISTI, Daejeon, Korea

**Professor, Dept. of Data & High Performance Computing Science, UST-KISTI, Daejeon, Korea

[Abstract]

In this paper, we propose CANVAS (Creative ANalytics enVironment And System), an analytics system of the National Research Data Platform (DataON). CANVAS is a personalized analytics cloud service for researchers who need computing resources and tools for research data analysis. CANVAS is designed in consideration of scalability based on micro-services architecture and was built on top of open-source software such as eGovernment Standard framework (Spring framework), Kubernetes, and JupyterLab. The built system provides personalized analytics environments to multiple users, enabling high-speed and large-capacity analysis by utilizing high-performance cloud infrastructure (CPU/GPU). More specifically, modeling and processing data is possible in JupyterLab or GUI workflow environment. Since CANVAS shares data with DataON, the research data registered by users or downloaded data can be directly processed in the CANVAS. As a result, CANVAS enhances the convenience of data analysis for users in DataON and contributes to the sharing and utilization of research data.

▶ **Key words:** CANVAS, Cloud Analysis Environment, Korea Research Data Platform, DataON, Workflow, JupyterLab

[요 약]

이 논문에서 우리는 국가연구데이터플랫폼 (DataON)의 분석서비스인 CANVAS (Creative ANalytics enVironment And System)를 제안한다. CANVAS는 연구데이터 분석 자원과 도구가 필요한 연구자들을 위한 개인화된 분석 클라우드 서비스이다. CANVAS는 마이크로서비스 아키텍처 기반으로 확장성을 고려하여 설계하였으며 전자정부프레임워크인 Spring 프레임워크, Kubernetes, JupyterLab 등의 오픈소스 소프트웨어를 이용하여 구축하였다. 구축된 시스템은 여러 사용자에게 개인화된 분석환경을 제공하며 고성능 클라우드 인프라 (CPU·GPU)를 활용하여 고속의 대용량 데이터 분석이 가능하다. 구체적으로 JupyterLab 이나 GUI 워크플로우 환경에서 데이터 모델링 및 처리가 가능하다. CANVAS는 DataON과 데이터가 공유되므로 사용자가 등록하거나 다운로드 받은 연구데이터는 CANVAS에서 바로 분석을 수행할 수 있다. 이로서 CANVAS는 DataON 사용자의 데이터 분석 편의성을 높이고 연구데이터 공유·활용 활성화에 기여한다.

▶ **주제어:** CANVAS, 클라우드 분석환경, 국가연구데이터플랫폼, DataON, 워크플로우, JupyterLab

-
- First Author: Seongchan Kim, Corresponding Author: Sa-kwang Song
 - Seongchan Kim (sckim@kisti.re.kr), Dept. of Machine Learning Data Research, KISTI; Dept. of Data & High Performance Computing Science, UST-KISTI
 - Sa-Kwang Song (esmallj@kisti.re.kr), Research Data Commons Team, KISTI; Dept. of Data & High Performance Computing Science, UST-KISTI
 - Received: 2021. 09. 30, Revised: 2021. 10. 22, Accepted: 2021. 10. 22.

I. Introduction

데이터의 분석과 활용을 통해 새로운 이론이나 현상을 발견하는 연구 방법론을 데이터기반 연구(Data-intensive Research)라고 한다. 이는 '데이터 중심연구'라고 불리어지는 제 4세대 연구 방법론이며 현대의 과학기술 최신 연구 방법론이다. 데이터 중심연구를 가능하게 하는 원천은 연구 데이터의 급격한 증가이다. re3data.org에 등록된 전세계 연구데이터 리포지토리(Repository)의 수는 2021년 5월 기준 3667개 이상이며 지난 5년간 꾸준히 증가하고 있다.

한편, 공적자금을 투입해 수행된 연구과제에서 생산된 데이터를 디지털화하여 개방 및 공유하고 공동연구를 활성화함으로써 과학연구를 발전시키고자 하는 오픈사이언스 실현을 위해 주요 선진국들은 여러 정책들을 추진하고 있다. 우리나라도 최근 과기정통부 주도로 국가 연구데이터의 공유·활용 활성화를 추진하고 있으며[1], 국가과학기술연구회에서도 출연(연) 발전방안의 일환으로 연구데이터 빅데이터화 추진 및 대응에 나서고 있다[2].

과학기술정보·데이터 연계·융합 인프라 및 공유 체계 구축의 임무를 지닌 한국과학기술정보연구원(KISTI)은 오픈 사이언스 패러다임에 대응하고 연구데이터의 디지털 공유·활용 체계 구축을 위해 노력하여 왔다[3]. 특별히 국가 연구데이터 공유·활용 전략과 출연(연) 연구데이터 빅데이터화 정책에 대응하기 위해 국가 연구데이터 공유·확산 체제 구축하였다. 세부적으로는 연구데이터 관리·공유 정책과 연구데이터 리포지터리(Institutional Data Repository, IDR) 보급 및 확산, 국가연구데이터플랫폼(DataON) 구축 및 운영을 2018년부터 수행해오고 있다[4, 5].

DataON¹⁾은 연구데이터 검색, 등록·공유, 관리, 연계, 분석 등의 다양한 서비스를 제공하고 있다. 먼저 검색서비스에서는 데이터셋, 표, 그림, 소프트웨어 등의 다양한 형식의 연구데이터를 패싯(Facet) 검색, 맵 검색 등의 편의 기능과 함께 제공하고 있다. 또한 사용자가 작성한 연구데이터를 업로드·저장·출판 등의 편리한 관리를 위해 마이드라이브(MyDrive) 서비스를 제공하고 있으며, 데이터 등록 기능을 통하여 DataON 리포지토리에 메타데이터와 함께 원시데이터를 게시하여 타 연구자들에게 검색 가능한 자료원으로 등록할 수 있도록 하였다. 이때 DataON 데이터 관리자는 사용자가 등록한 데이터의 무결성을 검토한 후 등록을 최종승인하게 된다. 연구데이터 연계 측면에서는 흩

어져 있는 연구데이터 리포지토리를 DataON에 등록하여 메타데이터를 연계하는 기능을 제공하고 있다. 연계 기능은 각 기관의 데이터 리포지토리를 국가연구데이터플랫폼에 등록하여 다양한 데이터가 검색이 가능하도록 기능을 제공하고 있다. DataON은 다양한 형태의 데이터 메타데이터 스키마를 통합하여 관리할 수 있는 통합 메타데이터 스키마를 정립하였으며, 연계 시 각 리포지토리의 메타데이터 스키마와 통합메타데이터 스키마간 매핑(mapping)을 수행하여 메타데이터의 연계를 수행하고 있다. 데이터 분석기능으로는 워크플로우 기반의 GUI 개발환경 및 JupyterLab 기반의 대화형 개발환경에서 분석이 가능한 개인분석환경을 제공하고 있다.

본 연구에서는 국가차원의 연구데이터 공유·확산체제 활성화 일환으로 구축된 DataON의 연구데이터 분석서비스인 CANVAS (Create Analysis environment and System)²⁾의 구축 현황과 이슈에 대해서 소개한다. CANVAS는 연구데이터 분석 자원·도구가 필요한 연구자들을 위하여 개인화된 분석 클라우드를 제공하는 서비스이다. CANVAS는 DataON에서 공유된 혹은 연구자들이 업로드한 데이터를 바로 분석할 수 있게끔 하는 체계를 제공한다. 분석 수행 후 연구결과 및 데이터는 및 다시 DataON에 공유할 수 있도록 하여 연구데이터 공유 원스탑(One stop) 체계를 완성하도록 구현되었다. CANVAS가 갖는 특징은 다음과 같다.

- JupyterLab 활용 개인 맞춤형 분석환경 구성 및 분석작업 수행
- GUI 기반 워크플로우 모델링 및 분석작업 수행
- 분석작업 실행 상태 및 결과 모니터링
- 고성능 클라우드 인프라(CPU·GPU) 활용 고속·대용량 분석
- 분석 결과 다운로드 및 가시화
- 다양한 분야·종류의 분석도구(소스코드, 애플리케이션, 워크플로우)
- DataON과 밀접 연결되어 등록하거나 다운로드 받은 연구데이터를 분석환경(CANVAS)에서 바로 분석

본 논문은 2장에서는 관련연구를 3장에서는 CANVAS의 시스템 아키텍처와 제공하는 서비스를 소개한다. 4장에서 결론 및 향후연구를 제시한다.

1) <https://dataon.kr>

2) <https://dataon.kisti.re.kr/canvas>

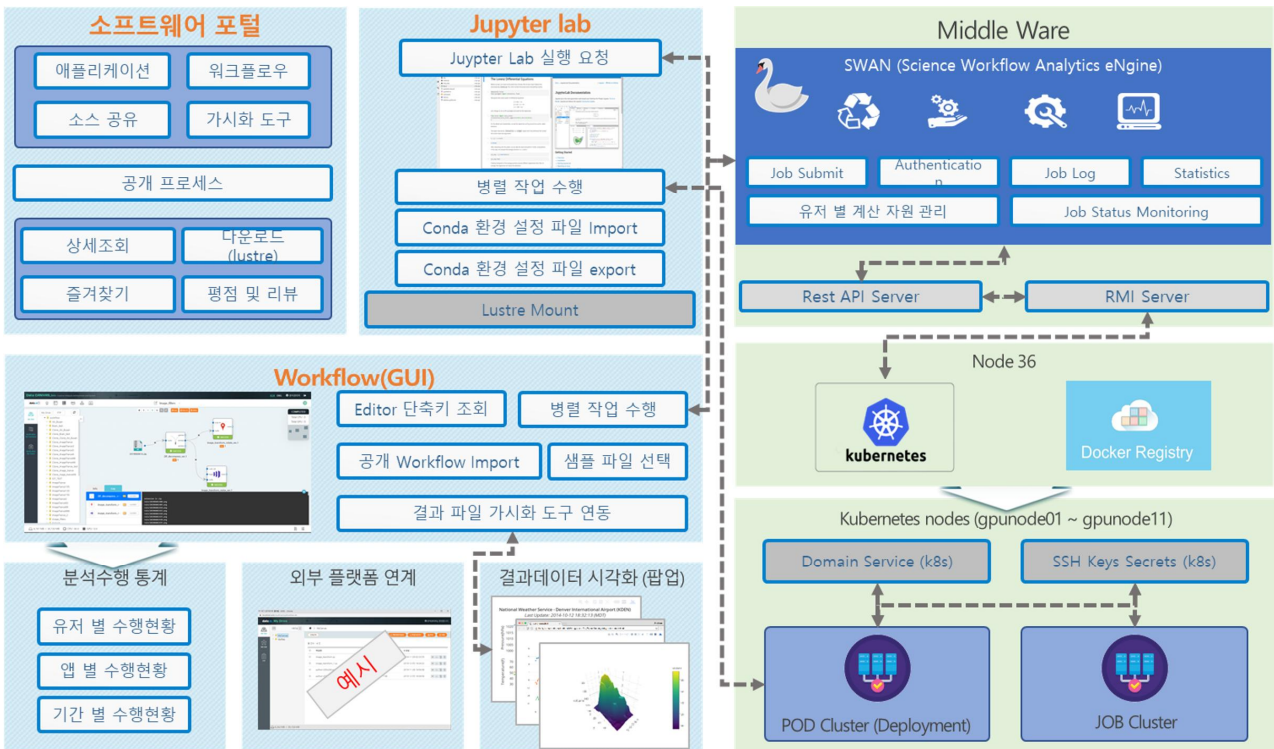


Fig. 1. CANVAS System Architecture

II. Related Work

이 장에서는 국내외 공공 및 민간분야의 데이터플랫폼에 대해 살펴본다. 데이터플랫폼이 제공하는 여러 기능 중 데이터 분석기능에 한정하여 동향을 분석한다.

국내 공공분야에서는 AI 허브와 데이터스토어가 있다. AI 허브(aihub.or.kr)는 한국지능정보사회진흥원(NIA)가 구축한 AI 기술 및 제품 개발에 필요한 AI 인프라(데이터, SW API, 컴퓨팅 자원)를 지원하는 AI 통합 플랫폼이다[6]. AI 허브에서는 AI 컴퓨팅 지원의 일환으로 GPU 기반의 컴퓨팅자원을 지원하며 사용자가 신청서를 접수하여 선정되면 접속하여 사용할 수 있는 리눅스 기반의 클라우드 서버를 제공한다. 또한 AI SW 지원으로 언어처리, 음성지능, 시각지능, 대화처리등 API를 제공하여 사용자들의 데이터 분석을 지원하고 있다. 데이터스토어(datastore.or.kr)는 데이터를 온라인으로 판매하거나 구매할 수 있게 하는 데이터 오픈마켓으로 한국데이터산업진흥원이 구축 및 운영하고 있다[7]. 데이터스토어는 데이터 간편분석 기능을 제공하고 있으며 이는 엑셀파일로 이루어진 샘플데이터를 업로드 하면 차트생성 및 맵핑과 포매팅 등의 기능을 제공한다.

민간분야에서는 한국데이터거래소(KDX, kdx.kr)가 KDX에서 구매하고 판매하는 데이터와 로컬 데이터를 인

스턴스로 전송하여 데이터를 분석할 수 있는 환경을 제공한다[8]. Hadoop과 Spark에 기반하여 분석툴(RStudio, Jupyter) 및 시각화툴(Zeppelin)을 제공하고 있다. 데이터 분석툴 인스턴스는 유료이며 사용시간에 비례하여 요금을 책정하고 있고, 데이터 시각화툴은 24종의 차트를 만들 수 있도록 지원하고 있다.

이밖에 공공데이터포털(NIA)[9], 서울시 열린데이터 광장[10] 등의 서비스에는 공공기관에서 취득한 데이터를 국민들에게 개방하여 파일, 오픈API 등의 방식으로 제공하고 있다. 하지만 이 포털들은 시각화 이외의 다양한 데이터 분석기능은 제공하고 있지 않다. 각 도메인 별로 의료 도메인에서는 임상데이터를 분석하여 사례들을 공유할 수 있는 플랫폼을 구현한 사례가 있으며[11], 지적분야(GIS)에서도 지역공간정보를 효율적으로 분석하기 위해 데이터 기반의 분석 플랫폼을 개발하여 활용을 활성화 하고자 하였다[12].

해외 사례로 EU의 EOSC(European Open Science Cloud)는 EU의 과학 연구를 지원하기 위해 연구데이터를 검색하고 분석도구를 제공할 수 있게끔 하는 연합된 개방형 환경을 제공한다[13]. 이를 위해 EOSC 포털에서는 EOSC 정보 및 리소스에 대한 관문으로 마켓플레이스에서는 데이터를 처리 및 분석(Processing & Analysis) 할 수 있는 서비스를 제공하고 있다. EOSC의 특징은 각 서비스 프로바이더(Provider)들이 분석 자원이나 도구들을 등록

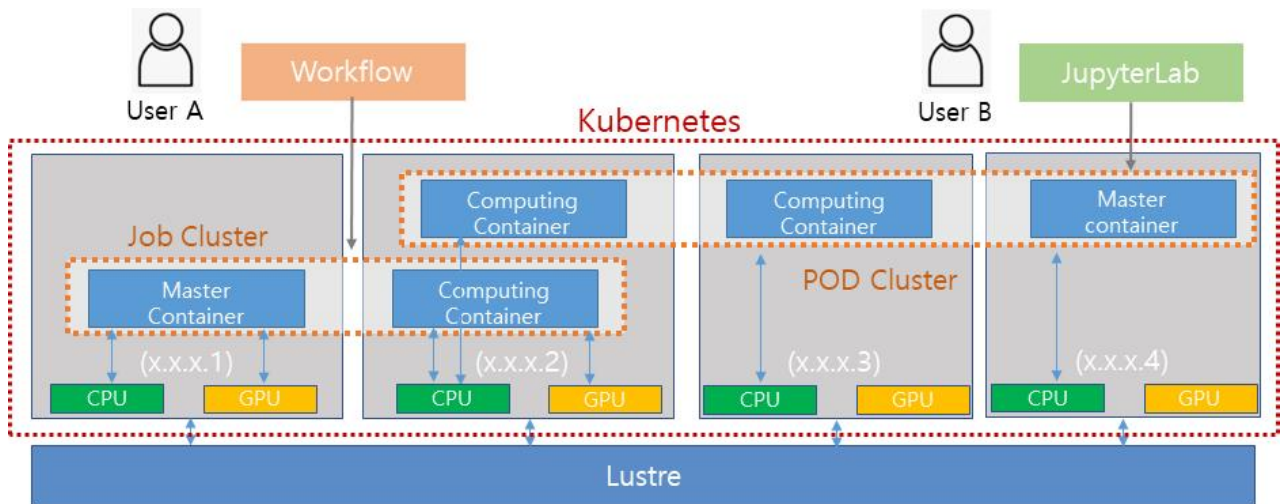


Fig. 2. Multi-user Senario

하고 제공할 수 있다는 것이며 사용자들은 등록된 도구들을 이용해 분석을 수행한다.

호주의 ARDC(Australian Research Data Commons)는 연구데이터를 다루는 포털인 RDA(Research Data Australia), 연구 클라우드 연합체계인 Nectar Research Cloud, Identifier 서비스 등을 제공하고 있다[14]. 이 중 Nectar Research Cloud는 컴퓨팅 인프라와 소프트웨어를 연구 커뮤니티에게 제공한다. 연구자들은 Nectar cloud dashboard를 통해 The University of Melbourne, Monash University 등에서 제공하는 컴퓨팅 인프라에 접속하여 분석을 수행할 수 있다.

해외사례의 마지막 사례는 GOSC(Global Open Science Cloud)이다. GOSC는 최근들어 각국의 오픈사이언스 클라우드를 표준화 하고 서로 연결하려는 움직임속에서 제안된 체제이다. GOSC에서는 분석자원, 데이터, 모델 등을 공유하고 협업할 수 있도록 연결을 목표로 하고 있다.

III. The Proposed Scheme

국가연구데이터플랫폼의 분석서비스(CANVAS)는 연구데이터를 분석할 수 있는 환경을 연구자들에게 제공하는 플랫폼이다. CANVAS는 다음의 목표를 달성하도록 서비스를 디자인 하고 시스템을 구축하였다.

- JupyterLab을 활용한 다양한 분석환경 구성 및 분석작업 수행
- GUI 기반 워크플로우 모델링 및 분석작업 수행
- 분석작업 실행 상태 및 결과 모니터링
- 고성능 클라우드 인프라(CPU-GPU) 활용 고속-대용

량 분석

- 분석 결과 다운로드 및 가시화 기능 제공
- 다양한 분야-종류의 분석도구(소스코드, 애플리케이션, 워크플로우) 제공

1. System Architecture

그림 1은 CANVAS의 시스템의 아키텍처를 보여준다. 기본적으로 CANVAS는 3개의 GPU(V100)가 탑재된 Linux 서버 11대를 분석자원으로 활용한다. 이 분석자원들은 Kubernetes를 활용하여 On premise 방식으로 Private cloud로 구성하였다. Kubernetes는 컨테이너를 배포하는데 필요한 오케스트레이션 및 관리기능들을 제공하며, 각 사용자의 요청한 분석작업 및 환경은 Pod나 Job으로 생성하여 Kubernetes가 관리하도록 하였다. 사용자가 JupyterLab 서비스를 로딩하면 해당 Docker Image로 Pod를 생성하며 Kubernetes가 관리하게 된다. 이때 사용자에게 할당된 저장소 (Lustrre에 사용자 계정에 할당된 약 100Mb(혹은 그 이상)의 공간) 와 연동이 되어 사용자 파일을 접근하여 처리할 수 있다.

Workflow 분석환경에서는 사용자가 구성한 분석작업을 실행하게 되면 미들웨어인 SWAN(Science Workflow Analytics eNgine)이 작업을 받아서 처리하게 된다. SWAN은 해당 작업을 Rest API를 통해서 Kubernetes Job으로 변환하여 분석을 수행하게 된다. 이때에도 Kubernetes가 해당 Job cluster를 관리한다.

CANVAS는 복수노드에서 실행 가능한 Pod Cluster나 Job Cluster를 지원하도록 하였다. 그림 2에서는 여러명의 사용자가 요청한 Pod나 Job Cluster를 복수의 노드에서 수행하는 모습을 보여주고 있다. 사용자 A는

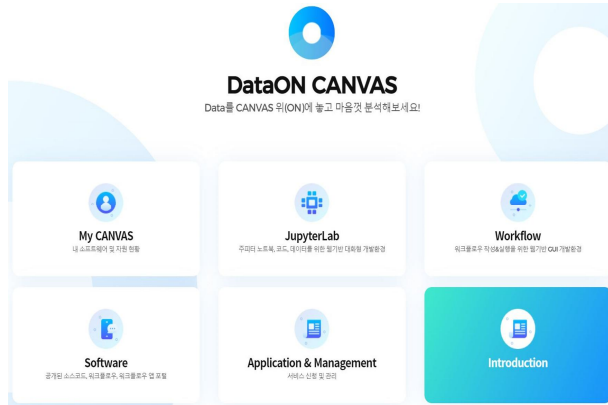


Fig. 3. Main page of CANVAS

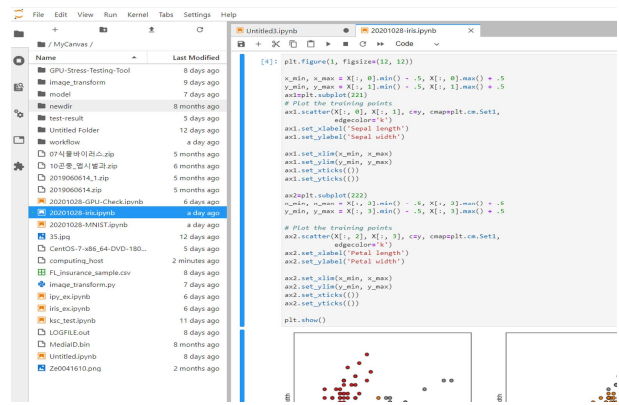


Fig. 4. JupyterLab

Workflow 분석환경에서 2개의 Node에서 CPU와 GPU를 동시에 사용하는 Job Cluster를 수행하고 있으며 사용자 B는 JupyterLab 환경에서 Node 3개의 CPU만 활용하는 Pod Cluster를 구성하여 사용하는 모습을 보여주고 있다. 이와 같은 메커니즘으로 동시 접속한 여러명의 사용자 작업을 관리하기 위해서 Kubernetes의 자원 관리 기능을 활용한다. Kubernetes는 각 Node의 활용도를 모니터링 하여 Utilization ratio가 적은 Node에게 작업을 자동으로 할당한다.

2. Services

DataON CANVAS는 아래 나열한 주요 서비스를 제공하고 있다. 그림 3은 CANVAS의 메인페이지를 보여주고 있으며 메인페이지에서는 각 서비스로의 진입이 가능하다.

- My CANVAS: 내 소프트웨어 및 자원 현황
- JupyterLab: 주피터 노트북, 코드, 데이터를 위한 웹 기반 대화형 개발 환경
- Workflow(GUI): 워크플로우 작성 및 실행을 위한 웹 기반 GUI 개발 환경
- Software 포털: 공개된 소스코드, 워크플로우, 워크플로우 앱 포털
- Application & Management: 서비스 신청 및 관리

2.1 JupyterLab

JupyterLab은 주피터 노트북, 코드, 데이터를 위한 웹 기반 대화형 개발환경이다. 사용자의 손쉬운 데이터 분석 및 프로그래밍을 포함한 분석모델 개발을 위해 JupyterLab을 제공한다(그림 4). 사용자는 셀(Cell) 단위의 코드(Python) 실행할 수 있으며 제공받은 가상환경

(JupyterLab을 실행하고 있는 Pod)에 터미널을 통해 접근할 수도 있다. 서비스 승인 시 부여받은 유형별(CPU·GPU) 계산자원(Pod)에 접근하여 분석작업을 실행하거나 JupyterLab을 실행하고 있는 Pod 자체에서도 작업을 실행할 수 있다. JupyterLab 사용자는 본인의 분석환경을 구성하기 위해 Conda 가상환경³⁾을 이용하여 파이썬 기반의 환경을 구성할 수 있다. CANVAS에서는 기본 Conda 환경을 제공하여 분석을 수행할 수 있도록 하였지만 사용자 개인만의 개발환경을 구성하고 분석을 처리 할 수 있도록 하였으며 구성환경을 export 하거나 import하여 환경 구성의 자유도를 높이도록 하였다. 또한 DataON 마이드라이브와 연동되어 DataON에 올려진 데이터를 CANVAS에서 바로 탐색하고 분석에 활용할 수 있도록 하였다.

2.2 Workflow

Workflow 분석환경은 앱의 연결구조인 워크플로우 개발을 위한 JSplumb 및 웹기반의 GUI 개발환경이다(그림 5). 이 개발환경은 상대적으로 비전문가의 분석작업의 용이성 및 분석작업 수행의 시각화를 위해서 제공된다.

Workflow는 데이터와 분석 앱을 수행절차대로 나열한 작업절차의 묶음이며, 이 환경에서는 Drag & Drop 방식으로 워크플로우를 작성한다. 소프트웨어 포털에 공개된 앱을 사용자가 가져다가 본인의 워크플로우를 구성할 때 배치 및 연결하여 사용할 수 있다. 데이터는 DataON에서 다운로드한 데이터나 MyDrive의 사용자 데이터를 가져와서 워크플로우에 배치할 수 있으며, 사용자 편의성을 위해 앱의 입력으로 작동할 수 있게끔 연결을 지원하고 있다. 연결된 데이터와 앱들은 워크플로우로 저장되며 각각의 워크플로우는 워크플로우(앱) 상세조회 및 실행로그 확인

3) Conda: 다양한 프로그래밍 언어 (Python, R, Ruby, Lua, Scala, Java, JavaScript, C / C ++, FORTRAN 등)에 대한 패키지, 종속성 및 환경 관리 툴

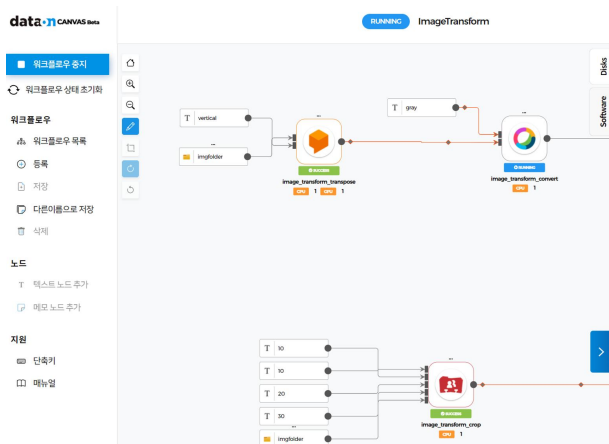


Fig. 5. Workflow Composing and running panel

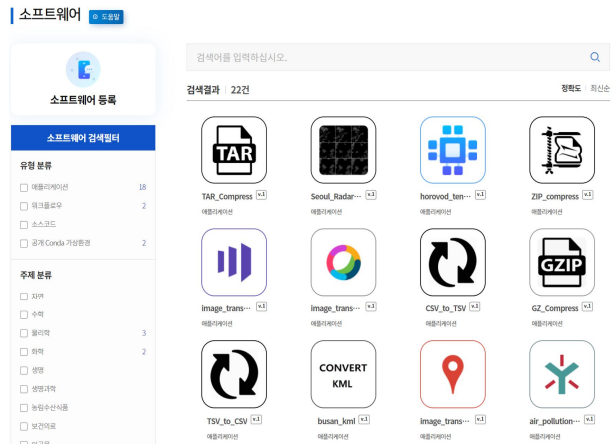


Fig. 6. Software Portal

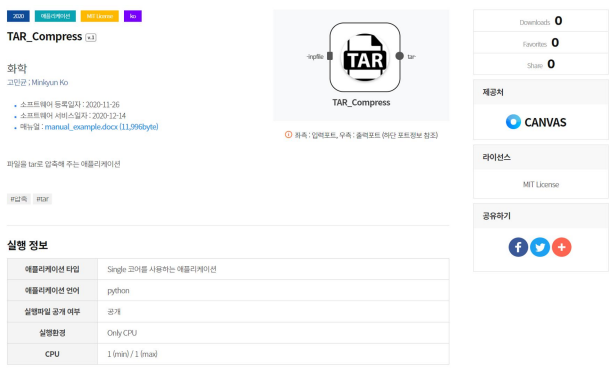


Fig. 7. Detailed Page of Software

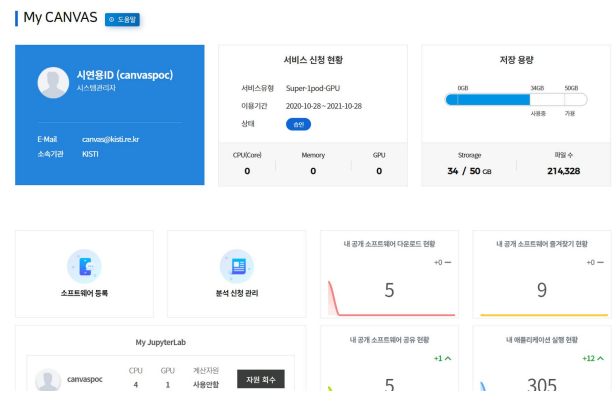


Fig. 8. My CANVAS: Resource Monitoring Page

이 가능하다. 워크플로우 불러오기, 실행, 저장, 삭제, 이력 조회 등의 기본기능 및 워크플로우 작성 편의기능(확대 축소, 단축키, 메모 등) 또한 제공하여 사용자 편의성을 극대화 하도록 하였다.

Workflow 분석환경에서 만들어진 Workflow 나 JupyterLab에서 생성된 소스코드들은 개발자가 소프트웨어 포털에 등록하여 공유할 수 있으며, 또한 개발자들은 포털에 공유된 앱이나 워크플로우를 다운로드 받아 본인의 분석 환경에서 수행하거나 재개발을 할 수 있다.

2.3 Software Portal

그림 6에서 보여지는 Software 포털은 제공자에 의해 공개되어 CANVAS 혹은 외부시스템에서 활용 가능한 소스코드, 워크플로우, 워크플로우 앱의 정보를 등록하고 제공하는 포털이다. CANVAS에서는 3 종류의 소프트웨어(앱): 소스코드, 워크플로우, 워크플로우를 등록하고 검색할 수 있다. 주제·유형 별 소프트웨어 검색이 가능하며 소프트웨어 상세보기(실행파일 미리보기, 포트정보, Conda 가상환경 정보)가 가능하다. 소프트웨어 다운로드, 즐겨찾기,

SNS 공유, 평가 및 리뷰작성 기능 또한 제공한다. 유사 카테고리 및 개발자 소프트웨어 추천기능도 갖추고 있다.

그림 7은 앱이나 워크플로우를 클릭했을 때 나타나는 상세페이지를 보여주고 있다. 상세페이지에는 개발자가 작성한 각 소프트웨어의 설명과 실행정보, 입력, 출력 포트, 라이선스 등에 관한 정보가 나타난다. 또한 이 소프트웨어의 조회, 다운로드, 좋아요 수 등의 부가정보가 나타난다.

2.4 My CANVAS

My CANVAS는 신청·승인된 서비스(분석자원) 현황, SW 등록현황, 분석작업 수행 현황 등을 모니터링 할 수 있는 개인 페이지이다. 이곳에서 분석 서비스 신청 및 관리, 서비스 신청현황 및 저장용량 조회 등이 가능하다. 또한 내가 등록한 소프트웨어, 리뷰, 즐겨찾기 조회를 할 수 있으며, 내 소프트웨어 다운로드, 공유현황도 조회 할 수 있다. 워크플로우 수행내역 및 Conda 가상환경 정보도 볼 수 있다.

IV. Conclusions

국가연구데이터플랫폼(DataON)은 연구데이터를 검색, 공유, 관리, 연계 하는 것 뿐 아니라 연구데이터의 활용을 극대화하기 위하여 연구데이터 분석환경인 CANVAS를 구축하여 연구자에게 제공하였다. 연구자들에게 클라우드 기반의 데이터 분석을 지원함으로써 연구 생산성을 높이고 종국에는 국가 R&D 투자 효율성 향상에 기여하였다.

향후 발전방향으로는 분석서버 및 저장소 등 분석자원의 증설이 필요하다. 또한 더 많은 사용자 풀(Pool) 확보 및 서비스의 활용도를 높이고 앱 및 데이터가 공유되도록 하는 이 필요하다. 또한 사용자들이 분석결과를 시각화하거나 직접 데이터를 편집할 수 있는 기능 도입을 고려하고 있다. 나아가서는 기관이나 사용자들이 데이터 뿐 아니라 자신의 분석자원(모델, 하드웨어 인프라 등)을 등록하여 공유·활용 할 수 있도록 하는 체계인 KRDC(Korea Research Data Commons) 구축을 고려하고 있다.

ACKNOWLEDGEMENT

This research was supported by Korea Institute of Science and Technology Information (KISTI).

(연구데이터와 인프라의 공유·활용 체제 구축, K-21-L01-C04-S01)

REFERENCES

- [1] National Science and Technology Council, "Strategies for sharing and utilizing research data to promote innovative growth," Jan. 2018.
- [2] National Science and Technology Research Society (NST), "Science and technology contribution (research) development plan," Jan. 2018.
- [3] M. Choi, and S. Lee, "Research Data Management Status of Science and Technology Research Institutes in Korea," Data Science Journal, Vol. 19, No. 1, pp. 1-11, Aug. 2020. DOI: 10.5334/dsj-2020-029
- [4] M. Cho, S. Kim, J. Kim, S.-K. Song, "Development of Pilot Service for Sharing Research Data," Proceedings of the Korea Computer Congress, pp. 1132-1134. Korea, Dec. 2019.
- [5] M. Cho, H. Yim, S. Kim, J. Kim, S.-K. Song, "National Research Data Platform Service Strategy for Pandemic Response," Proceedings of the Korea Computer Congress, pp. 1034-1035, Dec. 2020.
- [6] National Information Societ Agency (NIA), AI Hub, <https://aihub.or.kr/>
- [7] KOREA Data Agency, Data Store, <https://www.datastore.or.kr/>
- [8] KDX. Korea Data Exchange, <https://kdx.kr/main>
- [9] National Information Societ Agency (NIA), Public Data Portal, <https://www.data.go.kr/>
- [10] Seoul Metropolitan Government, Seoul Open Data Plaza, <https://data.seoul.go.kr/>
- [11] J.-H. Kim, "Development of Clinical Analytics Platform for the Sharing and Analysis of Clinical Cases," The Journal of Next-Generation Convergence Technology Association, Vol. 4, No. 1, pp. 5-14, Feb. 2020. DOI: 10.33097/jncta.2020.04.01.5
- [12] J.-S. Lim, S. Kim, S. H. Bae, K.-H. Kim, D.-K. Won, "Research on regional spatial information analysis platform about NTIS raw data," Journal of Cadastre & Land InformatiX, Vol. 50, No. 2, pp. 21-35, Dec. 2020. DOI: 10.22640/LXSIRI.2020.50.2.21
- [13] European Open Science Cloud (EOSC), <https://eosc-portal.eu/>
- [14] Australian Research Data Commons (ARDC), <https://ardc.edu.au/>

Authors



Seongchan Kim received the B.S. degree in Computer Science from Chonbuk National University, Jeonju, South Korea, in 2004, and the M.S. degree in Information and Communication Engineering and the Ph.D.

degree in Knowledge Service Engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2010 and 2017, respectively. Dr. Kim is currently a senior researcher in the Department of Machine Learning Data Research, Korea Institute of Science and Technology Information, Daejeon. Also, He is an Adjunct Professor in Data & High Performance Computing Science at University of Science and Technology (UST), Daejeon. His research interests include machine learning and deep learning applications for science and technology information.



Sa-kwang Song received the B.S. degree in statistics and the M.S. degree in computer science from Chungnam National University (CNU), Daejeon, South Korea, in 1997 and 1999, respectively, and the Ph.D. degree in

computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, in 2011. Dr. Song is currently a Principal Researcher and the Team Leader of the Research Data Commons Team, Research Data Sharing Center, Korea Institute of Science and Technology Information (KISTI), Daejeon, and also a Professor with the Department of Data and High Performance Computing Science, University of Science and Technology (UST), Daejeon. His current research interest includes artificial intelligence, machine learning, big data, text mining, natural language processing, information retrieval, and semantic web.