

신경회로망 이해 및 분석을 위한 멀티 모달 융합 DB 구축

한국뇌연구원
김남욱

통합 분석 DB station 소개

연구배경

- 차세대 바이오 분석 기술의 발전으로 다양한 영상, 이미지, 분자 데이터 등이 생성되고 그 양이 기하급수적으로 증가하는 추세이다. 이러한 방대한 뇌의 기능, 구조, 분자 데이터 등 뇌 회로망 데이터를 통합적으로 처리 및 분석 가능한 *SW* 기술의 부재로 인해 기존 연구의 한계점이 있었다. 특히 뇌에서 의사 결정과 관련한 전전두엽 특화 신경회로에 대한 분자수준에 대한 연구는 최근 수행되고 있으며, 여러 수준의 데이터를 시스템 차원에서 통합적으로 분석하기 위한 DB 통합 플랫폼 개발이 필요하다. 본 프로젝트는 뇌 의사결정에 있어서 전전두엽 신경회로로부터 생산되는 다양한 이미지, 활성 및 분자 데이터들을 수집 및 처리하고 구조에 따른 통합적으로 분석 및 가시화 기술을 개발하여 DB화함으로써 전전두엽 뇌 신경회로 규명을 위한 인터랙티브 데이터 플랫폼을 구축하는 것이다.

최종목표: 의사결정에 중요한 전전두엽 신경회로망 규명을 위해 멀티모달 정보의 저장, 추출, 분석이 가능한 웹 기반 인터랙티브 DB 플랫폼 구축

:클라우드 기반으로 뇌의 구조적, 기능적 분석 결과 및 연결성 수치화, 시각화 데이터를 효율적으로 저장하고 다양한 어플리케이션들과 통합을 지원하는 Data Station을 구축하며 이를 활용한 글로벌 뇌 연구 협력 인프라 구축

○ 목표 1. Cloud-Native 인프라 고도화 및 초고속 대용량 처리 시스템

- Cloud native 고도화, 대용량 데이터 압축 및 전송 플랫폼, 분석 파이프라인

○ 목표 2. 웹 기반 인터랙티브 플랫폼 및 분석 파이프라인 구축/서비스

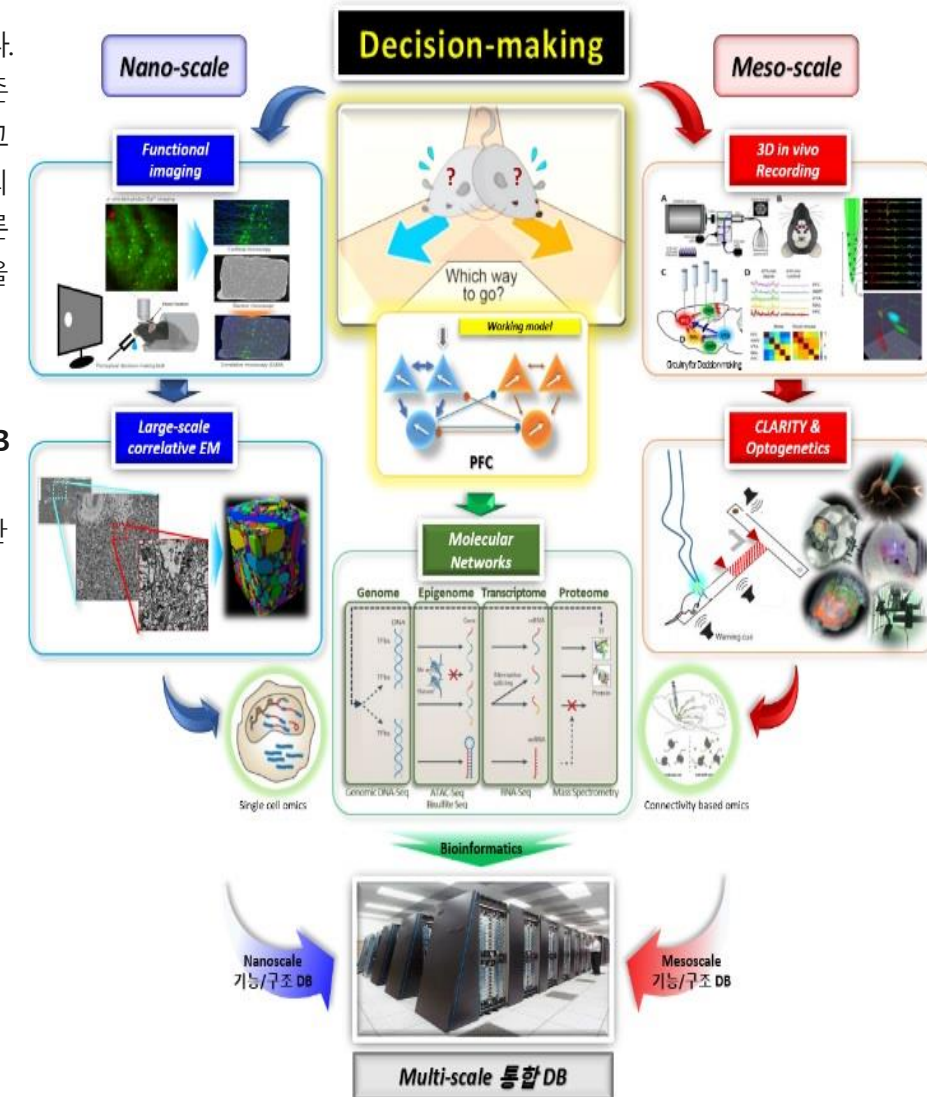
- 바이오 분석을 위한 jupyterlab(연구자 교육용 포함), genepattern과 galaxy 기반 파이프라인 구축

○ 목표 3. 국가 뇌 연구 네트워크 구축 및 운영

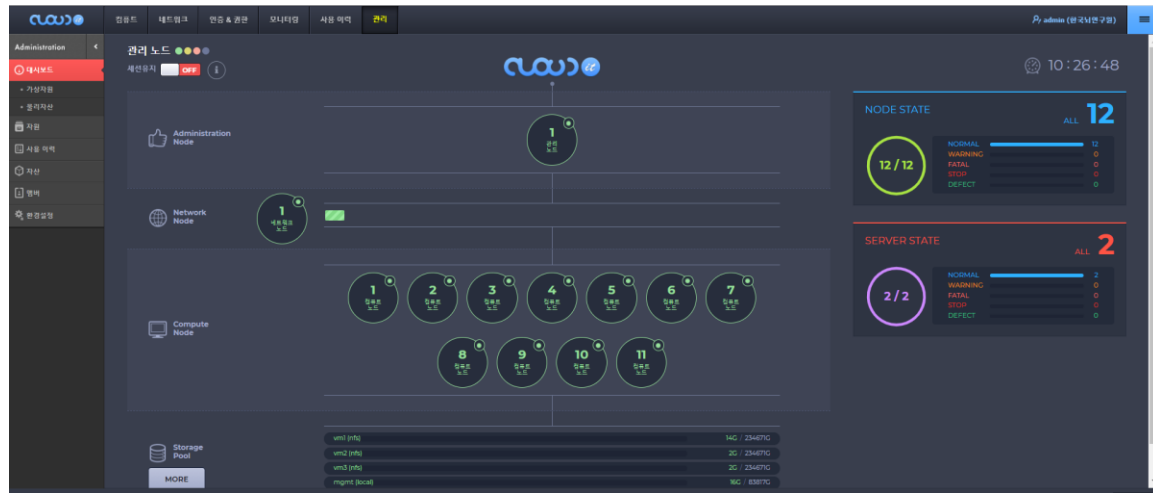
- 국제 뇌연구 기관 협력을 위한 국내 인프라 활용, 뇌과학 데이터 공유를 지원하는 인프라 구축

○ 목표 4. 멀티스케일 영상/이미지 통합분석 플랫폼 및 파이프라인 구축

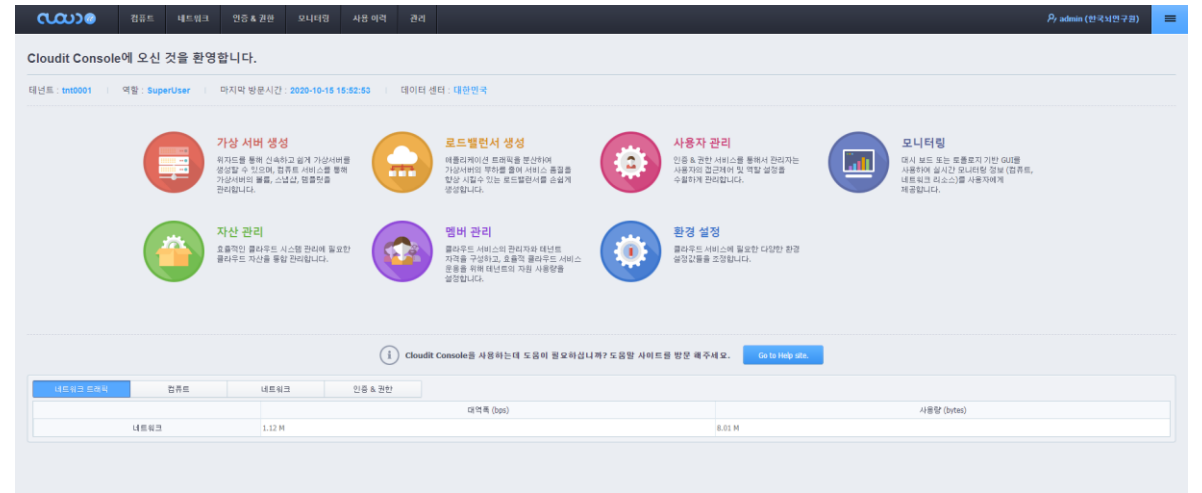
- 웹 기반 멀티스케일 이미징 분석 고속화/가시화/자동화기술 구현



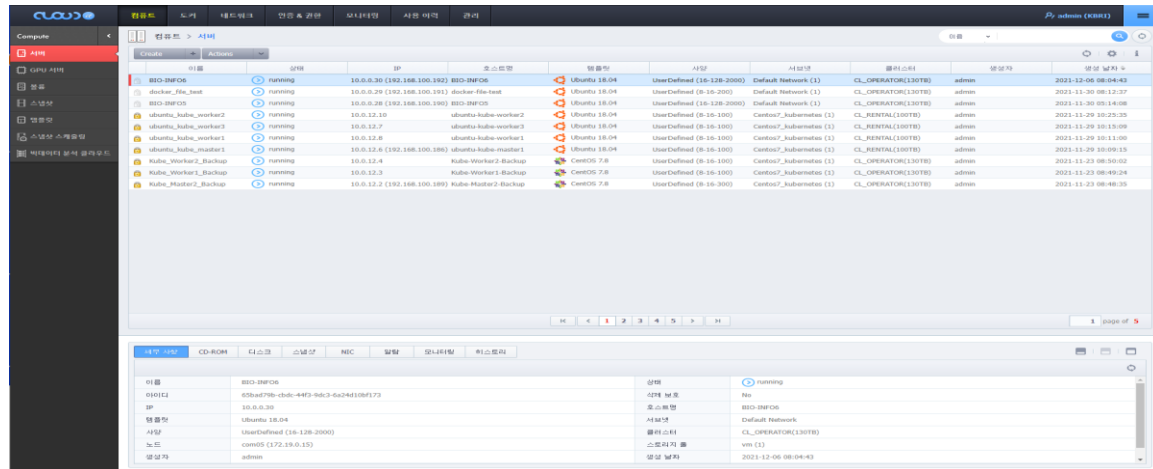
클라우드 플랫폼 (이노그리드사의 프라이빗 클라우드 도입)



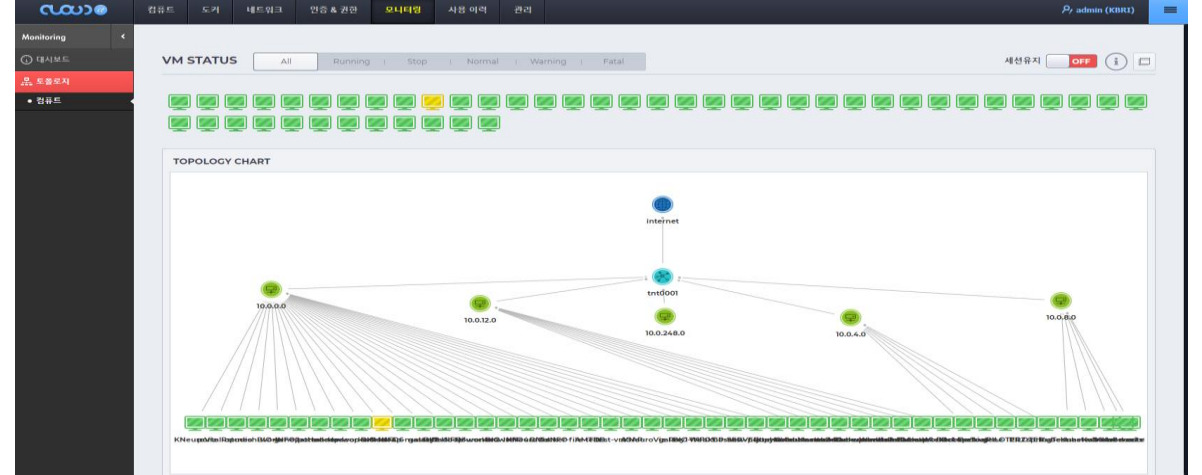
서버 자원 상태



가상화 자원 카테고리



VM 자원 제어



네트워크 자원 모니터링

통합 분석 DB station 클라우드 인프라 실장도



<랙 전면(기존)>



<랙 전면(좌)>



<랙 전면(우)>

DATA												
Port	1	3	5	7	9	11	13	15	17	19	21	23
Node	MGR 1	MGR 3	COM 2	COM 4	COM 6	COM 8	COM 10	COM 12	COM 14			
Port	2	4	6	8	10	12	14	16	18	20	22	24
Node	MGR 2	COM 1	COM 3	COM 5	COM 7	COM 9	COM 11	COM 13	-			

STORAGE												
port	1	3	5	7	9	11	13	15	17	19	21	23
Node	MGR 1	MGR 3	COM 2	COM 4	COM 6	COM 8	COM 10	COM 12	COM 14			
Port	2	4	6	8	10	12	14	16	18	20	22	24
Node	MGR 2	COM 1	COM 3	COM 5	COM 7	COM 9	COM 11	COM 13	-			

[illegible]

도커/쿠버네티스 환경 구축

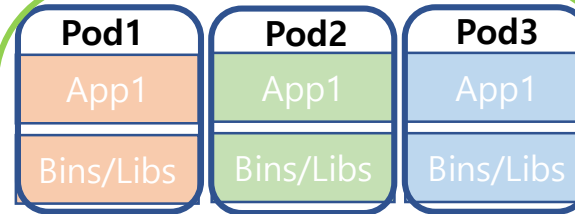


Kubernetes



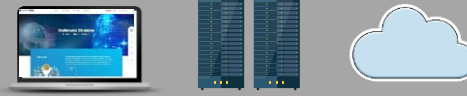
Cloud

Worker Node 1

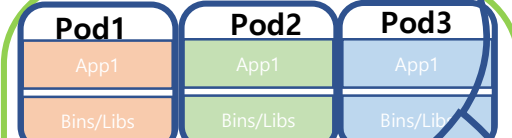


Operating System

Infrastructure



Worker Node N



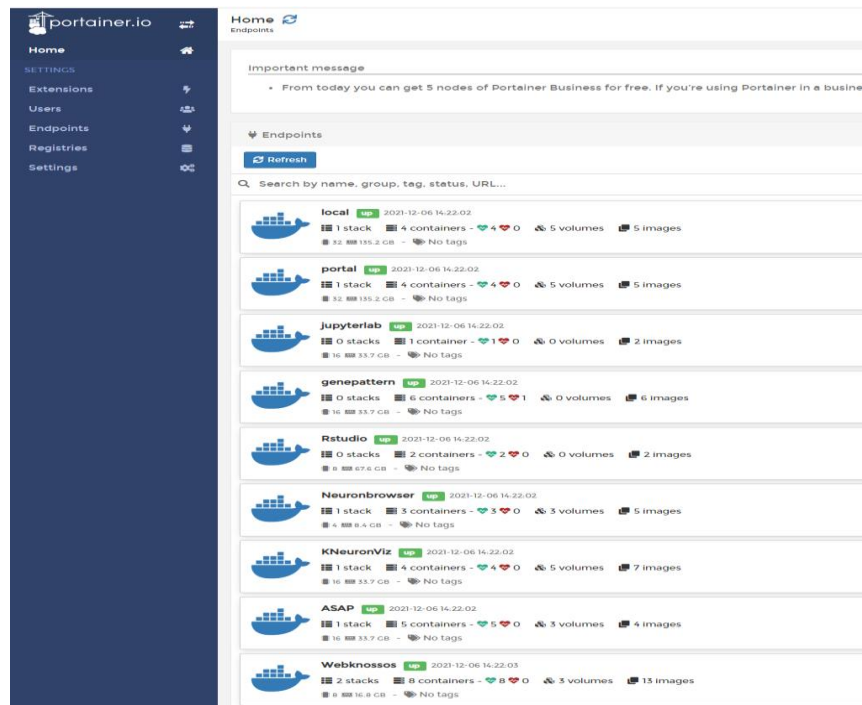
Operating System

Infrastructure

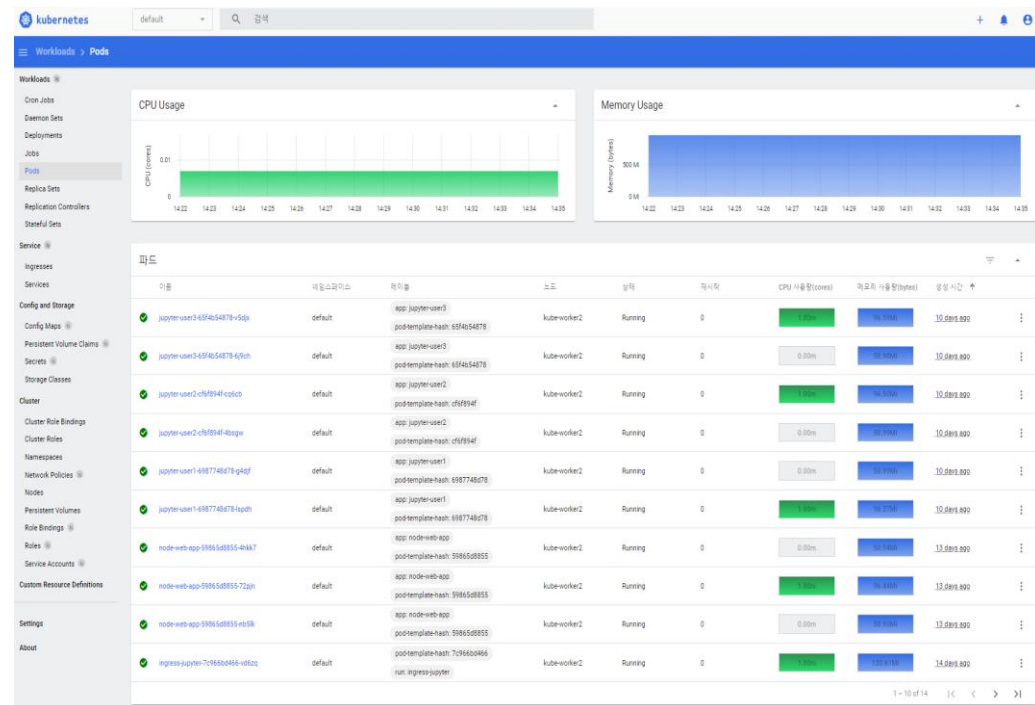


Docker

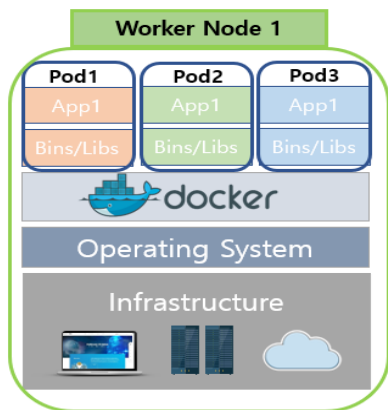
도커/쿠버네티스 환경 구축



Docker 이미지 기반
쿠버네티스 활성화
및 app 배포



각 Bio 데이터 분석
Application
도커 이미지 생성
및 관리



Docker



Kubernetes

App: KNeuroViz (도커 컨테이너 기반 뉴런 이미지 데이터 가시화 분석- appendix)

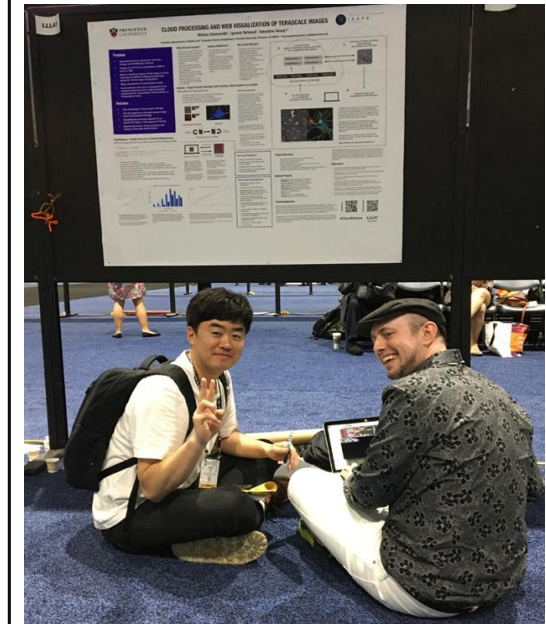
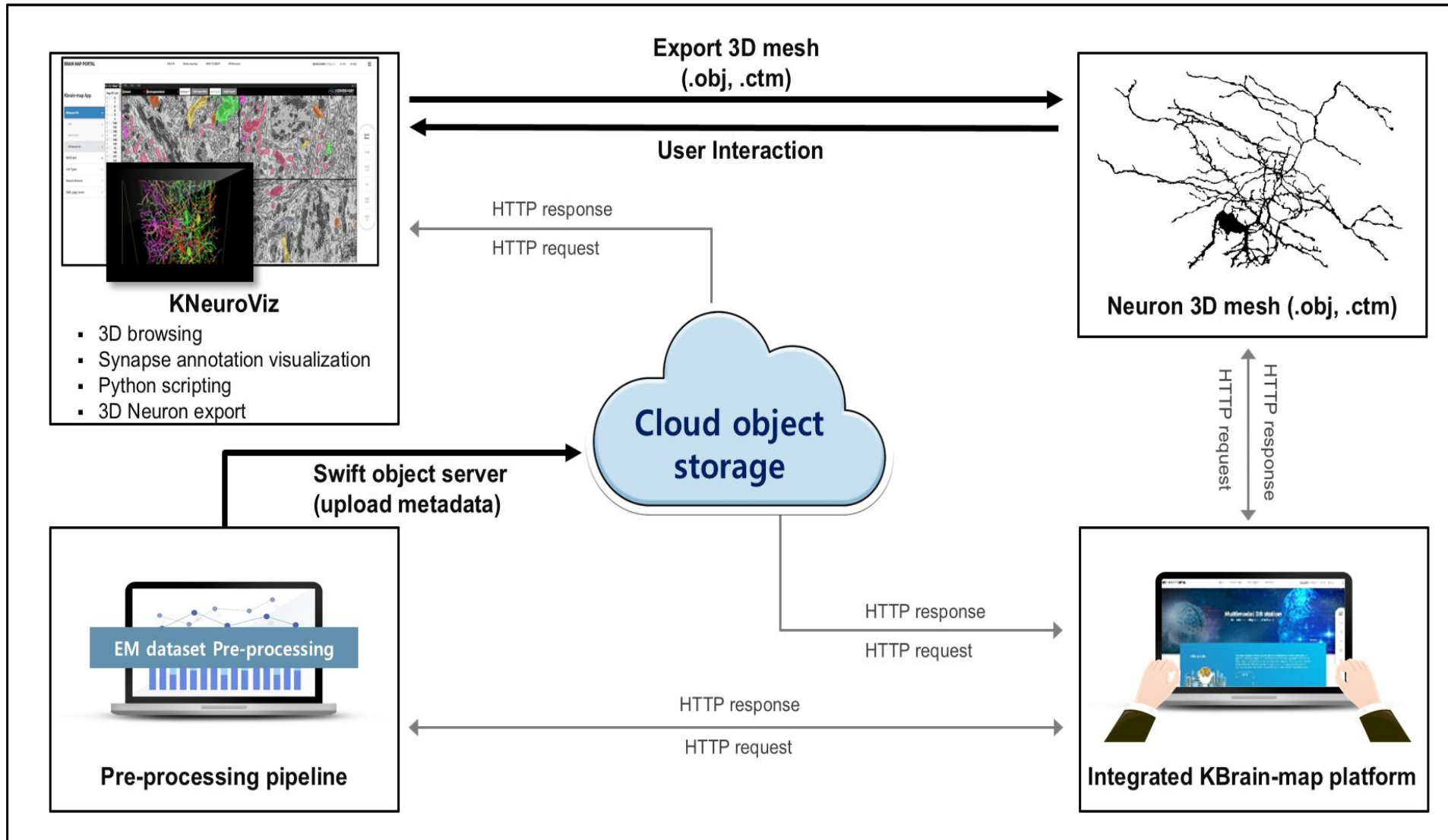
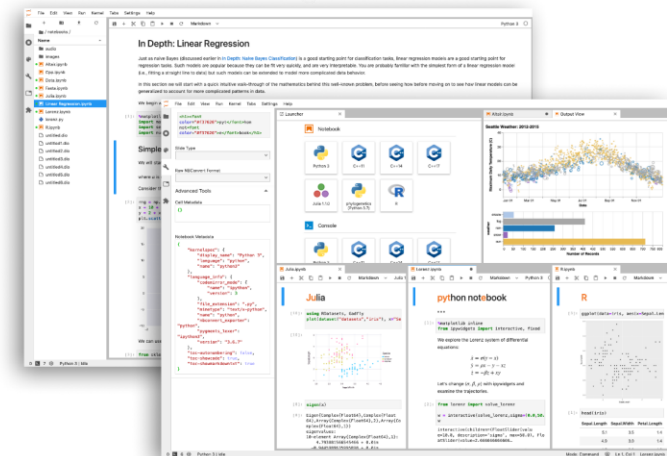
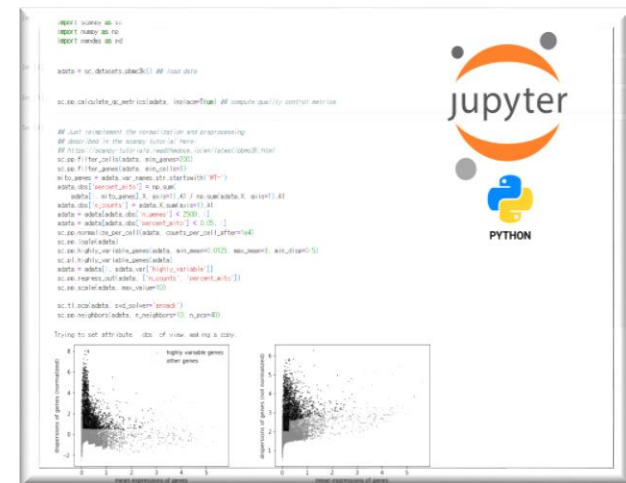
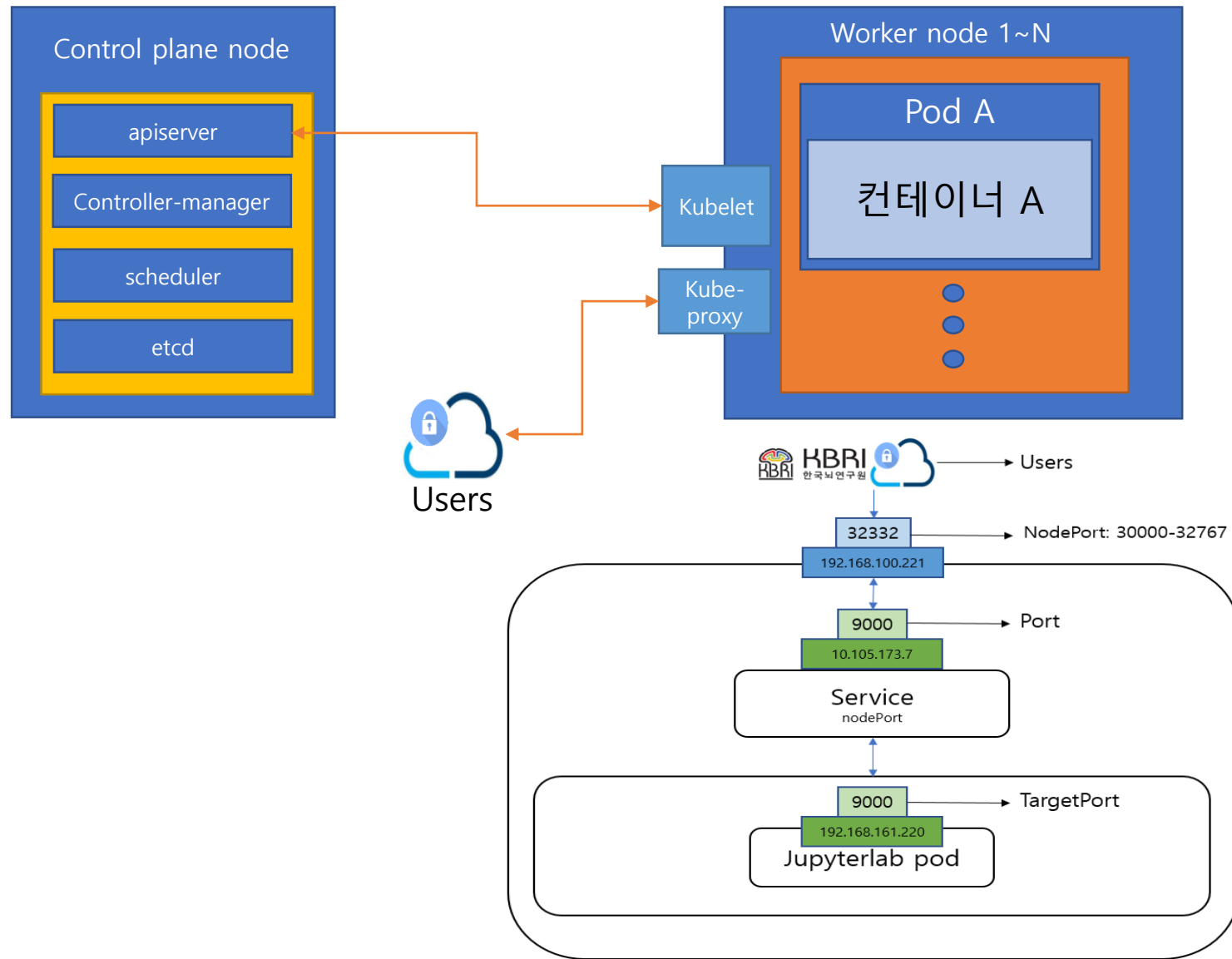


Figure 1:
Will Smith, KBrain-map
University of California, Berkeley

쿠버네티스 기반 연구자 교육용 주피터랩 컨테이너



기술적으로 차별화시키고자 했던 핵심요소

• 도커/쿠버네티스 기반 바이오 분석 tool 오케스트레이션 배포

- 뇌 영역의 바이오 데이터는 다양한 뇌 구조 scale 영역 측면의 연구가 이루어지기 때문에, 다양한 분석 S/W들이 존재합니다. 처음에는 S/W들이 늘어나면서 필요한 자원도 지속적으로 늘어나게 되고, 마찬가지로 서버 또한 증가하면서 관리하기에는 상당히 어려웠습니다. 이 문제를 해결하고자 도커 컨테이너를 사용하였고, 컨테이너 또한 효율적으로 관리하기 위해 컨테이너 오케스트레이션 툴인 쿠버네티스를 활용하여 S/W를 컨테이너화하여 좀 더 효율적으로 관리 및 배포할 수 있는 시스템을 구축하였습니다.

• (대용량)3D 뉴런 이미지 실시간 가시화 viewer

- 약 1.2TB 대용량 이미지 데이터를 웹에서 실시간 3D 가시화 및 분석 서비스하는 것이 핵심이었습니다. 컴퓨터비전 관련 기술동향을 조사하면서 3D 렌더링 오픈소스를 조사 및 KISTI 슈퍼컴퓨터 담당 연구원께 자문을 구하면서, 이를 빠르게 실시간 렌더링 및 웹 서비스를 생각하였지만 많은 문제점이 있었습니다. 꾸준히 기술동향 및 논문을 살피면서 프린스턴대학교의 세바스찬 랩에서 유사하게 연구하고 있는 것을 알게되었습니다. 그래서 미국 SFN 학회에 참석하여 승랩의 William Smith 연구원을 통해 복셀 기반 원본 3D 이미지를 작은 3D chunk(cubic)단위로 슬라이싱 및 웹 포털 브라우저 display에 보여지는 영역만 3D chunk 데이터의 ROI 좌표 기반 데이터 stream 방식의 I/O처리로 실시간 렌더링 가시화를 할 수 있었습니다.

• 바이오 분석 파이프라인

- 뇌 영역의 다양한 멀티스케일(Macro-Meso-Micro/Nano 단위) 기반의 처리 flow를 5단계로 Data ingest->Data storage->Data processing->Data Analysis->Visualize 구성하며, 이러한 방식을 이용하여 효율적인 데이터 처리 기반 인터페이스와 사용자에게 최적화 분석 서비스를 표준화 파이프라인을 구현하였습니다.

과제(프로젝트)를 통한 경험과 교훈

- 프로젝트를 통해 클라우드를 운영하면서 클라우드의 네이티브와 도커/쿠버네티스의 오케스테이션 기술을 활용한 적용 방법과 대용량 2D/3D 이미지 및 텍스트 데이터 처리를 위한 개발 방법론 등 다수의 앱을 별도 분리하여 MSA(Micro service architecture) 서비스 방식에 대해서 많은 지식과 경험을 가졌습니다.
- 그리고, 국내에서 아직 알려지지 않아서 조사하는데 시간과 여러 가지 기술을 파악하는데 있어서 많은 어려움이 있었지만 완수를 다 하고자 하는 프로젝트의 갈망과 열의를 다하며 여러모로 팀원(with 용역 업체)과 함께 화기애애한 분위기 속에서 다양한 의견 제시와 협동하여 많은 어려움도 금세 극복할 수 있었습니다.
- 마지막으로 사용자에게 신뢰성있는 프로그램을 배포하기 위해선 개발자의 진심과 혼이 담겨져 있어야 한다고 생각했습니다. 그 만큼 많은 노력과 열정을 다해 만들었기에 신뢰성과 사용자의 요구사항에 만족이 뒤따른다고 생각합니다.

Appendix

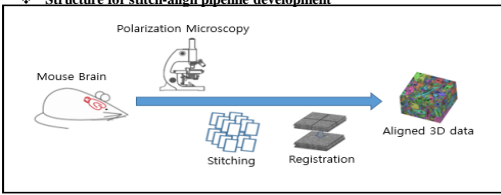
Abstract

Various types of image are exponentially increasing by development of optics and optical technologies in neuroscience field and these data are characterized by big-data, which requires high-throughput and high-speed processing technologies. High performance computing and resources are essential to convert the series of high-resolution 2D images to 3D. However, this process has limitation because it is very dependent on the software installed in the image instruments and their location. In this study, we propose a web tool to convert 2D sliced images into 3D data on the web which enables the users to analyze their data regardless of the location where they are and the software what they have. This tool increases the optimizing efficiency of file classification and sorting by the enhanced stitching and provides the flexibility of image combination through Fourier transform algorithm. The combined images are matched with the z-axis using rigid transformation method. The benefit of this system is to reduce the availability of resources and the risk of data missing. This tool will be installed on the web DB of Neural Circuit Research Group and open to users in the web service.

Keyword(s) : Big data, Fourier transform, stitch, align

Introduction

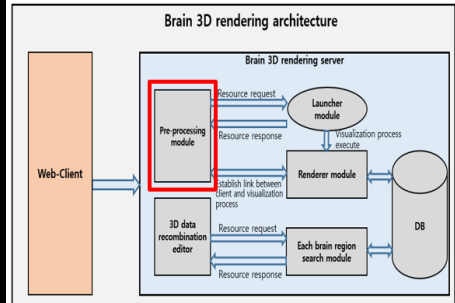
❖ **Structure for stitch-align pipeline development**



The big data are being generated by new optical equipment technologies in neuroscience such as monitoring tools for neural activity and imaging tools for circuit formation. Polarization microscopy such as confocal and light sheet microscope are widely used for neural network and generates high-resolution images. It is still necessary to develop an efficient 3D creation because it is nonexchangeable imaging file format in huge volume affecting data analysis. In this study, we propose the web-based automated stitch-align software to convert 2D sliced images into 3D data.

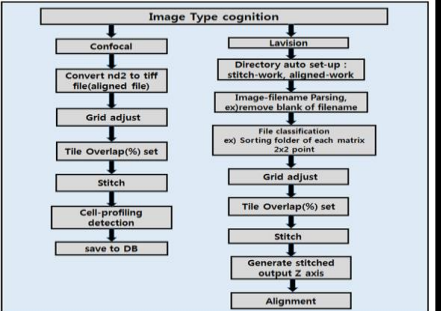
Method

❖ **Infrastructure system for Web-based Brain Bio-Data**



The goal of this system is building web-based interactive DB platform for understanding and analyzing of neural circuit

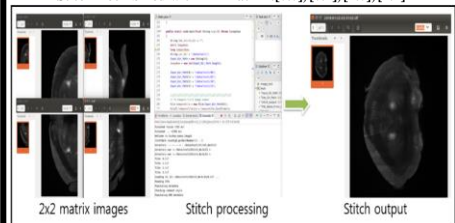
❖ **Preprocessing module : Automated stitch-align pipeline**



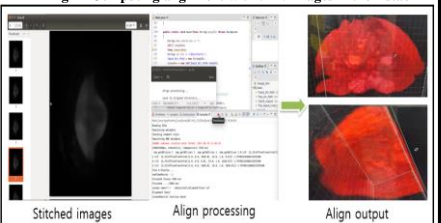
2D images derived from polarization microscopy are input and pre-processing module is performed for each type of inputted images.

Results

❖ **Stitch – combined with 2x2 matrix :[0,0],[0,1],[1,0],[1,1]**



❖ **Align – Computing alignment with 5440 images in the z-stack**

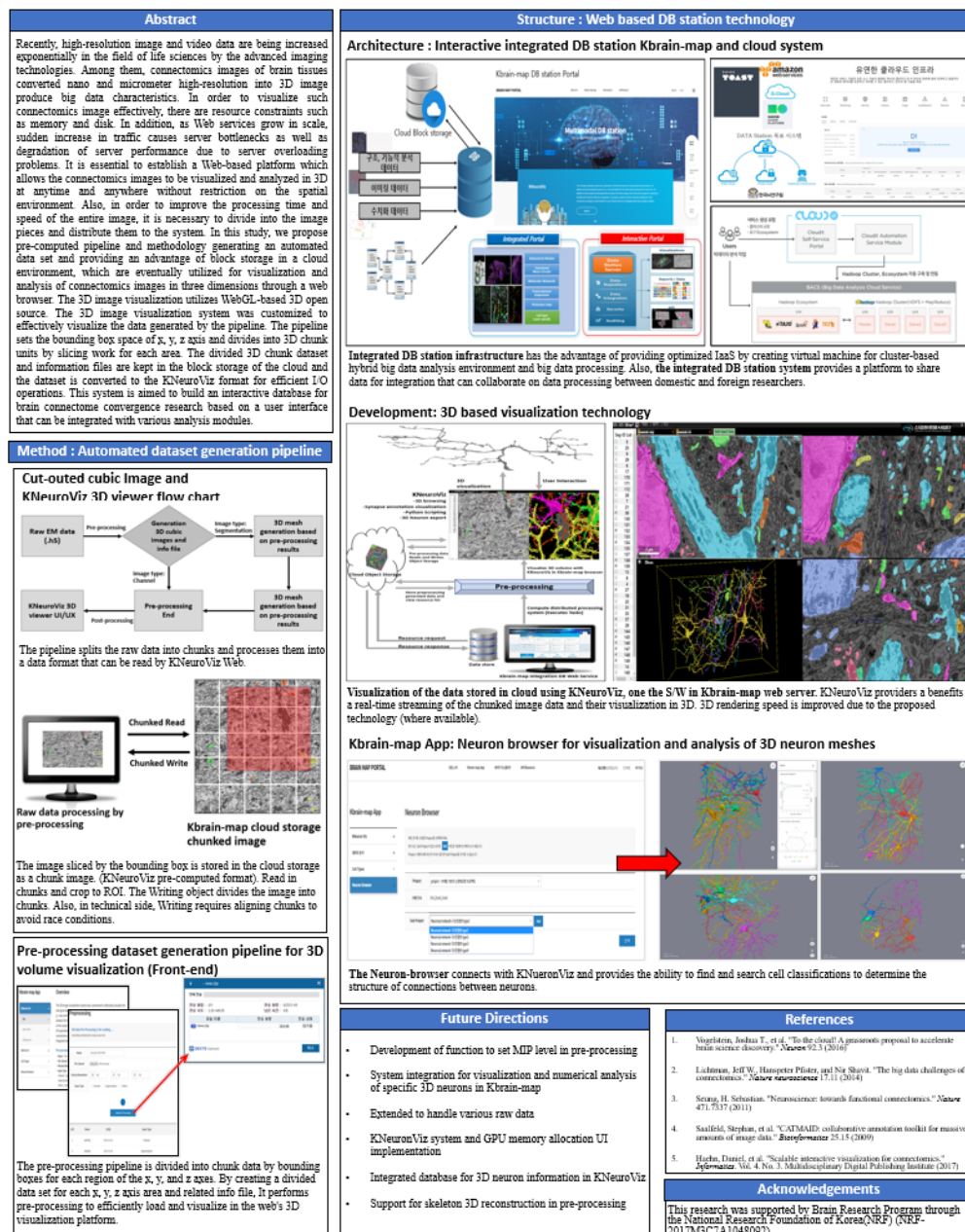


Conclusion

- In proposed method, this pipeline optimized high-performance computing resource, required to process big-data.
- To perform stitch and alignment, we proposed an efficient alignment and stitching method in terms of wide area and speed.
- This strategy showed that 2D slice images were efficiently converted to 3D data.

References

- [1] J.L. Mather leghal., et al. Image stitching and 2D to 3D Image Reconstruction for Abnormal Activity Detection. International journal of computer applications. vol 133, no17, 2016.
- [2] Ehsam Adel., et al. Image Stitching based on Feature Extraction Techniques: A Survey. Vol 99, no 6, 2014.
- [3] Scale-invariant feature transform (https://en.wikipedia.org/wiki/Scale-invariant_feature_transform)
- [4] Fourier transform (https://en.wikipedia.org/wiki/Fourier_transform)
- [5] Rigid transformation (https://en.wikipedia.org/wiki/Rigid_transformation)

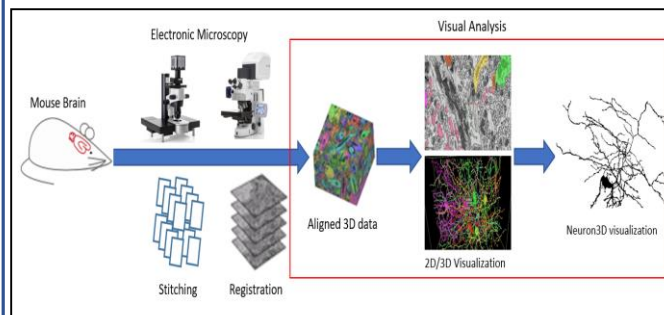


Abstract

In the field of connectome research, there is an ongoing need for analytical techniques to process the massive data obtained using high-resolution microscope imaging technology. In terms of computer science, the development of original technology for technological advancement, simulation of research results, and scalability for additional supplementation and improvement requires the development of third party and open source-based tools. In this study, we propose pre-computed pipeline and methodology generating an automated data set and providing an advantage of block storage in a cloud environment, which are eventually utilized the images produced by electronic microscopy (EM) to be visualized and analyzed in three dimensions through a web browser, KBrain-map platform. We implemented the open sources and computer vision libraries in this pipeline to detect neurons, synaptic connectivity, and neural structure in terabyte-scale EM data. This platform includes an automated pre-processing pipeline for EM images with high-capacity storage space. In addition, we developed the KNeuroViz, an analytical solution for post-processing, for web-based 3D visualization and analysis of neurons. This solution is a modified software of Neuroglancer and optimized to the web-based system. This will be implanted into KBrain-map platform, eventually. In current study, we propose the KBrain-map platform which is a cloud-based platform and includes the pipeline to visualize neurons and synapses in 3D and analyze their connectivity, efficiently. This system will be continuously integrated with various analysis modules providing an interactive platform for brain research.

Introduction

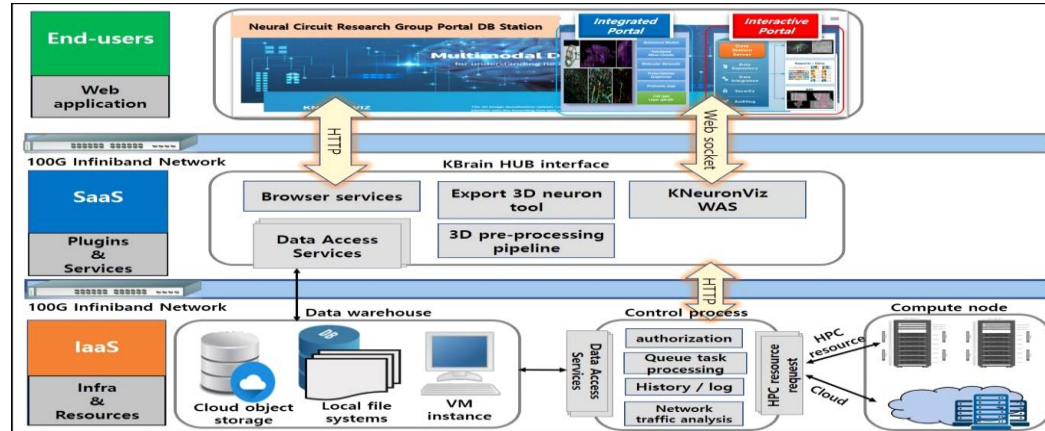
EM(Electron microscope) image analysis process and visual analysis flow chart



The big data are being generated by new optical equipment technologies in neuroscience such as monitoring tools for neural activity and imaging tools for circuit formation. Electron microscope is widely used for neural network and generates high-resolution images. It is still necessary to develop an efficient 3D creation because it is nonexchangeable imaging file format in huge volume affecting data analysis. In this study, we propose the integrated KBrain-map platform as a SaaS that can be used for efficient processing, storage, management, interactive visualization, and analysis.

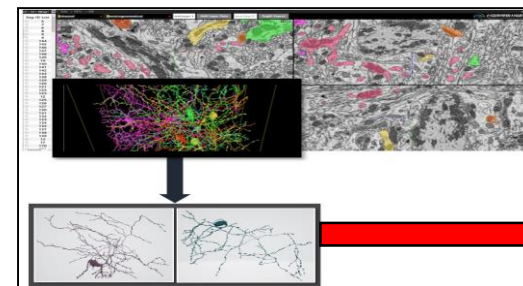
Web based Integrated DB station system

Architecture : Interactive integrated DB station KBrain-map and cloud system



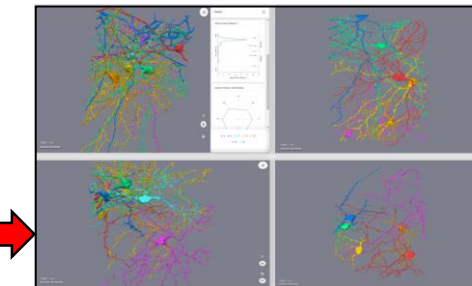
The integrated DB station, which provides cloud-based IaaS and SaaS models, is composed of 3 major layers: (1) The end-users (Web application) level, which provides access to KBrain-map, (2) the SaaS (plugins & services) level, and (3) the IaaS (infrastructure as a service, Infra & Resource) level.

KBrain-map App: KNeuroViz for visualization of 3D neuron meshes



KNeuroViz shows neuron viewer for 3D analysis and visualization of neural circuit networks. In the bottom 3D mesh display of figure, the 3D image of each 2 neurons is the result of visualization in KNeuroViz from an exported neuron image file (.obj or .ctm) to a 3D neuron mesh.

KBrain-map App: Neuron browser for analysis of 3D neuron



The Neuron Browser connects to KNeuroViz and provides the ability to find cell classifications, perform statistical analysis, and determine the structure of connections between neurons.

References

1. Vogelstein, Joshua T., et al. "To the cloud! A grassroots proposal to accelerate brain science discovery." *Neuron* 92.3 (2016)
2. Lichtman, Jeff W., Hanspeter Pfister, and Nir Shavit. "The big data challenges of connectomics." *Nature neuroscience* 17.11 (2014)
3. Seung, H. Sebastian. "Neuroscience: towards functional connectomics." *Nature* 471.7337 (2011)
4. Saalfeld, Stephan, et al. "CATMAID: collaborative annotation toolkit for massive amounts of image data." *Bioinformatics* 25.15 (2009)
5. Haehn, Daniel, et al. "Scalable interactive visualization for connectomics." *Informatics*, Vol. 4, No. 3. Multidisciplinary Digital Publishing Institute (2017)

Acknowledgements

This research was supported by KBRI basic research programs (20-BR-02-03 & 20-BR-04) and National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (NRF-2015M3C7A1029037 & 2017M3C7A1048092).

Github URL

- <https://github.com/Virusuki>
-> enjoy devops!