수행 프로젝트:
바이오 기반 대용량 데이터 분석을 위한 통합 분석 시스템 구축 (통합 분석 DB station 포털)

2021/12

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 S/W 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 인터액티브 데이터 플랫폼을 구축하는 것이다.

최종목표: 의사결정에 중요한 전전두엽 신경회로망 규명을 위해 멀티모달 정보의 저장, 추출, 분석이 가능한 웹 기반 인터액티브 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

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

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

o 목표 2. 웹 기반 인터액티브 플랫폼 및 분석 파이프라인 구축/서비스

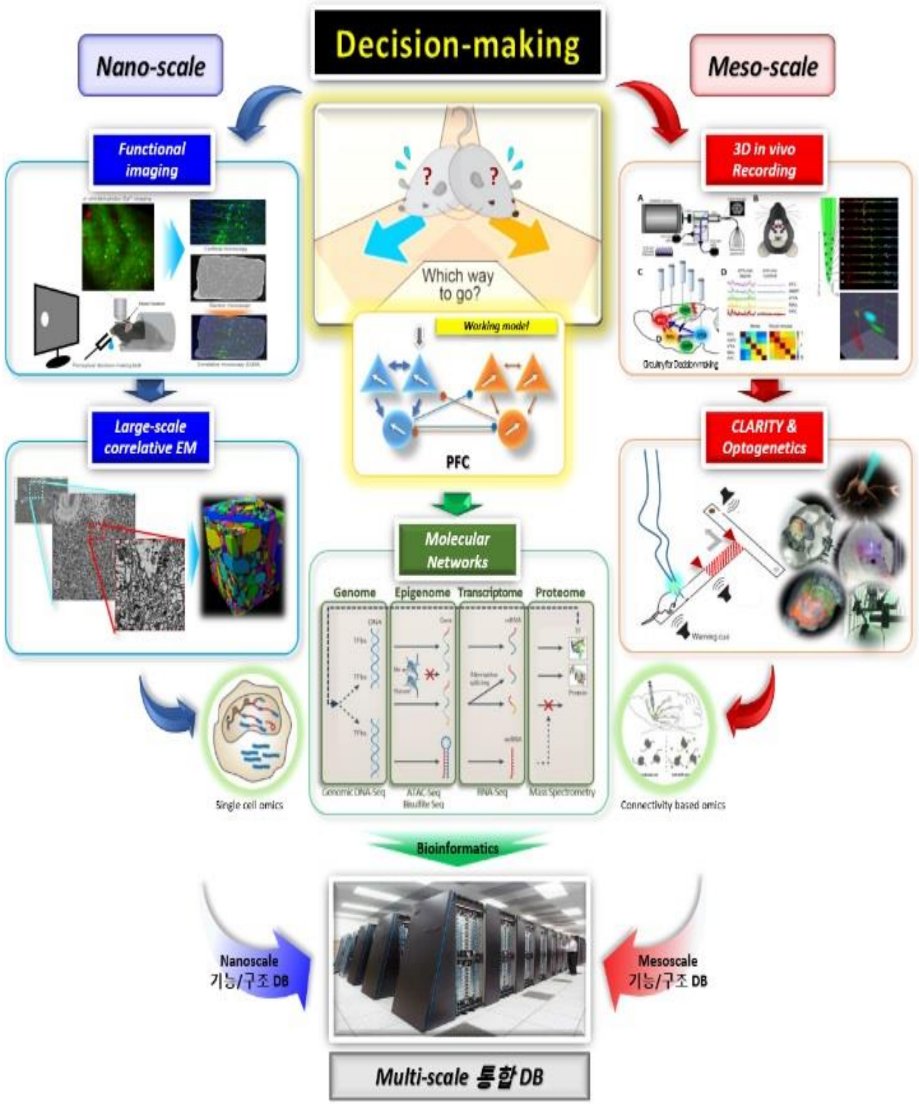
- 바이오 분석을 위한 jupyterlab(연구자 교육용 포함), genepattern과 galaxy 기반 파이프라인 구축

o 목표 3. 국가 뇌 연구 네트워크 구축 및 운영

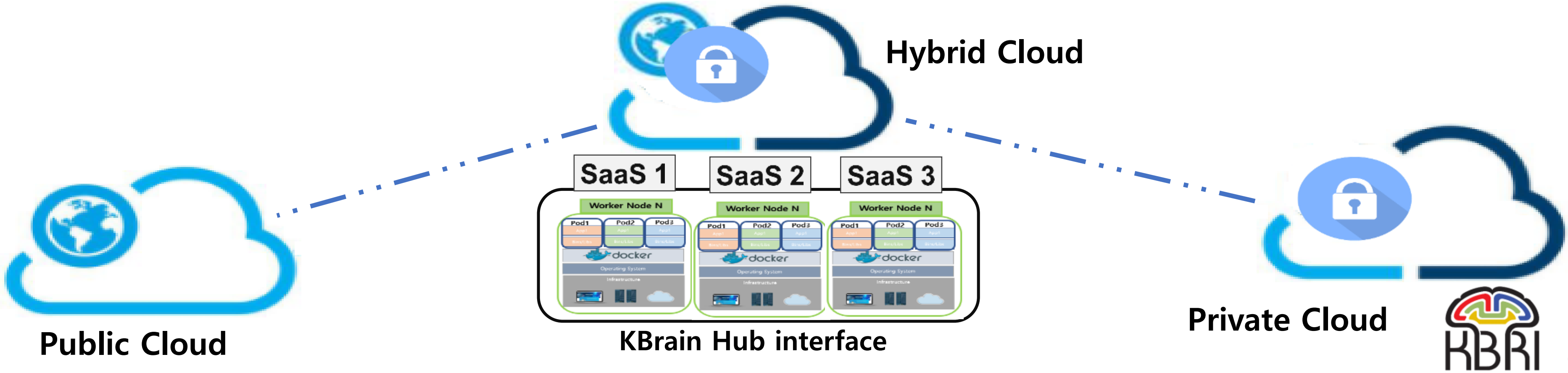
- 국제 뇌연구 기관 협력을 위한 국내 인프라 활용, 뇌과학 데이터 공유를 지원하는 인프라 구축

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

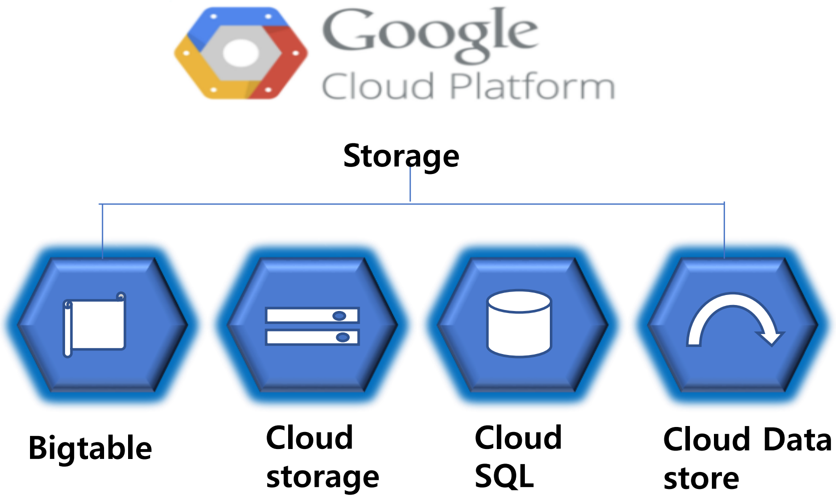
- 웹 기반 멀티스케일 이미징 분석 고속화/가시화/자동화기술 구현

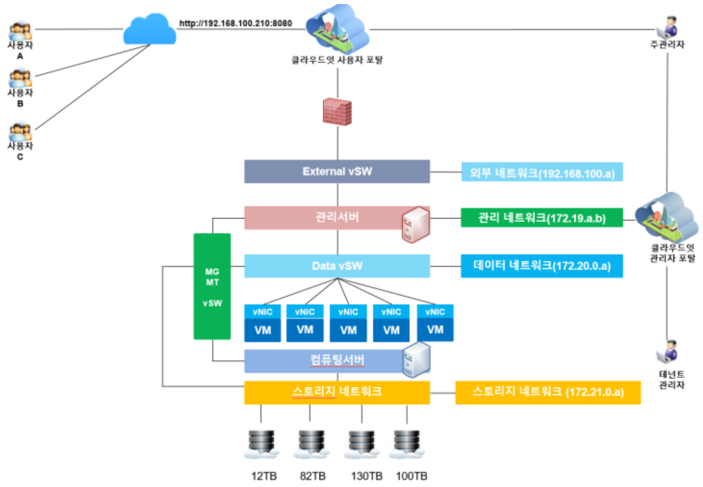


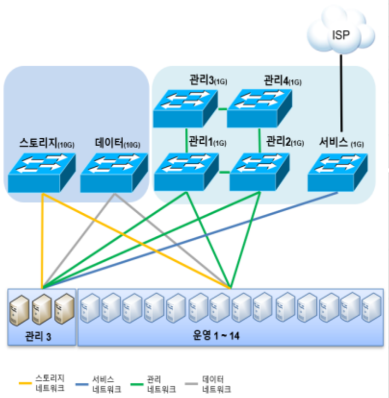
하이브리드 클라우드 환경
Pysical diagram



KBRI 한국뇌연구원







범 례			
☁	인터넷		
	1G 스위치		
	10G 스위치		
	관리 서버		
	운영 서버		

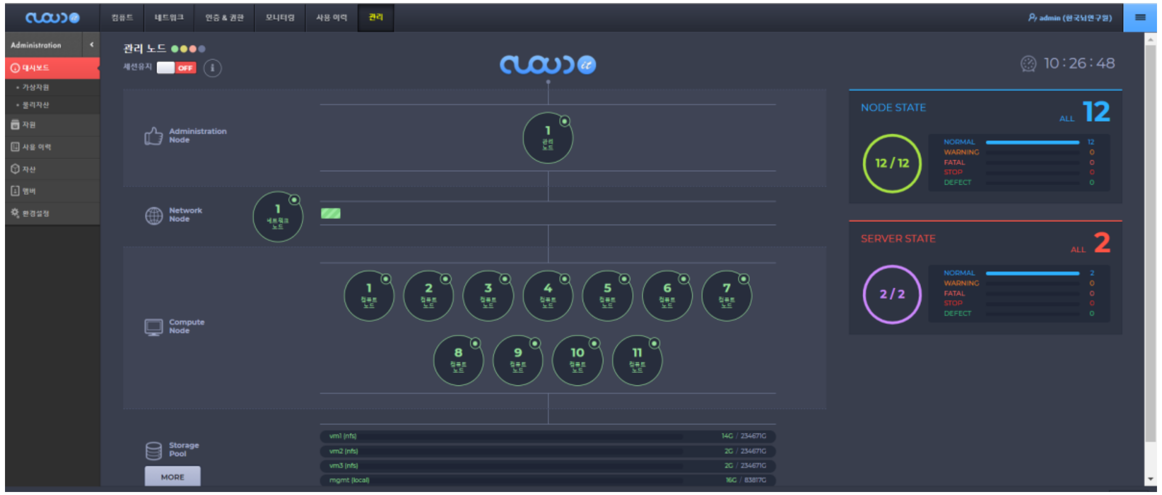
구성 장비 소계

구 분	수량	비고
1G 스위치	5 대	서비스/ 관리1/관리2/관리3/관리4
10G 스위치	2 대	데이터/스토리지
관리서버	3 대	관리 서버 3대 (12TB MGMT Pool 2대) (82TB MGMT Pool 1대)
운영 서버	14 대	운영 노드 14대 (130TB VM Pool 7대) (100TB VM Pool 7대)

Logical Structure

Pysical Structure

클라우드 플랫폼 (이노그리드사의 프라이빗 클라우드 도입)



NODE STATE ALL 12

12 / 12

NORMAL	12
WARNING	0
FATAL	0
STOP	0
DEFECT	0

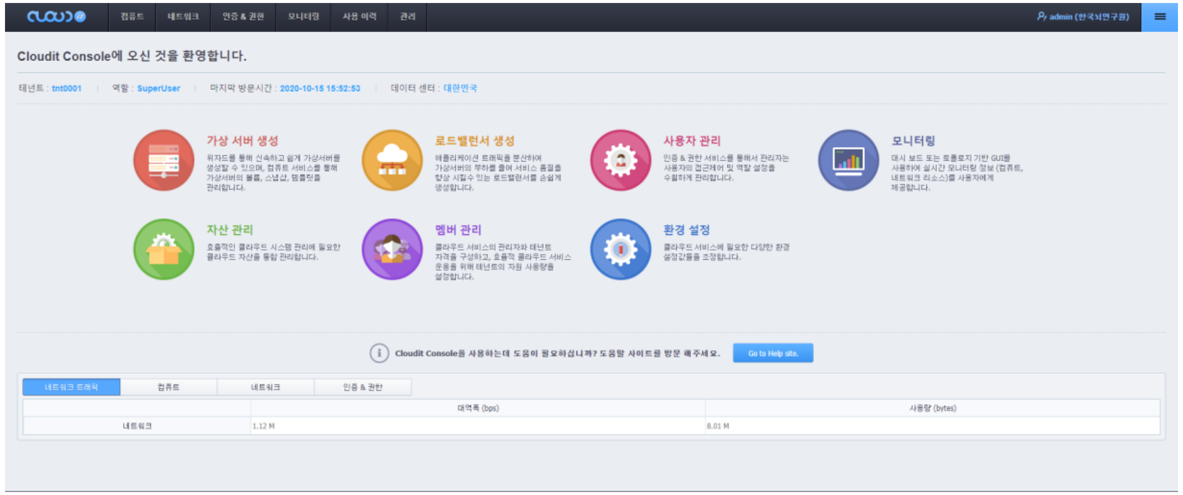
SERVER STATE ALL 2

2 / 2

NORMAL	2
WARNING	0
FATAL	0
STOP	0
DEFECT	0

vm1 (nfs)	14G / 234671G
vm2 (nfs)	2G / 234671G
vm3 (nfs)	2G / 234671G
mgmt (local)	16G / 83817G

서버 자원 상태



Cloudit Console에 오신 것을 환영합니다.

테넌트 : tnt0001 | 역할 : SuperUser | 마지막 방문시간 : 2020-10-15 15:52:53 | 데이터 센터 : 대한민국

가상 서버 생성
위자드를 통해 신속하고 쉽게 가상서버를 생성할 수 있으며, 컴퓨트 서비스를 통해 가상서버의 볼륨, 스냅샷, 템플릿을 관리합니다.

로드밸런서 생성
애플리케이션 트래픽을 분산하여 가상서버의 부하를 줄여 서비스 품질을 향상 시킬수 있는 로드밸런서를 손쉽게 생성합니다.

사용자 관리
인증 & 권한 서비스를 통해서 관리자는 사용자의 접근제어 및 역할 설정을 수월하게 관리합니다.

모니터링
대시 보드 또는 토폴로지 기반 GUI를 사용하여 실시간 모니터링 정보 (컴퓨트, 네트워크 리소스)를 사용자에게 제공합니다.

자산 관리
효율적인 클라우드 시스템 관리에 필요한 클라우드 자산을 통합 관리합니다.

멤버 관리
클라우드 서비스의 관리자와 테넌트 자격을 구성하고, 효율적 클라우드 서비스 운용을 위해 테넌트의 자원 사용량을 설정합니다.

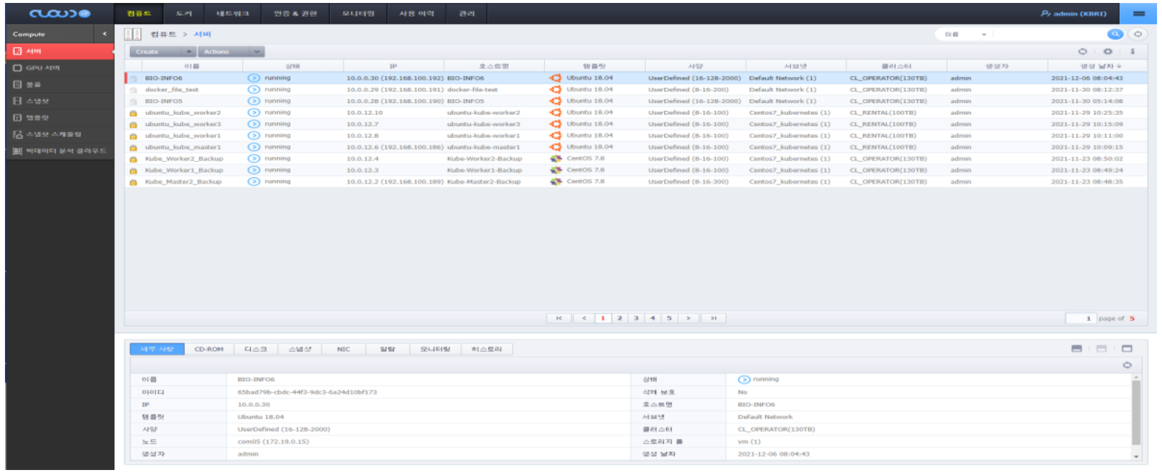
환경 설정
클라우드 서비스에 필요한 다양한 환경 설정값들을 조정합니다.

Cloudit Console을 사용하는데 도움이 필요하십니까? 도움말 사이트를 방문 해 주세요. Go to Help site.

네트워크 트래픽 | 컴퓨트 | 네트워크 | 인증 & 권한

	대역폭 (bps)	사용량 (bytes)
네트워크	1.12 M	8.01 M

가상화 자원 카테고리



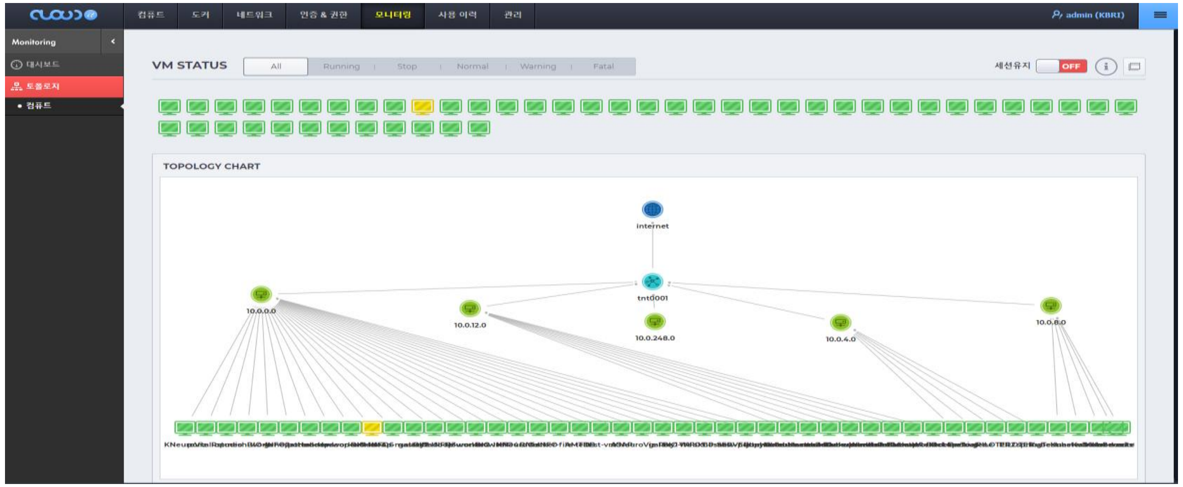
이름	상태	IP	호스트명	템플릿	사양	서브넷	클러스터	생성자	생성 날짜
BIO-INFO6	running	10.0.0.30 (192.168.100.192)	BIO-INFO6	Ubuntu 18.04	UserDefined (16-128-2000)	Default Network (1)	CL_OPERATOR(130TB)	admin	2021-12-06 08:04:43
docker_file_test	running	10.0.0.29 (192.168.100.191)	docker-file-test	Ubuntu 18.04	UserDefined (8-16-200)	Default Network (1)	CL_OPERATOR(130TB)	admin	2021-11-30 08:12:37
BIO-INFO5	running	10.0.0.28 (192.168.100.190)	BIO-INFO5	Ubuntu 18.04	UserDefined (16-128-2000)	Default Network (1)	CL_OPERATOR(130TB)	admin	2021-11-30 05:14:08
ubuntu_kube_worker2	running	10.0.12.10	ubuntu-kube-worker2	Ubuntu 18.04	UserDefined (8-16-100)	Centos7_kubernetes (1)	CL_RENTAL(100TB)	admin	2021-11-29 10:25:35
ubuntu_kube_worker3	running	10.0.12.7	ubuntu-kube-worker3	Ubuntu 18.04	UserDefined (8-16-100)	Centos7_kubernetes (1)	CL_RENTAL(100TB)	admin	2021-11-29 10:15:09
ubuntu_kube_worker1	running	10.0.12.8	ubuntu-kube-worker1	Ubuntu 18.04	UserDefined (8-16-100)	Centos7_kubernetes (1)	CL_RENTAL(100TB)	admin	2021-11-29 10:11:00
ubuntu_kube_master1	running	10.0.12.6 (192.168.100.186)	ubuntu-kube-master1	Ubuntu 18.04	UserDefined (8-16-100)	Centos7_kubernetes (1)	CL_RENTAL(100TB)	admin	2021-11-29 10:09:15
Kube_Worker2_Backup	running	10.0.12.4	Kube-Worker2-Backup	CentOS 7.8	UserDefined (8-16-100)	Centos7_kubernetes (1)	CL_OPERATOR(130TB)	admin	2021-11-23 08:50:02
Kube_Worker1_Backup	running	10.0.12.3	Kube-Worker1-Backup	CentOS 7.8	UserDefined (8-16-100)	Centos7_kubernetes (1)	CL_OPERATOR(130TB)	admin	2021-11-23 08:49:24
Kube_Master2_Backup	running	10.0.12.2 (192.168.100.189)	Kube-Master2-Backup	CentOS 7.8	UserDefined (8-16-300)	Centos7_kubernetes (1)	CL_OPERATOR(130TB)	admin	2021-11-23 08:48:35

1 2 3 4 5 — 1 page of 5

세부 사항 | CD-ROM | 디스크 | 스냅샷 | NIC | 알람 | 모니터링 | 히스토리

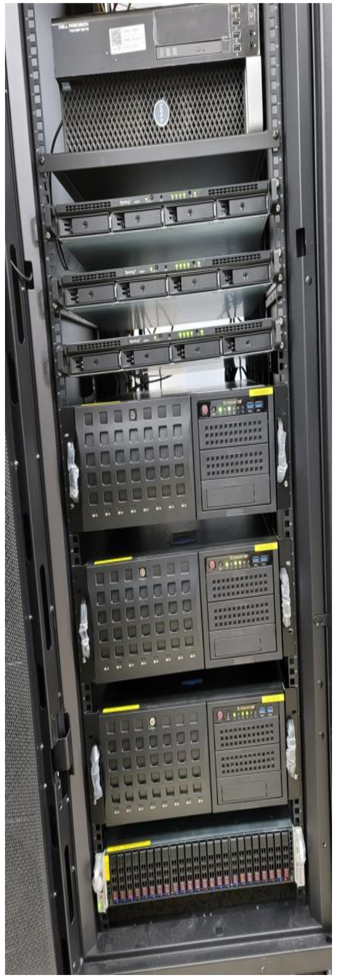
이름	BIO-INFO6	상태	running
아이디	65bad79b-cbdc-44f3-9dc3-6a24d10bf173	삭제 보호	No
IP	10.0.0.30	호스트명	BIO-INFO6
템플릿	Ubuntu 18.04	서브넷	Default Network
사양	UserDefined (16-128-2000)	클러스터	CL_OPERATOR(130TB)
노드	com05 (172.19.0.15)	스토리지 풀	vm (1)
생성자	admin	생성 날짜	2021-12-06 08:04:43

VM 자원 제어

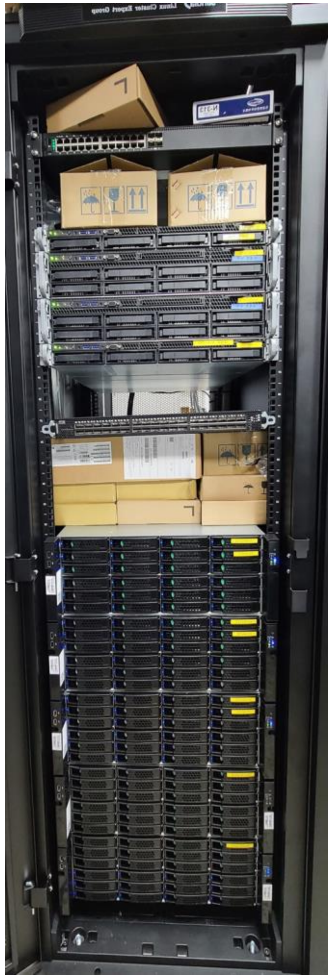


네트워크 자원 모니터링

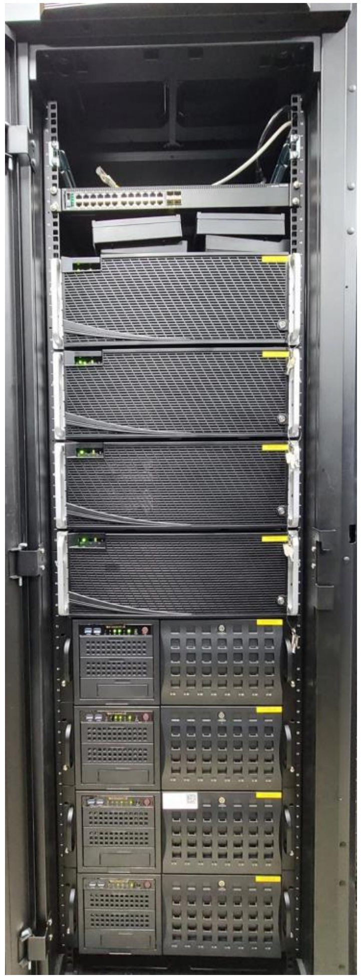
통합 분석 DB station 클라우드 인프라 실장도



<랙 전면(기존)>



<랙 전면(좌)>



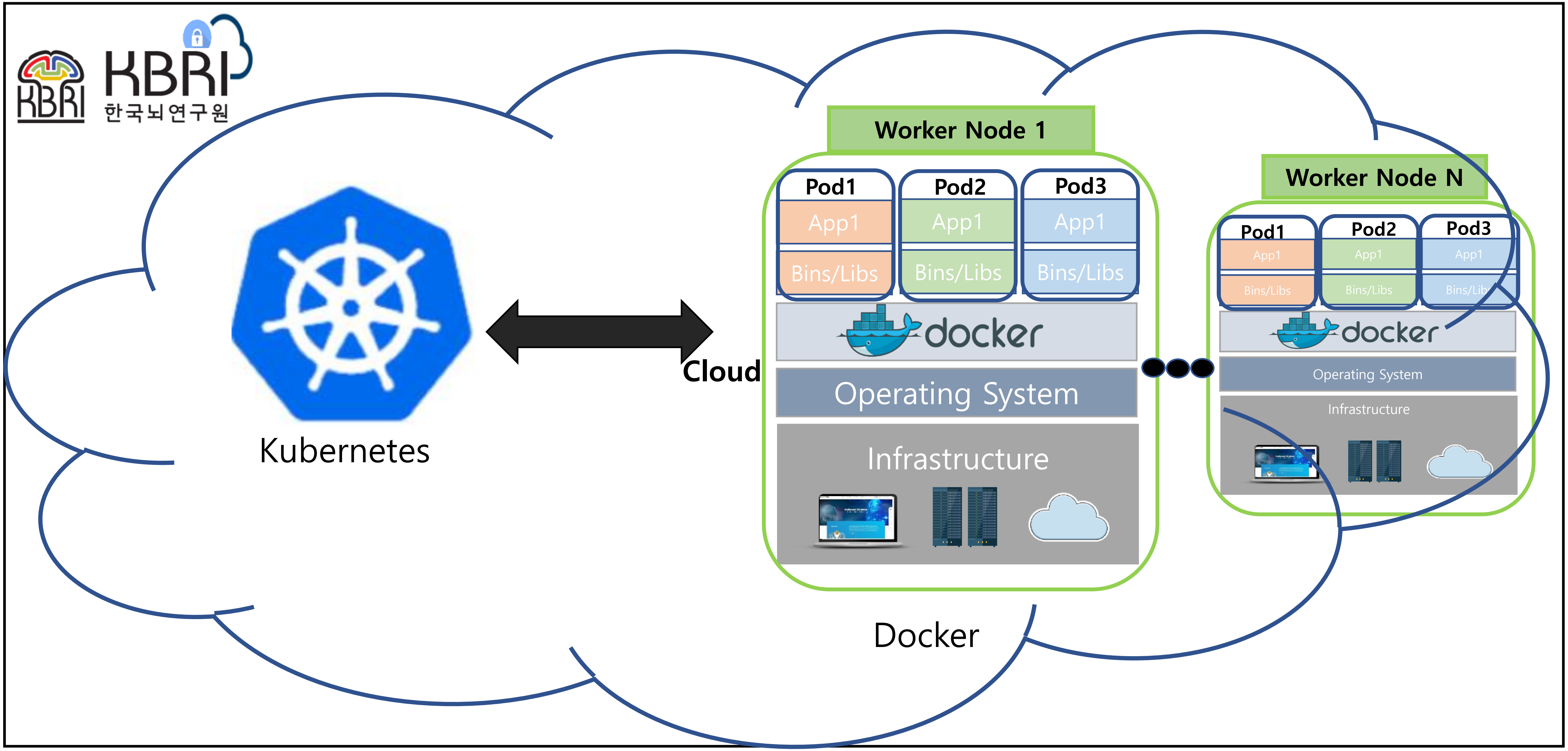
<랙 전면(우)>

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	-			

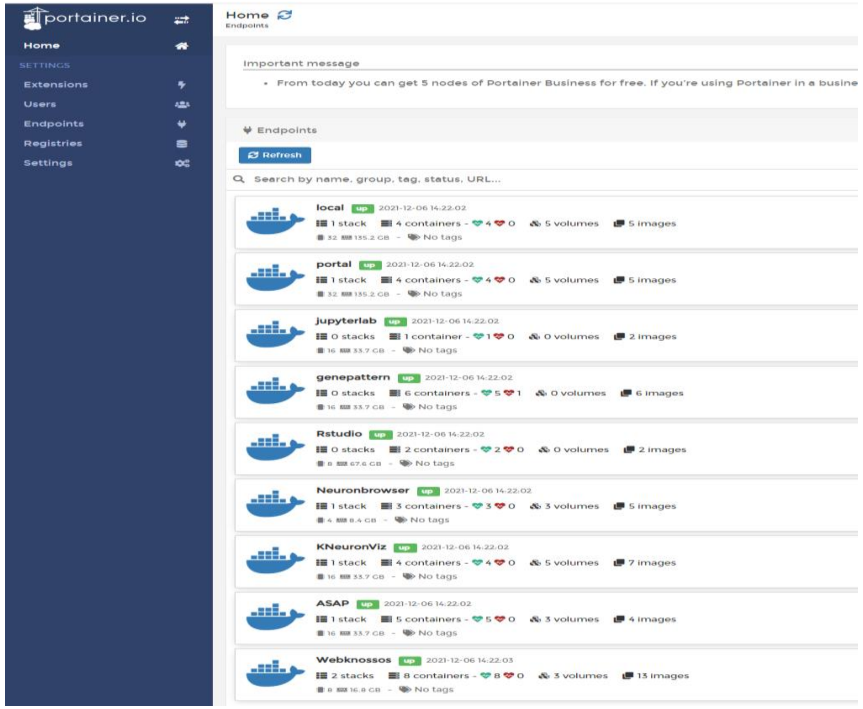
STORAGE												
port	1	3	5	7	9	11	13	15	17	19	21	23
Node	MGR 1	MGR 3	COM 2	COM 4	COM 6	COM 8	COM 10	COM 12	COM 14			
Port	2	4	6	8	10	12	14	16	18	20	22	24
Node	MGR 2	COM 1	COM 3	COM 5	COM 7	COM 9	COM 11	COM 13	-			

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

도커/쿠버네티스 환경 구축



도커/쿠버네티스 환경 구축



Home
Endpoints

Important message

- From today you can get 5 nodes of Portainer Business for free. If you're using Portainer in a busine

Endpoints

Search by name, group, tag, status, URL...

local up 2021-12-06 14:22:02
1 stack 4 containers - 4 0 5 volumes 5 images
32 135.2 GB - No tags

portal up 2021-12-06 14:22:02
1 stack 4 containers - 4 0 5 volumes 5 images
32 135.2 GB - No tags

jupyterlab up 2021-12-06 14:22:02
0 stacks 1 container - 1 0 0 volumes 2 images
16 33.7 GB - No tags

genepattern up 2021-12-06 14:22:02
0 stacks 6 containers - 5 1 0 volumes 6 images
16 33.7 GB - No tags

Rstudio up 2021-12-06 14:22:02
0 stacks 2 containers - 2 0 0 volumes 2 images
8 67.6 GB - No tags

Neuronbrowser up 2021-12-06 14:22:02
1 stack 3 containers - 3 0 3 volumes 5 images
4 8.4 GB - No tags

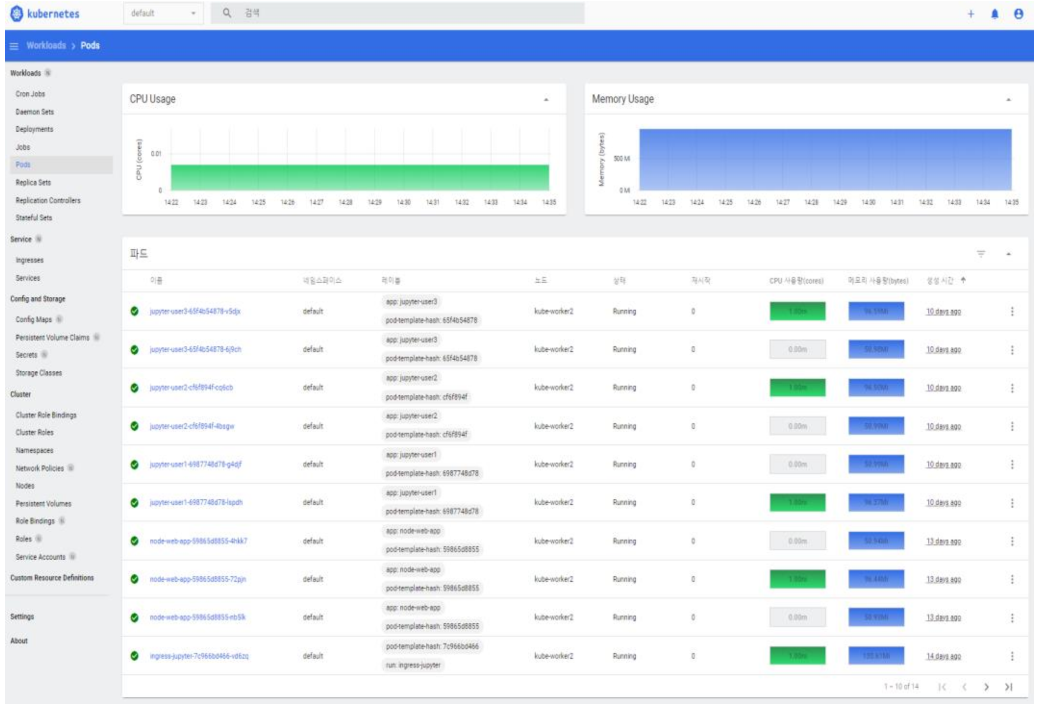
KNeuronViz up 2021-12-06 14:22:02
1 stack 4 containers - 4 0 5 volumes 7 images
16 33.7 GB - No tags

ASAP up 2021-12-06 14:22:02
1 stack 5 containers - 5 0 3 volumes 4 images
16 33.7 GB - No tags

Webknossos up 2021-12-06 14:22:03
2 stacks 8 containers - 8 0 3 volumes 13 images
8 16.8 GB - No tags

각 Bio 데이터 분석 Application 도커 이미지 생성 및 관리

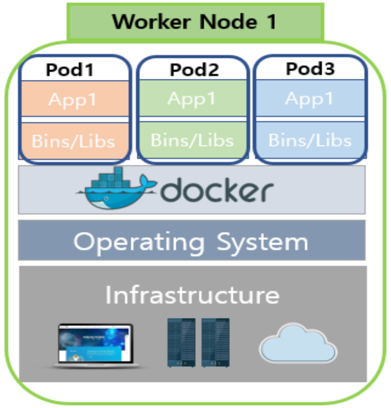
Docker 이미지 기반 쿠버네티스 활성화 및 app 배포



파드

이름	네임스페이스	레이블	노드	상태	재시작	CPU 사용량(cores)	메모리 사용량(bytes)	생성 시간
jupyter-user3-65f4b54878-v5djx	default	app: jupyter-user3 pod-template-hash: 65f4b54878	kube-worker2	Running	0	1.00m	[illegible]	10 days ago
jupyter-user3-65f4b54878-6j9ch	default	app: jupyter-user3 pod-template-hash: 65f4b54878	kube-worker2	Running	0	0.00m	[illegible]	10 days ago
jupyter-user2-cf6f894f-cq6cb	default	app: jupyter-user2 pod-template-hash: cf6f894f	kube-worker2	Running	0	1.00m	[illegible]	10 days ago
jupyter-user2-cf6f894f-4bsgw	default	app: jupyter-user2 pod-template-hash: cf6f894f	kube-worker2	Running	0	0.00m	[illegible]	10 days ago
jupyter-user1-6987748d78-g4djf	default	app: jupyter-user1 pod-template-hash: 6987748d78	kube-worker2	Running	0	0.00m	[illegible]	10 days ago
jupyter-user1-6987748d78-lspdh	default	app: jupyter-user1 pod-template-hash: 6987748d78	kube-worker2	Running	0	1.00m	[illegible]	10 days ago
node-web-app-59865d8855-4hkk7	default	app: node-web-app pod-template-hash: 59865d8855	kube-worker2	Running	0	0.00m	[illegible]	13 days ago
node-web-app-59865d8855-72pjn	default	app: node-web-app pod-template-hash: 59865d8855	kube-worker2	Running	0	1.00m	[illegible]	13 days ago
node-web-app-59865d8855-nb5lk	default	app: node-web-app pod-template-hash: 59865d8855	kube-worker2	Running	0	0.00m	[illegible]	13 days ago
ingress-jupyter-7c966bd466-vd6zq	default	pod-template-hash: 7c966bd466 run: ingress-jupyter	kube-worker2	Running	0	1.00m	[illegible]	14 days ago

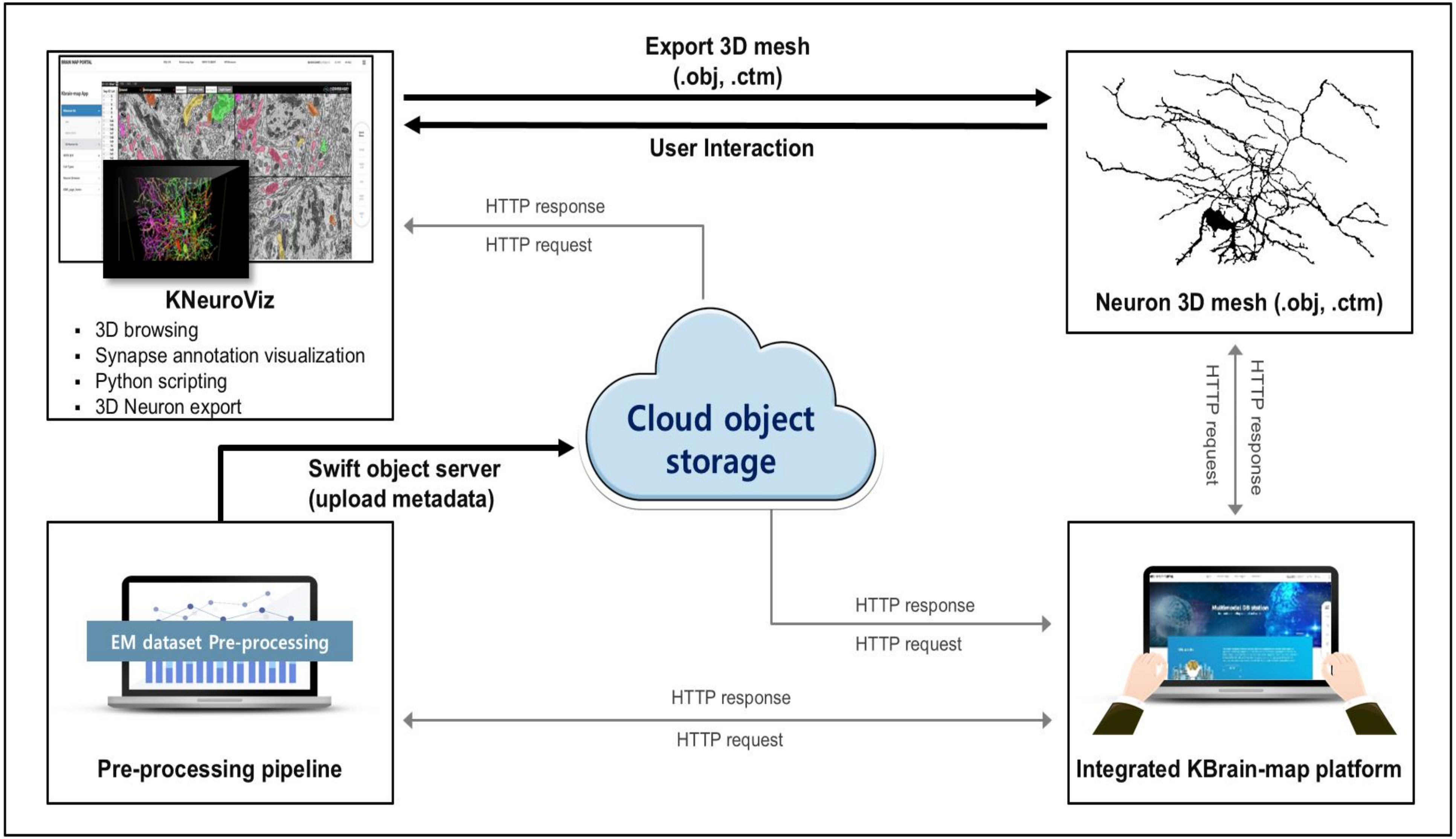
1 – 10 of 14



Docker

Kubernetes

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



Export 3D mesh (.obj, .ctm)

User Interaction

HTTP response

HTTP request

KNeuroViz

- 3D browsing
- Synapse annotation visualization
- Python scripting
- 3D Neuron export

Neuron 3D mesh (.obj, .ctm)

HTTP request

HTTP response

Cloud object storage

Swift object server (upload metadata)

EM dataset Pre-processing

Pre-processing pipeline

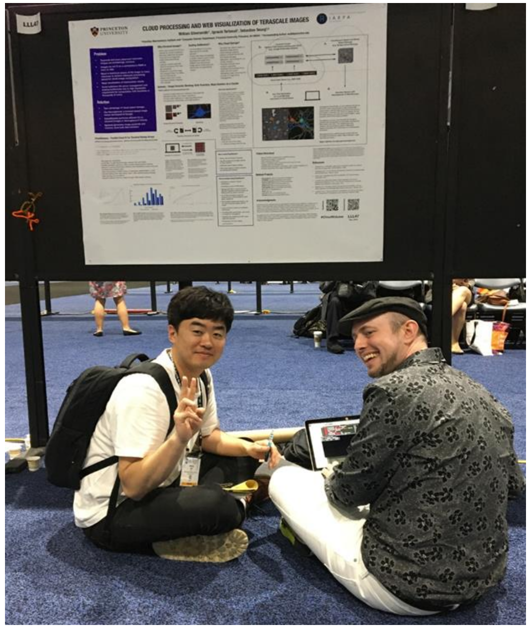
HTTP response

HTTP request

HTTP response

HTTP request

Integrated KBrain-map platform

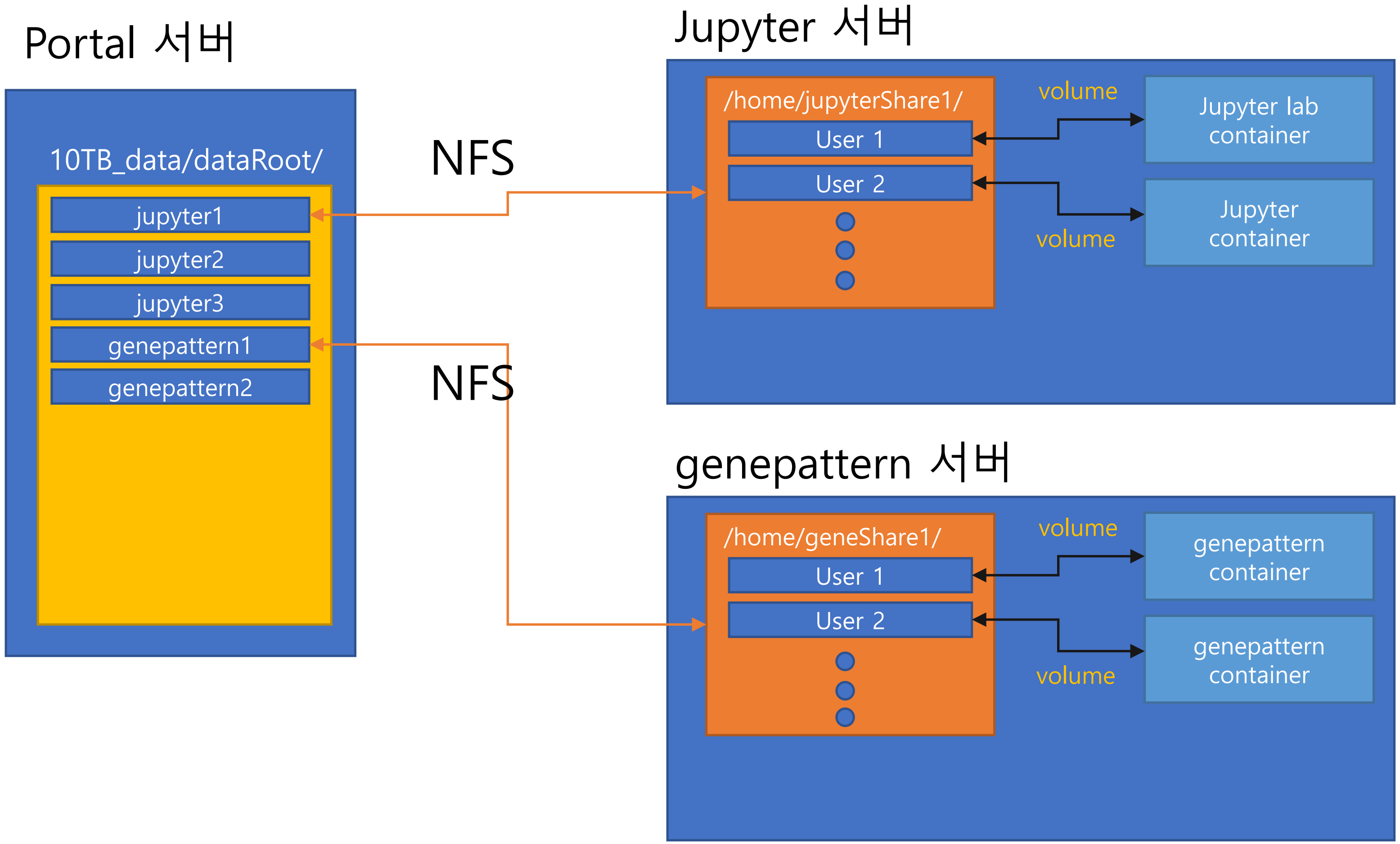


Helper:

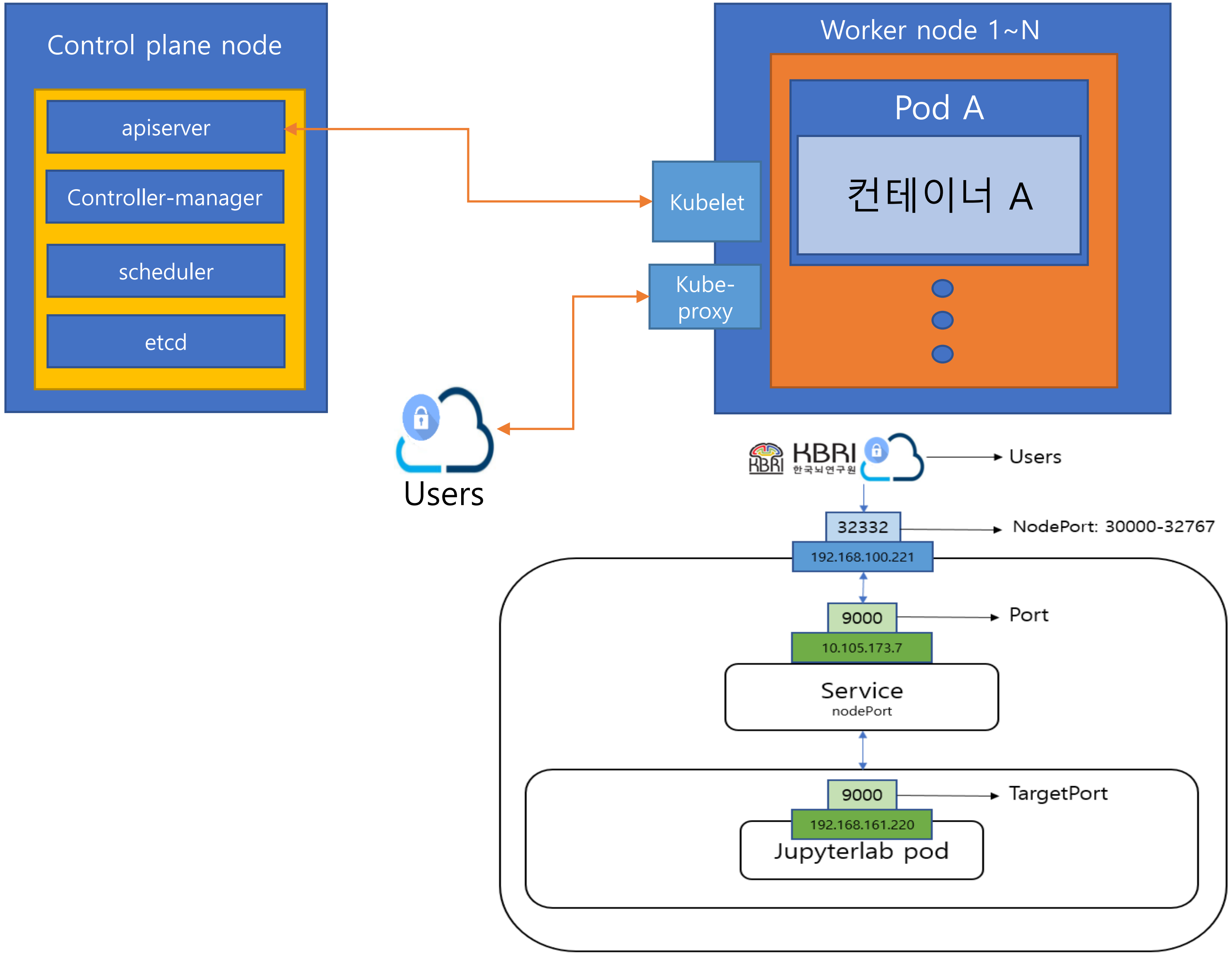
William-silversmith

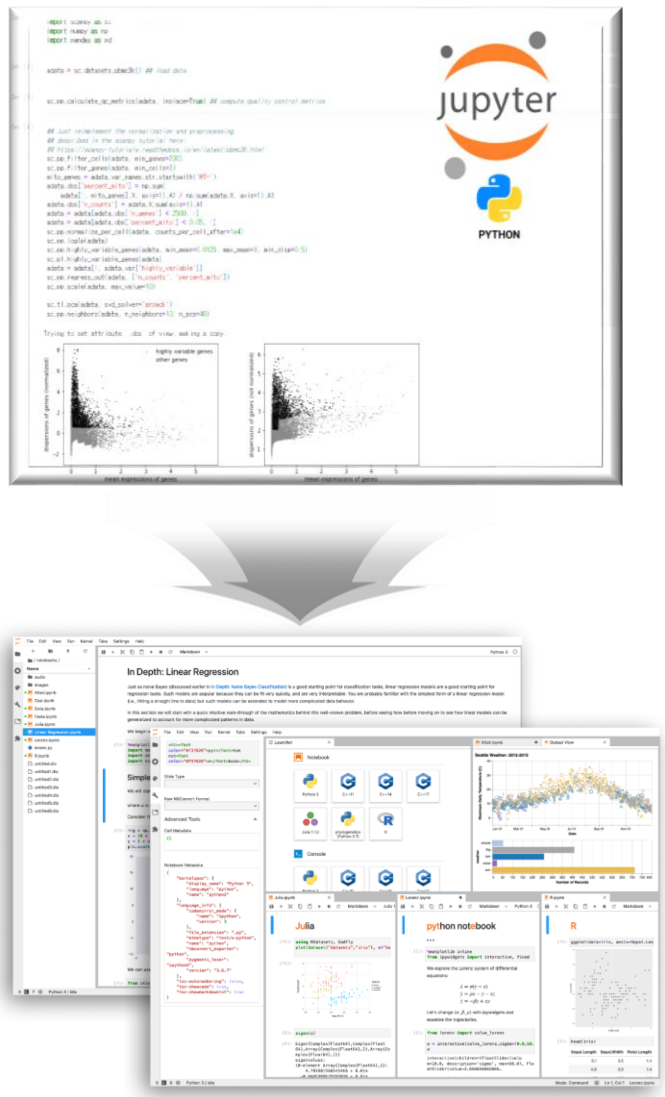
Princeton University: Seung Lab

바이오 분석 파이프라인 App (Jupyterlab 기반 바이오 분석 파이프라인) – 구성도



쿠버네티스 기반 연구자 교육용 주피터랩 컨테이너





```
import scanpy as sc
import numpy as np
import pandas as pd

adata = sc.datasets.pbmc3k() ## load data

sc.pp.calculate_qc_metrics(adata, inplace=True) ## compute quality control metrics

## Just reimplement the normalization and preprocessing
## described in the scanpy tutorial here-
## https://scanpy-tutorials.readthedocs.io/en/latest/pbmc3k.html
sc.pp.filter_cells(adata, min_genes=200)
sc.pp.filter_genes(adata, min_cells=3)
mito_genes = adata.var_names.str.startswith('MT-')
adata.obs['percent_mito'] = np.sum(
    adata[:, mito_genes].X, axis=1).A1 / np.sum(adata.X, axis=1).A1
adata.obs['n_counts'] = adata.X.sum(axis=1).A1
adata = adata[adata.obs['n_genes'] < 2500, :]
adata = adata[adata.obs['percent_mito'] < 0.05, :]
sc.pp.normalize_per_cell(adata, counts_per_cell_after=1e4)
sc.pp.log1p(adata)
sc.pp.highly_variable_genes(adata, min_mean=0.0125, max_mean=3, min_disp=0.5)
sc.pl.highly_variable_genes(adata)
adata = adata[:, adata.var['highly_variable']]
sc.pp.regress_out(adata, ['n_counts', 'percent_mito'])
sc.pp.scale(adata, max_value=10)

sc.tl.pca(adata, svd_solver='arpack')
sc.pp.neighbors(adata, n_neighbors=10, n_pcs=40)

Trying to set attribute `.obs` of view, making a copy.
```

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 관라히기 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 세바스찬승 랩에서 유사하게 연구하고 있는 것을 알게되었습니다. 그래서 미국 SFN 학회에 참석하여 승랩의 William-silversmith 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 오케스테레이션 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

Appendix. 1

2018 한국분자세포생물학회

Fast automated stitch-align pipeline and 3D web tool

Nam Uk Kim[1,2], Ju Yeon Choi[2], Sung Jin Jeong[1,2,*], Yu Jin Jang[2], and Byeong Soo Kang[3]
[1]Neural Circuit Research Group, Korea Brain Research Institute, Daegu, 41068, Korea,
[2]Molecular Aging & Development Laboratory, Korea Brain Research Institute, Daegu, 41068, Korea,
[3]R&D Center, SYSOFT, Daegu, 42988, Korea
*Corresponding author: sjjeong@kbri.re.kr

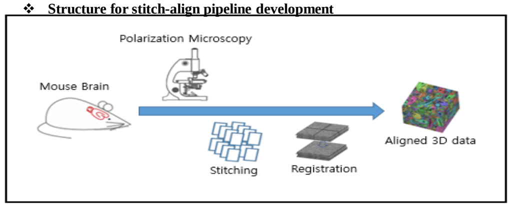
Abstract

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.

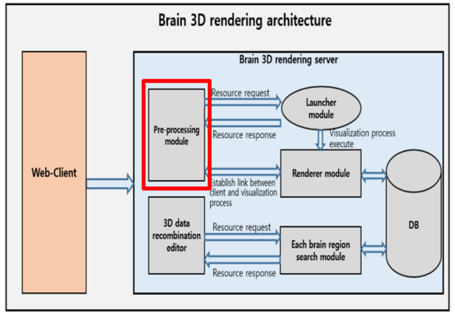
Polarization microscopy such as confocal and light sheet microscope are widely used for neural network and generates high-resolution images.

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 Big-Data**


Brain 3D rendering architecture
Web-Client
Brain 3D rendering server
Pre-processing module
Resource request
Resource response
Launcher module
Visualization process execute
Establish link between client and visualization process
Renderer module
3D data recombination editor
Resource request
Resource response
Each brain region search module
DB

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

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

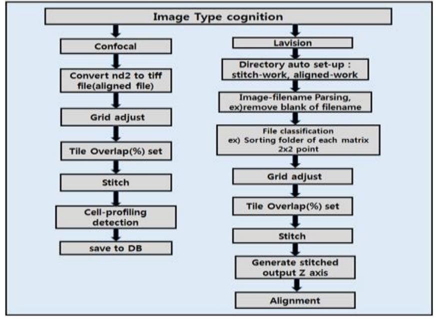
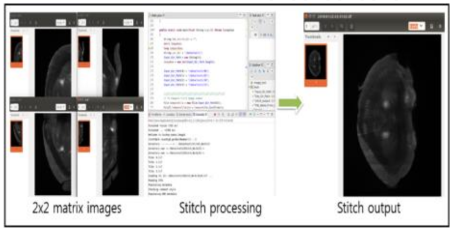


Image Type cognition
Confocal
Convert nd2 to tiff file(aligned file)
Grid adjust
Tile Overlap(%) set
Stitch
Cell-profiling detection
save to DB
Lavision
Directory auto set-up : stitch-work, aligned-work
Image-filename Parsing, ex)remove blank of filename
File classification ex) Sorting folder of each matrix 2x2 point
Grid adjust
Tile Overlap(%) set
Stitch
Generate stitched output Z axis
Alignment

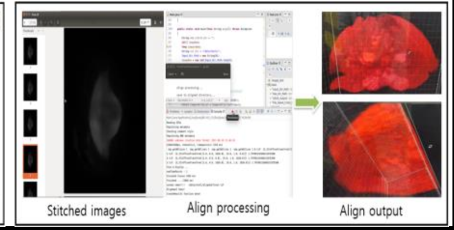
2D images derived from polarization microscopy are input and pre-processing module is performed for each type of inputted images.

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. Mazher lqbal., et al. Image stitching and 2D to 3D Image Reconstruction for Abnormal Activity Detection. International journal of computer applications. vol 133. no17. 2016.
[2] Ebtsam 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)

Appendix. 2

2019 한국뇌연구원(KBRI)

Automated data-set generation pipeline for 3D Neuron visualization and cloud processing in Kbrain-map DB station portal

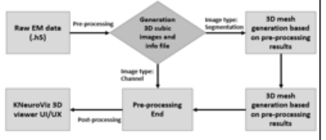
Nam Uk Kim[1,2], Byeong Soo Kang[3], and Sung-Jin Jeong[1,2,*]
Neural Circuit Research Group, Korea Brain Research Institute, Daegu, 41068, Korea,
[2]Molecular Aging & Development Laboratory, Korea Brain Research Institute, Daegu, 41068, Korea,
[3]R&D Center, SYSOFT, Daegu, 42988, Korea
[*]Corresponding author: sjjeong@kbri.re.kr

Abstract

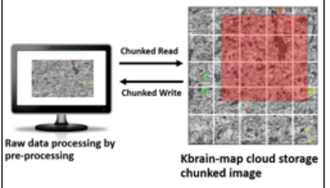
Recently, high-resolution image and video data are being increased exponentially in the field of life sciences by the advanced imaging technologies. Among them, connectomics images of brain tissues converted nano and micrometer high-resolution into 3D image produce big data characteristics. In order to visualize such connectomics image effectively, there are resource constraints such as memory and disk. In addition, as Web services grow in scale, sudden increase in traffic causes server bottlenecks as well as degradation of server performance due to server overloading problems. It is essential to establish a Web-based platform which allows the connectomics images to be visualized and analyzed in 3D at anytime and anywhere without restriction on the spatial environment. Also, in order to improve the processing time and speed of the entire image, it is necessary to divide into the image pieces and distribute them to the system. 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 for visualization and analysis of connectomics images in three dimensions through a web browser. The 3D image visualization utilizes WebGL-based 3D open source. The 3D image visualization system was customized to effectively visualize the data generated by the pipeline. The pipeline sets the bounding box space of x, y, z axis and divides into 3D chunk units by slicing work for each area. The divided 3D chunk dataset and information files are kept in the block storage of the cloud and the dataset is converted to the KNeuroViz format for efficient I/O operations. This system is aimed to build an interactive database for brain connectome convergence research based on a user interface that can be integrated with various analysis modules.

Method : Automated dataset generation pipeline

Cut-outed cubic Image and KNeuroViz 3D viewer flow chart

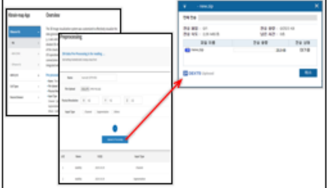

Raw EM data (.h5)
Pre-processing
Generation 3D cubic images and info file
Image type: Segmentation
3D mesh generation based on pre-processing results
Image type: Channel
3D mesh generation based on pre-processing results
Pre-processing End
Post-processing
KNeuroViz 3D viewer UI/UX

The pipeline splits the raw data into chunks and processes them into a data format that can be read by KNeuroViz Web.


Chunked Read
Chunked Write
Raw data processing by pre-processing
Kbrain-map cloud storage chunked image

The image sliced by the bounding box is stored in the cloud storage as a chunk image. (KNeuroViz pre-computed format). Read in chunks and crop to ROI. The Writing object divides the image into chunks. Also, in technical side, Writing requires aligning chunks to avoid race conditions.

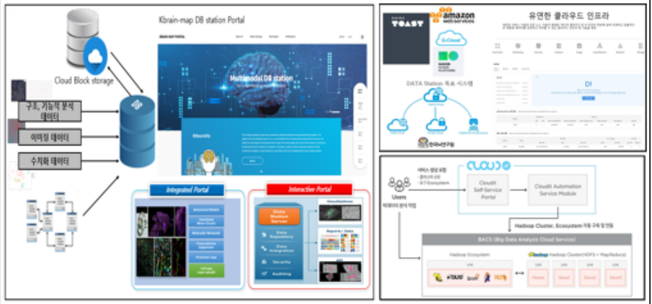
Pre-processing dataset generation pipeline for 3D volume visualization (Front-end)



The pre-processing pipeline is divided into chunk data by bounding boxes for each region of the x, y, and z axes. By creating a divided data set for each x, y, z axis area and related info file, It performs pre-processing to efficiently load and visualize in the web's 3D visualization platform.

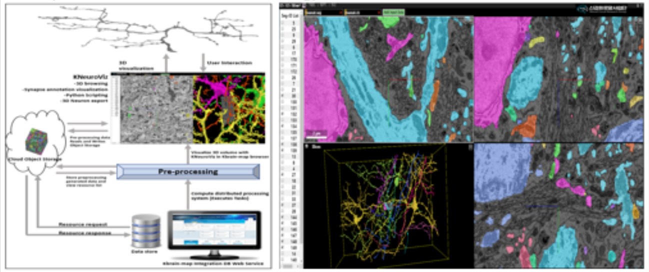
Structure : Web based DB station technology

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



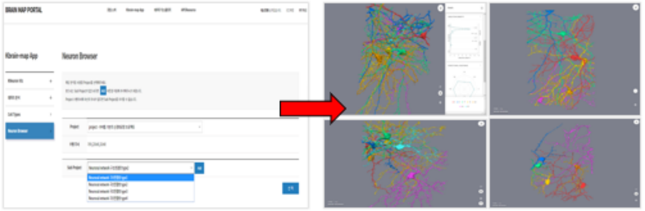
Integrated DB station infrastructure has the advantage of providing optimized IaaS by creating virtual machine for cluster-based hybrid big data analysis environment and big data processing. Also, **the integrated DB station system** provides a platform to share data for integration that can collaborate on data processing between domestic and foreign researchers.

Development: 3D based visualization technology



Visualization of the data stored in cloud using KNeuroViz, one the S/W in Kbrain-map web server. KNeuroViz providers a benefits a real-time streaming of the chunked image data and their visualization in 3D. 3D rendering speed is improved due to the proposed technology (where available).

Kbrain-map App: Neuron browser for visualization and analysis of 3D neuron meshes



The Neuron-browser connects with KNueronViz and provides the ability to find and search cell classifications to determine the structure of connections between neurons.

Future Directions

- Development of function to set MIP level in pre-processing
- System integration for visualization and numerical analysis of specific 3D neurons in Kbrain-map
- Extended to handle various raw data
- KNeuronViz system and GPU memory allocation UI implementation
- Integrated database for 3D neuron information in KNeuroViz
- Support for skeleton 3D reconstruction in pre-processing

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 Brain Research Program through the National Research Foundation of Korea(NRF) (NRF-2017M3C7A1048092)

Appendix. 3

2020 한국뇌신경과학회

Automated data-set generation pipeline for 3D Neuron visualization and cloud processing in Kbrain-map DB station portal

Nam Uk Kim[1,2], Byeong Soo Kang[4], and Sung-Jin Jeong[1,2,3,*]

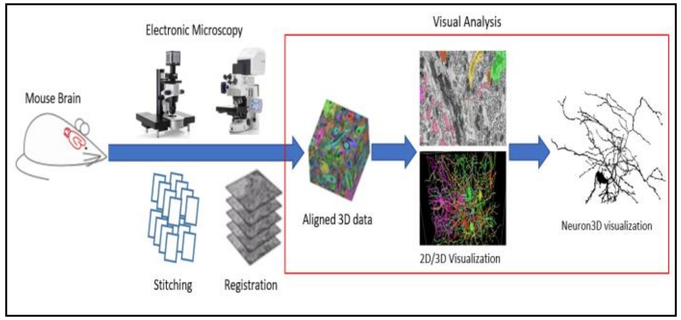
[1]Neural Circuit Research Group, Korea Brain Research Institute, Daegu, 41068, Korea,
[2]Molecular Aging and Development Laboratory, Research Group of Neural Development Disorders and Rare Diseases, Korea Brain Research Institute, Daegu, Republic of Korea,
[3]Department of Brain and Cognitive Sciences, Daegu Gyeongpuk Institute of Science and Technology, Daegu 42988, Republic of Korea,
[4]R&D Center, SYSOFT, Daegu, 42988, Republic of Korea
**Corresponding author: sjjeong@kbri.re.kr*

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

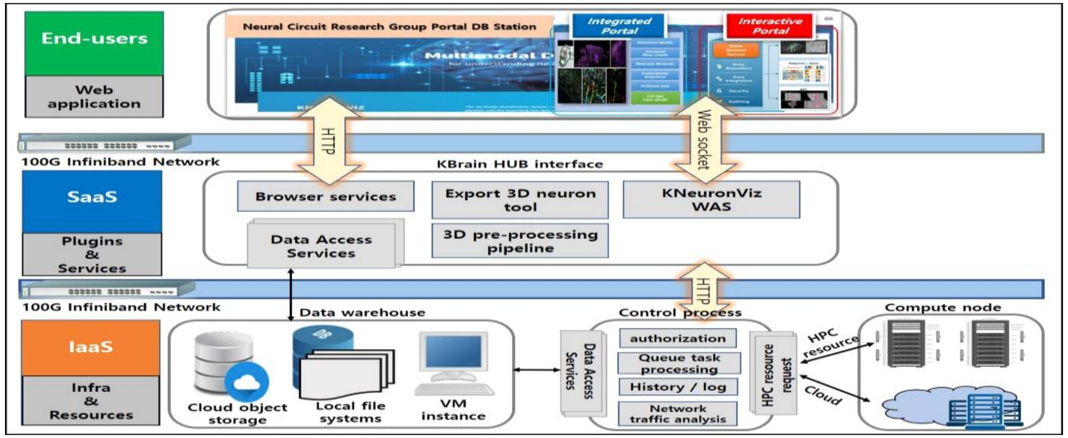
EM(Electron microscope) image analysis process and visual analysis flow chart



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

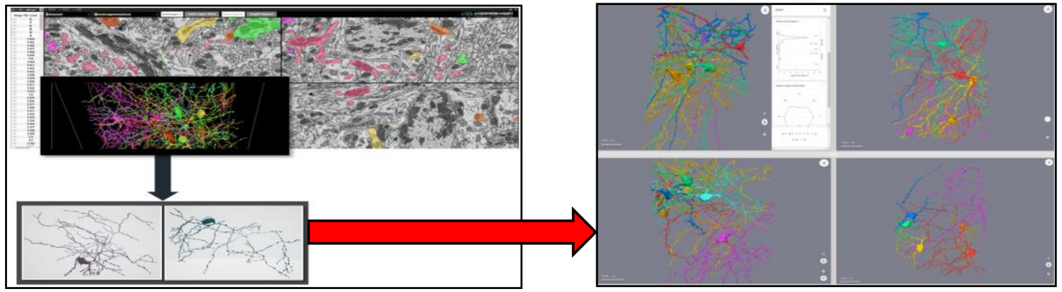
Architecture : Interactive integrated DB station KBrain-map and cloud system


End-users
Web application
Neural Circuit Research Group Portal DB Station
Integrated Portal
Interactive Portal
HTTP
Web socket
100G Infiniband Network
KBrain HUB interface
SaaS
Plugins & Services
Browser services
Export 3D neuron tool
KNeuronViz WAS
Data Access Services
3D pre-processing pipeline
HTTP
100G Infiniband Network
Data warehouse
Control process
Compute node
IaaS
Infra & Resources
Cloud object storage
Local file systems
VM instance
Data Access Services
authorization
Queue task processing
History / log
Network traffic analysis
HPC resource request
HPC resource
Cloud

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

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



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.

The Neuron Browser connects to KNueronViz 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!