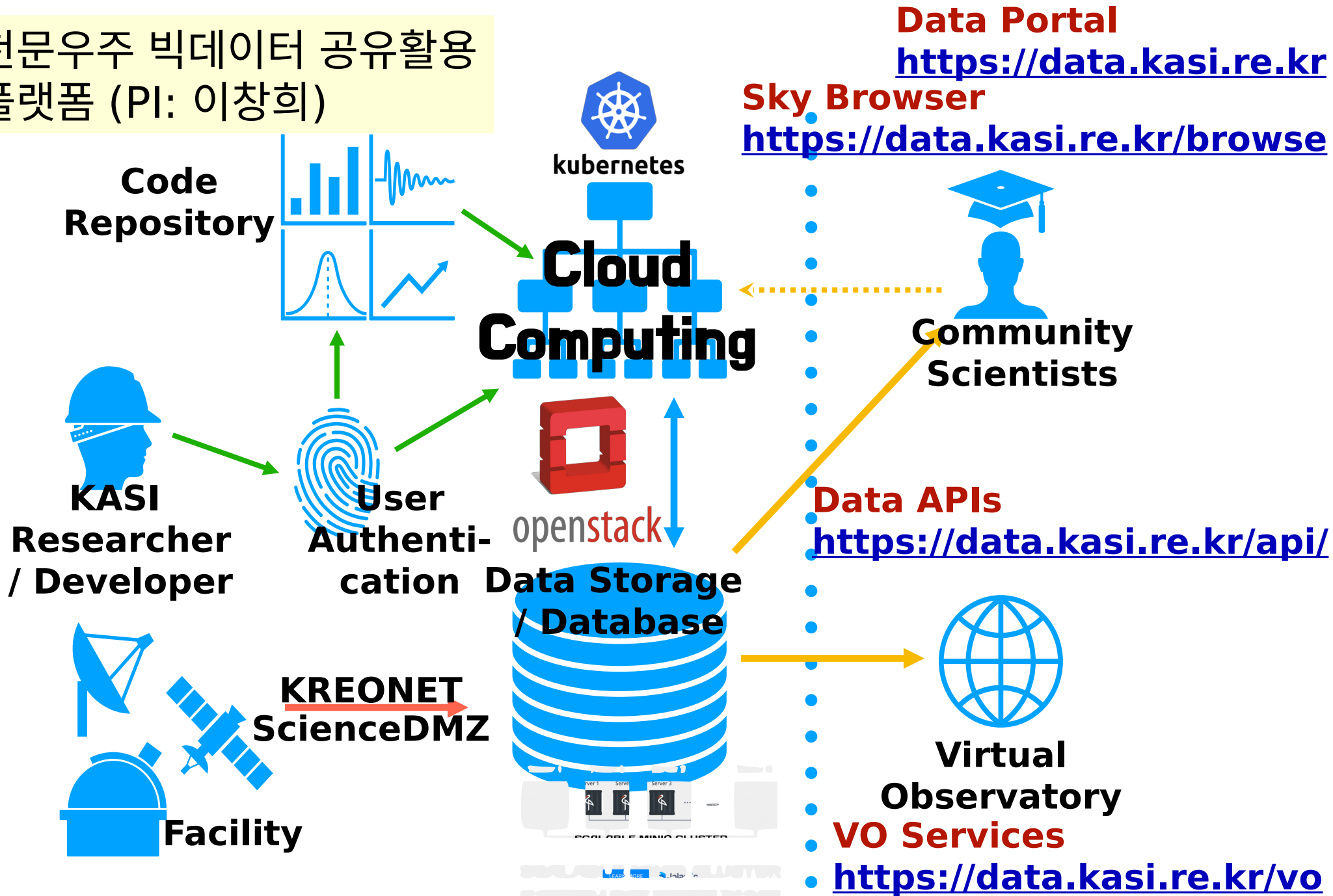


# KASI Science Cloud의 관측 빅데이터 분석 활용 사례 소개

신민수, 이창희, 이재준, 홍성용  
한국천문연구원

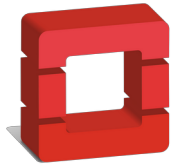
# KASI Science Cloud

천문우주 빅데이터 공유활용  
플랫폼 (PI: 이창희)



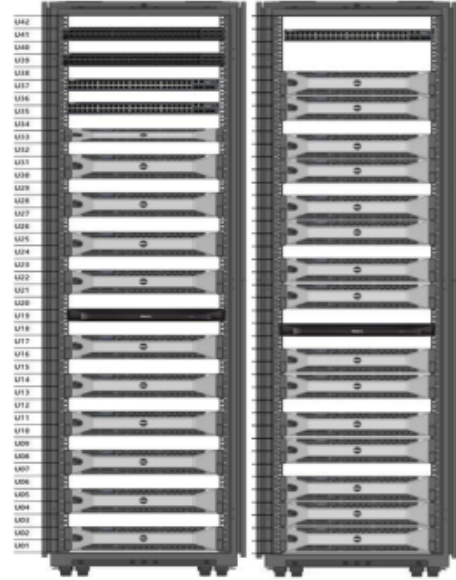
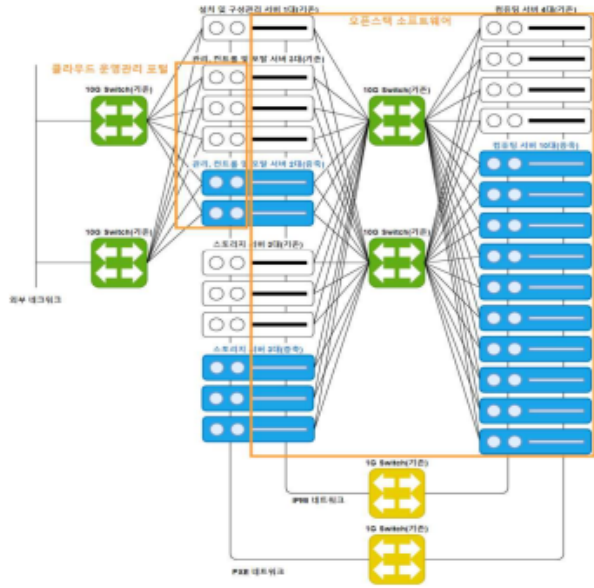
# Cloud computing

- Cloud computing: 미국 NIST에서 제시한 5가지 특징들은 1) on-demand self-service, 2) broad network access, 3) resource pooling, 4) rapid elasticity, and 5) measured service.
- KASI operation 주체로서의 최대 고려 사항: **on-demand self-service + rapid elasticity** (for KASI R&D) with maximum system **utilization**. → IaaS (+ PaaS; 쿠버네티스: 이재준 & 하둡: 홍성용) for KASI research.
- 해외 Private cloud computing IaaS 예들: NSF-funded Jetstream, NASA Center for Climate Simulation's Explore, UK STFC Cloud (for Euclid), UK IRIS and SKA, etc.



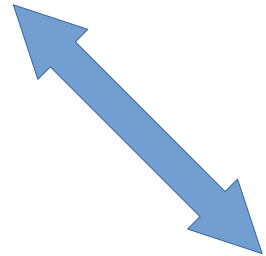
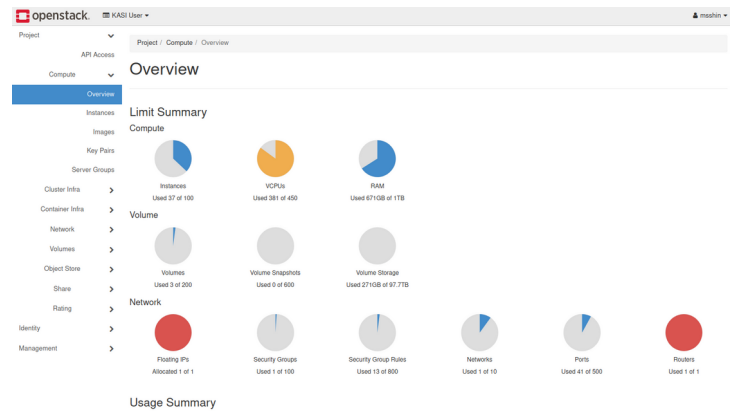
# OpenStack-based Private Cloud

openstack



- 클라우드 컴퓨팅 시스템의  
computing nodes : 672 vCPUs,  
3.5 TB RAM with
- 데이터 공유 공용스토리지 : > 1 PB

See the KASI how-to document:  
<https://data.kasi.re.kr/wiki/howto>



대용량 데이터 공유용 고밀도 오브젝트 스토리지

# 사례 1: 기계학습 활용을 위한 자료 처리

- 목적: 40만 개의 은하들에 대한 200만 개의 FITS 이미지 파일 처리를 통해서 기계학습 모델에 활용할 수 있는 NumPy 자료 생성.
- 방법: GNU Parallel을 이용한 distributed naively parallel processing + Gluster network distributed shared filesystem + analysis Python codes.

```
ubuntu@ml-run-master:~/ml$ df -h
Filesystem      Size  Used Avail Use% Mounted on
udev            16G   0    16G   0% /dev
tmpfs           3.2G  1.1M  3.2G   1% /run
/dev/vda1       243G  4.7G  238G   2% /
tmpfs           16G   0    16G   0% /dev/shm
tmpfs           5.0M   0    5.0M   0% /run/lock
tmpfs           16G   0    16G   0% /sys/fs/cgroup
/dev/vda15      105M  5.2M  100M   5% /boot/efi
/dev/loop0      56M   56M   0 100% /snap/core18/1997
/dev/loop1      71M   71M   0 100% /snap/lxd/19647
tmpfs           3.2G   0    3.2G   0% /run/user/0
ml-input-master:/gvol 7.8T  4.8T  3.1T  62% /mnt/gluster-input
ml-output-master:/gvol 18T  194G  18T   2% /mnt/gluster-output
tmpfs           3.2G   0    3.2G   0% /run/user/1000
/dev/loop3      54M   54M   0 100% /snap/snapd/19122
/dev/loop4      64M   64M   0 100% /snap/core20/1891
/dev/loop5      56M   56M   0 100% /snap/core18/2745
/dev/loop6      92M   92M   0 100% /snap/lxd/24061
ubuntu@ml-run-master:~/ml$
```

→ 128개 vCPU + 26TB  
Gluster volume.

## Galaxy Morphological Classification with Deformable Attention Transformer

Seokun Kang<sup>1</sup>, Minsu Shin<sup>2</sup>, Taehwan Kim<sup>1</sup>  
Artificial Intelligence Graduate School, UNIST<sup>1</sup>  
Korea Astronomy and Space Science Institute<sup>2</sup>



### Introduction

- Problem
  - Galaxy morphological classification is an important in astronomy
  - Most prior work study coarse-level morphological classification and use raster low-dynamic range images
- Challenges
  - Dataset is highly unbalanced.
  - Foreground objects exist in front of the target galaxy.
  - There are many non-values in the dataset that are harmful to training.

### Experiment Results

• The Classification Result of the Galaxy Zoo2 Image Dataset.

Model	Acc Top-1	Acc Top-2
Linformer [2]	80.427%	93.953%
DAT-Small	81.104%	94.373%
DAT-Base	82.222%	94.668%

◦ DAT-Base achieves 82.222% test set accuracy and outperforms Linformer (80.427% test set accuracy in our experiment, 80.55% is reported in [2])

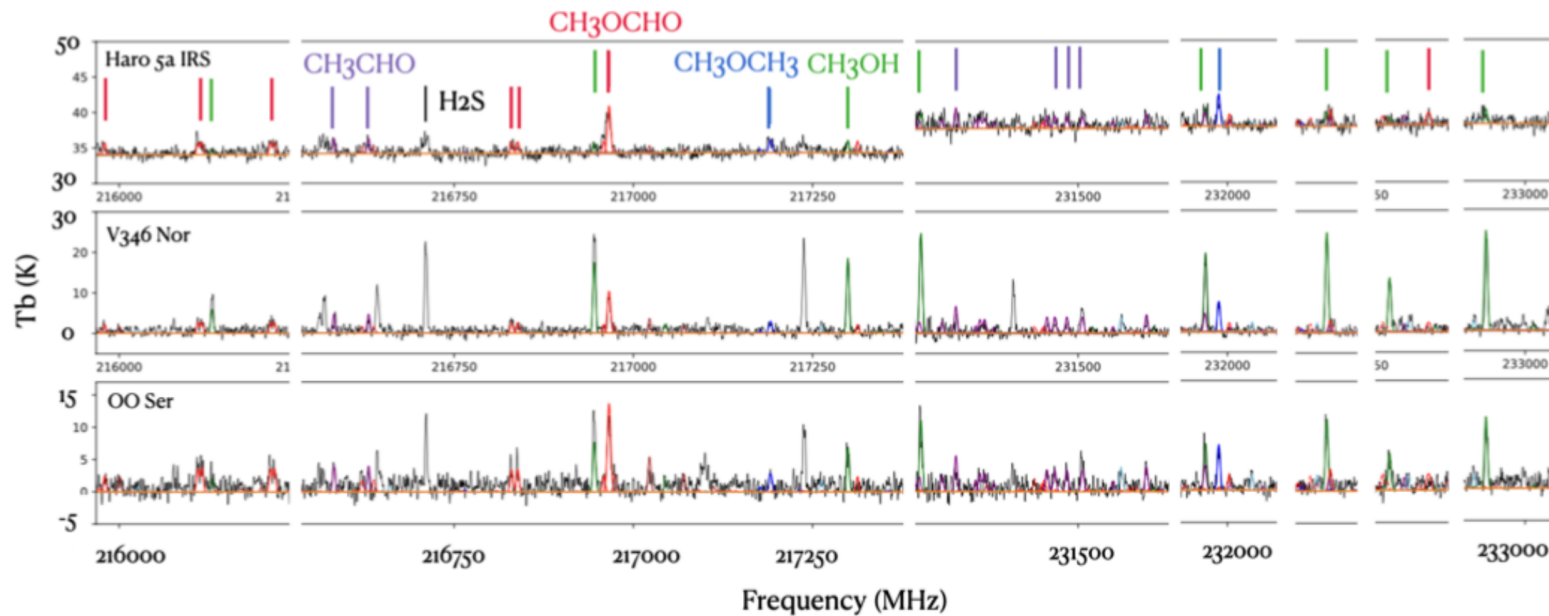
### Confusion Pattern



# 사례2: ALMA spectrum 분석

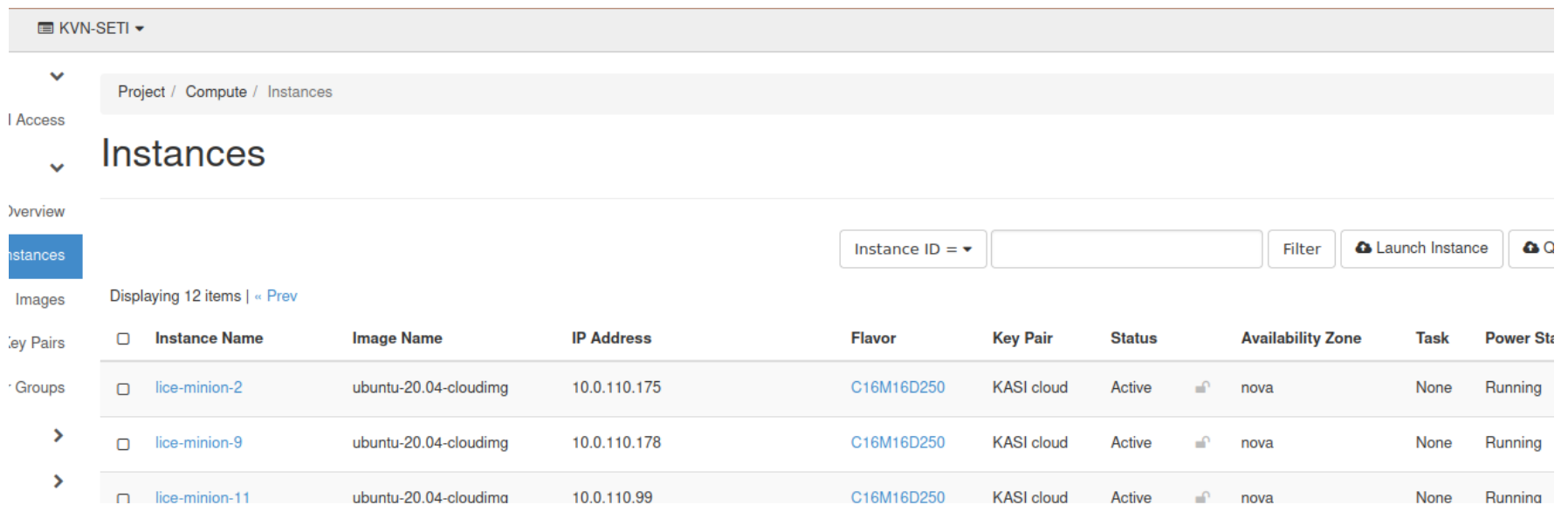
- 목적: ALMA spectrum 분석을 통한 유기분자 생성 연구.
- 방법: MPI-parallelized analysis code in Markov Chain Monte Carlo method.  
→ **96개 vCPU + 192GB RAM.**

여러분자들의 스펙트럼 피팅의 예



# 사례3: KVN 자료 상관처리 분석

- 목적: KVN 자료의 multi-center 상관처리를 통한 새로운 신호, 천체 검출 및 분석.
- 방법: MPI-parallelized 프로그램 + Gluster network distributed shared filesystem  
→ **256 vCPU + 256GB RAM + 45TB Gluster filesystem.**



The screenshot shows the OpenStack dashboard for the 'KVN-SETI' project. The 'Instances' tab is selected, displaying a table of 12 instances. The table columns are Instance Name, Image Name, IP Address, Flavor, Key Pair, Status, Availability Zone, Task, and Power State. Three instances are visible: 'lice-minion-2', 'lice-minion-9', and 'lice-minion-11', all running on the 'ubuntu-20.04-cloudimg' image, 'C16M16D250' flavor, and 'KASI cloud' key pair.

Instance Name	Image Name	IP Address	Flavor	Key Pair	Status	Availability Zone	Task	Power State
lice-minion-2	ubuntu-20.04-cloudimg	10.0.110.175	C16M16D250	KASI cloud	Active	nova	None	Running
lice-minion-9	ubuntu-20.04-cloudimg	10.0.110.178	C16M16D250	KASI cloud	Active	nova	None	Running
lice-minion-11	ubuntu-20.04-cloudima	10.0.110.99	C16M16D250	KASI cloud	Active	nova	None	Running