



2019년도 12월 한국생물공학회 교육 워크숍

생물공학×기계학습

- 일시 _ 2019년 12월 19일(목) ~ 20일(금)
- 장소 _ 한국과학기술회관 중회의실 (서울시 강남구)
- 주최 _ 한국생물공학회

튜토리얼 1

한국생명공학연구원
합성생물학전문연구단
김하성 선임연구원

목차

- Docker & Git
 - Python & Biopython
 - Machine learning & Deep learning
-
1. 실습 진행 docker (windows, MacOS) > colab
 2. jupyterlab, colab 실행 끝날때까지 기다리기
 3. 자료는 모두 배포 예정

딥러닝을 위한 컴퓨터 환경 설정

1. Download Anaconda
2. Install Anaconda & Python
3. Start and Update Anaconda
4. Install CUDA Toolkit & cuDNN
5. Create an Anaconda Environment
6. Install Deep Learning API's (TensorFlow & Keras)



Linux

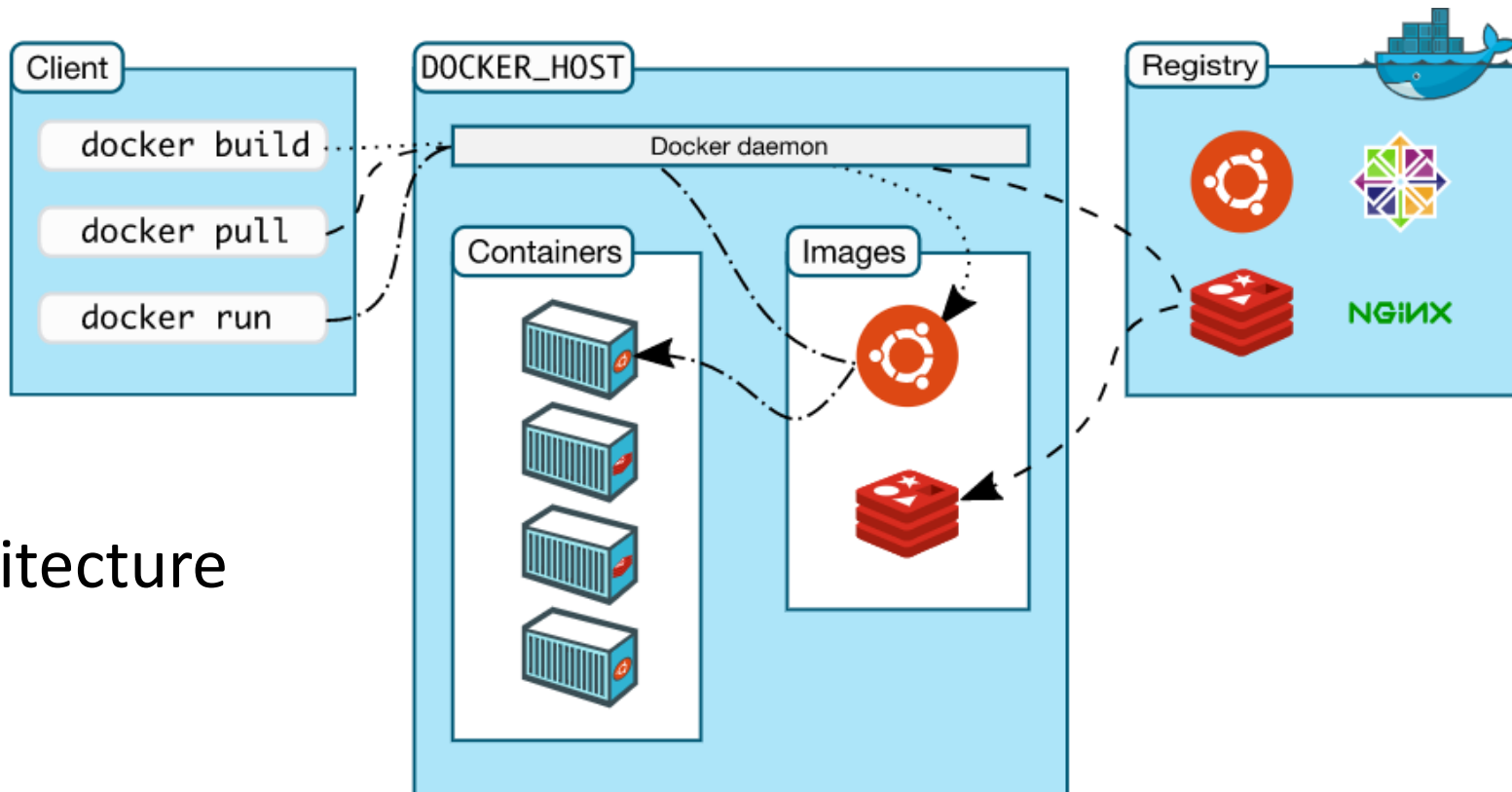
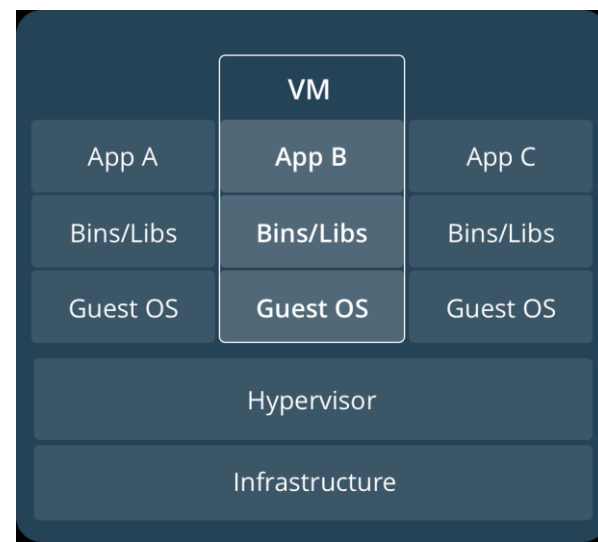
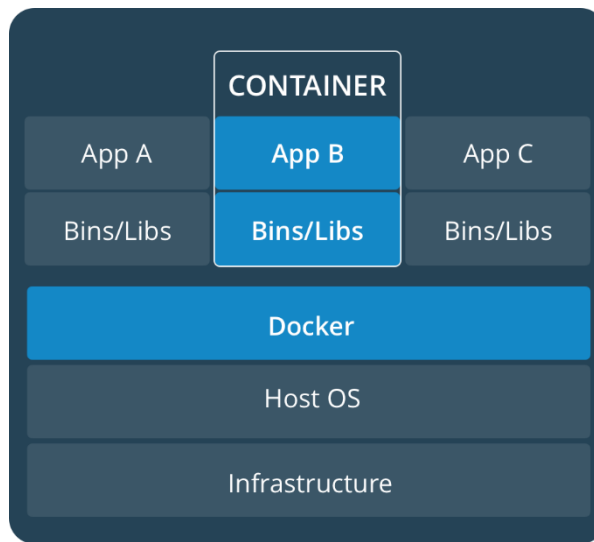
데이터 분석 자체 보다 더 복잡하고 까다로움



docker

- 리눅스 컨테이너 기반 오픈소스 가상화 플랫폼
 - 컨테이너 - 격리된 리소스 이용 프로세스 동작
 - 기존 가상화 기술 - 하드웨어 > Guest OS > 프로세스
 - Docker - 하드웨어 > OS > 도커엔진 > 프로세스
 - 개발환경 == 배포환경 (App + OS)
- ➔ 다른사람 개발환경 == 내 개발환경

Container vs. VM



Architecture

Example



NCBI Resources ☒ How To ☒

ORFfinder

PubMed ▼

Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for [Linux x64](#).

Examples (click to set values, then click Submit button) :

- [NC_011604](#) Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- [NM_000059](#); genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt

Try the two-factor authentication beta. [Learn more >](#)

[Explore](#)[Repositories](#)[Organizations](#)[Get Help](#)[haseong](#)[Docker EE](#)[Docker CE](#)[Containers](#)[Plugins](#)

Filters

1 - 2 of 2 results for **orffinder**. [Clear search](#)

Most Popular

Docker Certified

☐ ☒ Docker Certified

Images

☐ Verified Publisher
*Docker Certified And Verified Publisher
Content*



unlhcc/orffinder

By [unlhcc](#) • Updated 2 months ago

98
Downloads

Container

Linux

x86-64

```
PS C:\bioengml\datasets> docker pull unlhcc/orffinder
```

```
Using default tag: latest
```

```
latest: Pulling from unlhcc/orffinder
```

```
d8d02d457314: Pull complete
```

```
388cf7f58f53: Pull complete
```

```
a62e8a8a05cd: Pull complete
```

```
e7d70c76806b: Pull complete
```

```
Digest: sha256:cbe18a0055fa41074af946d40fb7da1fe8c683b2e5dc35115d165e9047e06e27
```

```
Status: Downloaded newer image for unlhcc/orffinder:latest
```

```
docker.io/unlhcc/orffinder:latest
```

```
PS C:\bioengml\datasets>
```

docker run with ORFfinder

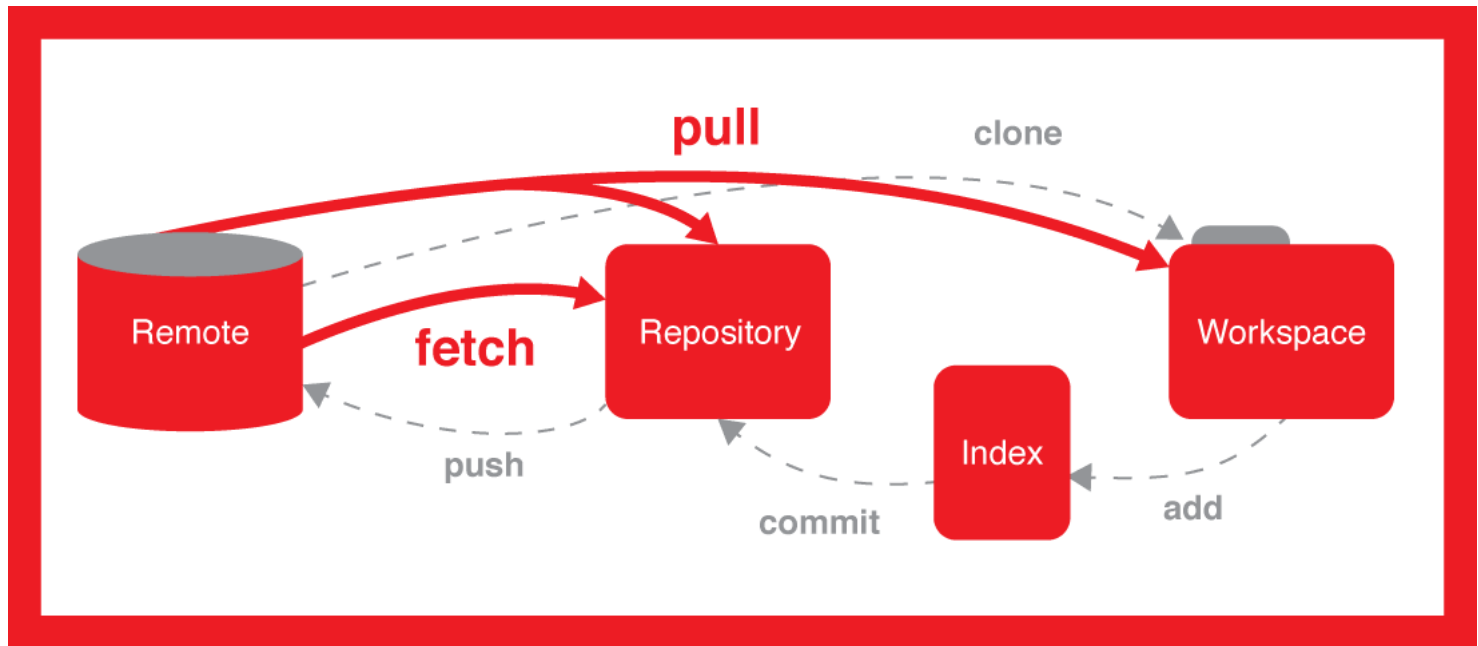
```
PS C:\bioengml\datasets> docker run --rm -v c:\bioengml:/home/bioengml -w /home/bioengml unlhcc/orffinder:latest ORFfinder -in ./datasets/NC_005816.fna -out ./datasets/orf
PS C:\bioengml\datasets> dir
```

디렉터리: C:\bioengml\datasets

Mode	LastWriteTime	Length	Name
-a----	2019-12-07 오후 11:15	235482	ls_orchid.gbk
-a----	2019-12-13 오후 9:33	1022	my_seq_records.fa
-a----	2019-12-07 오후 9:20	9853	NC_005816.fna
-a----	2019-12-07 오후 10:04	31838	NC_005816.gb
-a----	2019-12-13 오후 9:33	12932	opuntia.aln
-a----	2019-12-13 오후 9:33	352	opuntia.dnd
-a----	2019-12-08 오전 10:12	7292	opuntia.fasta
-a----	2019-12-18 오전 12:46	15842	orf
-a----	2019-12-08 오후 11:47	1055056	PF14532_full.txt

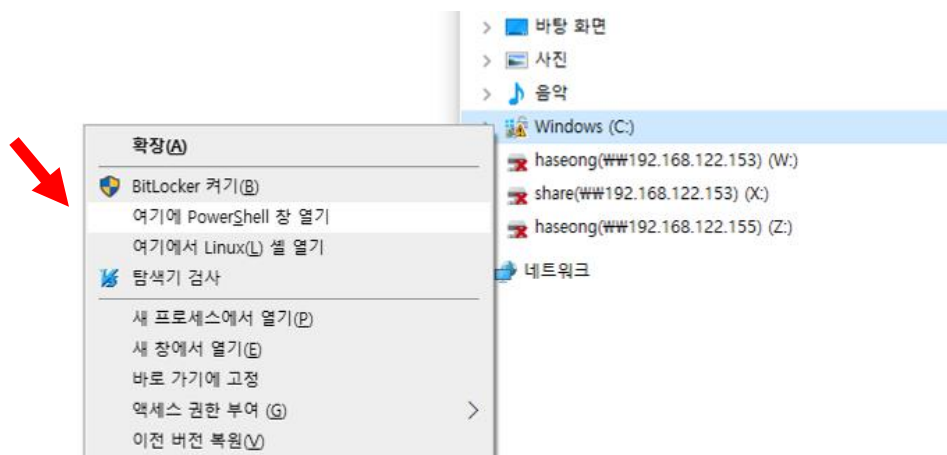


버전 관리 도구



git

- 브라우저의 c:W 드라이브에서 Shift + 마우스오른쪽버튼 클릭
PowerShell 실행



- 다음 명령으로 필요 파일들 복사 및 이미지 다운로드

수업 전
미리 실행

```
> git clone https://github.com/greendaygh/bioengml.git  
> cd bioengml  
> docker pull haseong/bioengml:v1  
> docker images
```

도커 컨테이너 실행

브라우저의 c:\Wbioengml 폴더에 Shift + 마우스 오른쪽 버튼 클릭
PowerShell 실행

```
Windows PowerShell
PS C:\bioengml> docker ps
CONTAINER ID          IMAGE                COMMAND              CREATED              STATUS
PORTS                NAMES
PS C:\bioengml> .\docker_run.bat

C:\bioengml>docker run --rm -it -d -p 8888:8888 --name bioengml -v c:\bioengml:/home/
bioengml haseong/bioengml:v1 jupyter lab --no-browser --allow-root --ip=0.0.0.0 --note
book-dir=/home/bioengml --NotebookApp.token=
72b6b5f619c49ce7fe2f7169e164867a8a226904f8ca490badf44e7fb1d4be52
PS C:\bioengml> docker ps
CONTAINER ID          IMAGE                COMMAND              CREATED
STATUS                PORTS                NAMES
72b6b5f619c4          haseong/bioengml:v1 "jupyter lab --no-br..." 2 minutes ago
Up 2 minutes          0.0.0.0:8888->8888/tcp bioengml
PS C:\bioengml>
```

git

- MacOS에는 git이 기본적으로 설치되어 있음
- 터미널 실행
- 다음 명령으로 필요 파일들 복사 및 이미지 다운로드 (약 5.3GB)

수업 전
미리 실행

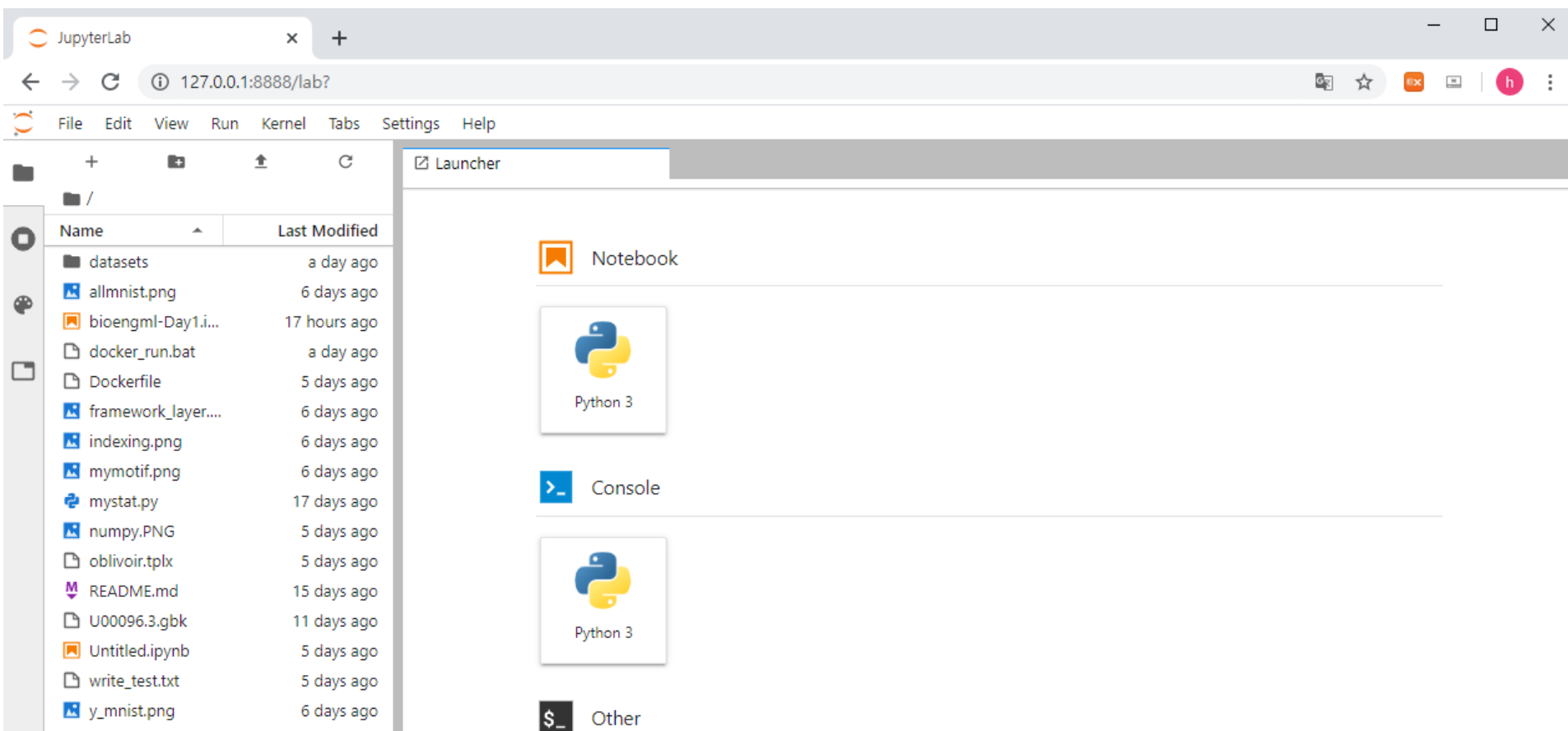
```
> cd /tmp/  
/tmp> git clone https://github.com/greendaygh/bioengml.git  
/tmp> cd bioengml  
/tmp/bioengml> docker pull haseong/bioengml:v1  
/tmp/bioengml> docker images
```

도커 컨테이너 실행

```
/tmp/bioengml> docker run --rm -it -d -p 8888:8888 --name bioengml -v  
/tmp/bioengml:/home/bioengml haseong/bioengml:v1 jupyter lab --no-browser --  
allow-root --ip=0.0.0.0 --notebook-dir=/home/bioengml --NotebookApp.token=
```

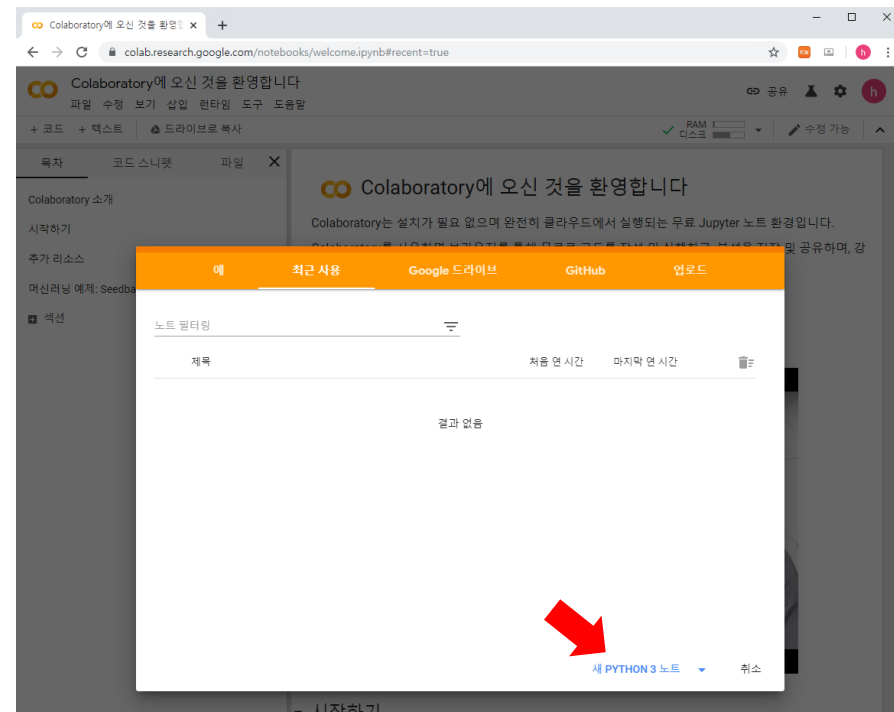
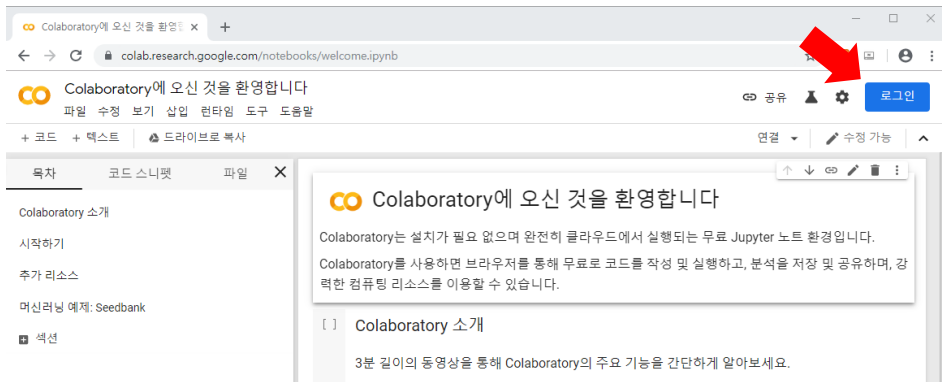
<https://github.com/greendaygh/bioengml>

Jupyter lab 실행

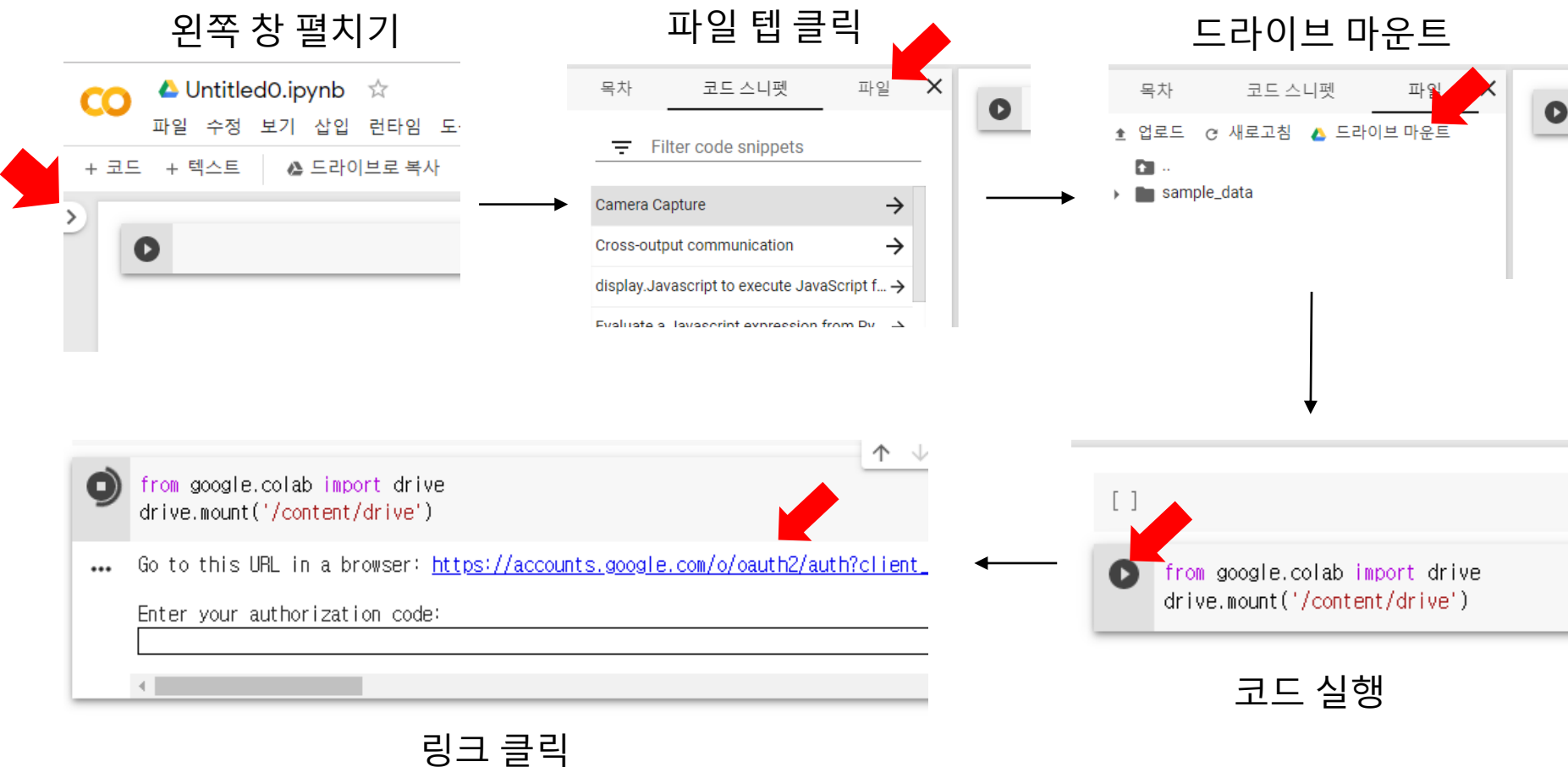


colab

- <http://colab.research.google.com>
- 구글 계정으로 로그인
- 새 PYTHON3 노트 시작



구글 드라이브 연결




2.


Google 계정으로 로그인




Google Drive File Stream이(가) 내
Google 계정에 액세스하려고 합니
다

 kimhaseong@gmail.com

이렇게 하면 Google Drive File Stream에서 다음 작
업을 할 수 있습니다.

 Google 드라이브 파일 보기, 수정, 생성, 삭제 ⓘ

 Google 포토의 사진, 동영상, 앨범을 봅니다. ⓘ

● 프로필 및 연락처와 같은 Google 사용자 정보 ⓘ
조회

● Google 드라이브 문서 보기, 수정, 생성, 삭제 ⓘ

Google Drive File Stream 앱을 신뢰할 수 있는지 확
인

민감한 정보가 이 사이트 또는 앱과 공유될 수 있습니다.
Google Drive File Stream의 [서비스 약관](#) 및
[개인정보처리방침](#)을 검토하여 내 데이터가 어떻게 처리
되는지 알아보세요. 언제든지 [Google 계정](#)에서 액세스
권한을 확인하고 삭제할 수 있습니다.

[타사 앱 권한 부여에 관한 위험 알아보기](#)

취소

허용

3.

Google
로그인

이 코드를 복사하여 애플리케이션으로 전환한 다음 붙여
넣으세요.

4/uAE-
y9EfzNfauR_LMHHb5ezpnHsUBrr6SxY2LMtNR6j

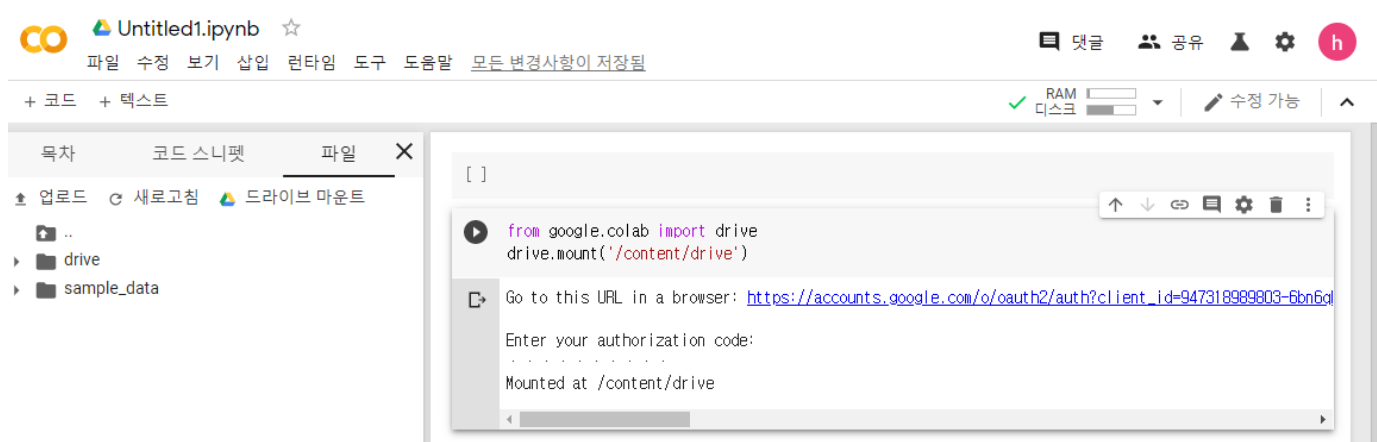


코드 붙여 넣고 Enter 입력

colab



구글 드라이브와 연동 완료



+ 코드 클릭



git

colab

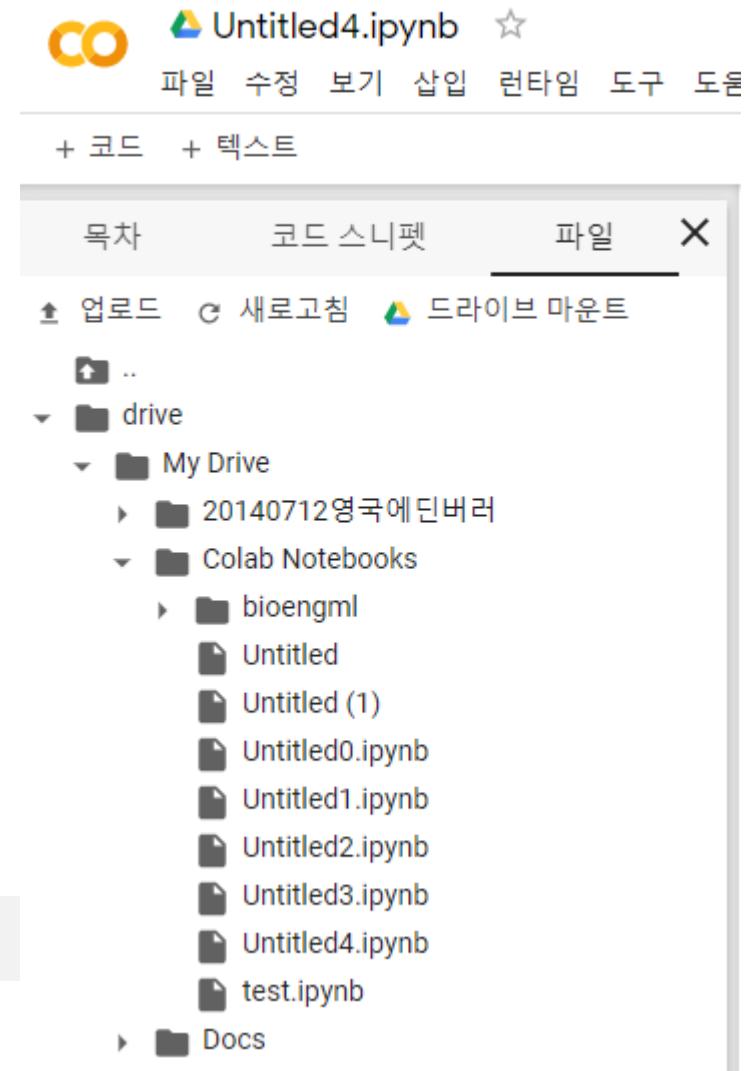
/content/drive/My Drive/Colab
Notebooks/bioengml 경로 있는지 확인

수업 내용 다운로드 (구글 드라이브
에서 삭제 후 다시 다운로드)

```
!git clone https://github.com/greendaygh  
/bioengml.git "/content/drive/My Drive/C  
olab Notebooks/bioengml"
```

biopython 모듈 설치

```
!pip install biopython
```



.ipynb 파일 열기

- colab 의 파일 > 드라이브에서 찾기
- 새창 열리면 bioengml 디렉토리 클릭
- bioengml-Day1.ipynb 파일 오른쪽 마우스버튼
- 연결 앱 > Google colaboratory