

# NSML, 자원 배분

2022 한국어 AI 경진대회

# 목차

1. NSML 대회용 사이트
2. NSML 구조
3. NSML 실행
4. NSML 자원 배분, 제약 조건
5. submit vs submit -t
6. submit 제약(시간 간격, 실행 시간)



# 경진 대회용 사이트

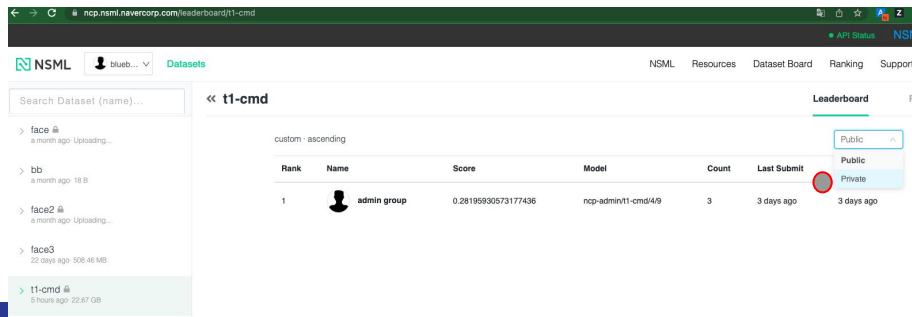
주소: <https://ncp.nsml.navercorp.com/>

문서: [https://n-clair.github.io/ai-docs/\\_build/html/ko\\_KR/index.html](https://n-clair.github.io/ai-docs/_build/html/ko_KR/index.html)

leaderboard (종합) : <https://ncp.nsml.navercorp.com/ranking>

leaderboard (상세) : 각 dataset 클릭하여 확인(session-id 및 checkpoint확인 가능)

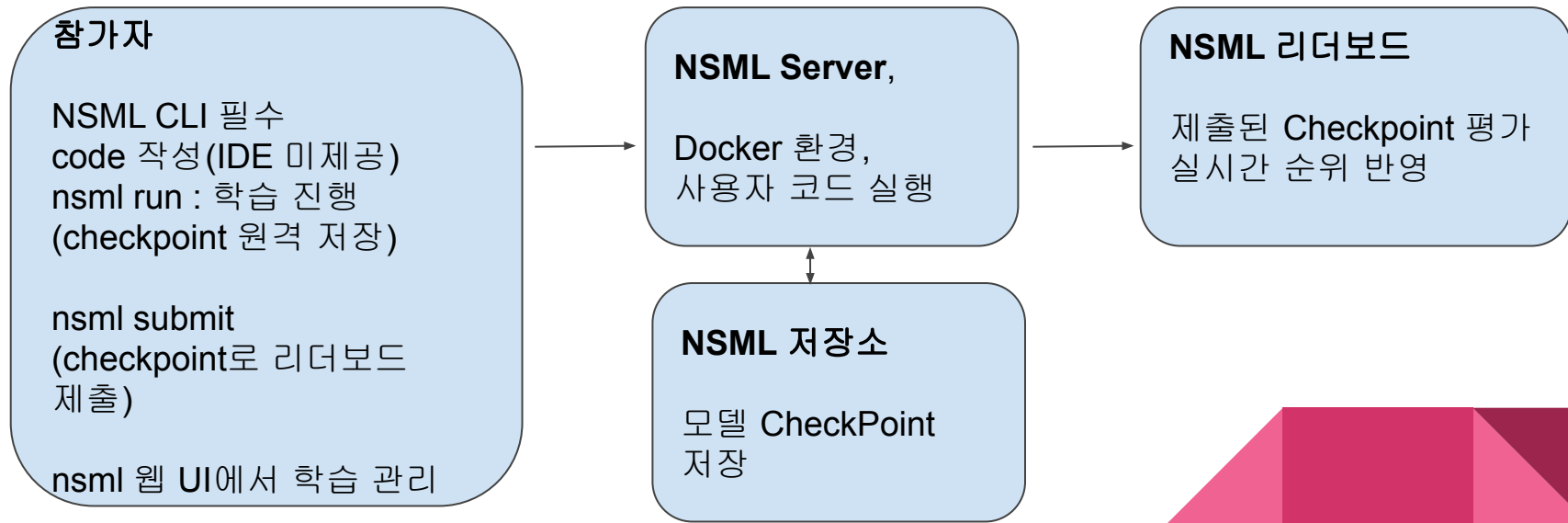
leaderboard (개인별) : 개인의 제출 기록을 확인할 수 있음



The screenshot shows the NSML Leaderboard interface for the 't1-cmd' dataset. On the left, a sidebar lists datasets: 'face', 'bb', 'face2', 'face3', and 't1-cmd' (selected). The main area displays a table of submissions for 't1-cmd' in 'custom' order, ascending. The table has columns for Rank, Name, Score, Model, Count, and Last Submit. The top entry is from the 'admin group' with a score of 0.28195930573177436. A red circle highlights the 'Public/Private' toggle switch, which is currently set to 'Public'.

Rank	Name	Score	Model	Count	Last Submit
1	admin group	0.28195930573177436	ncp-admin/t1-cmd/4/9	3	3 days ago

# NSML 구조



# NSML 실행(CLI help 및 문서 참고)

## baseline 실행 과정

1. git clone <https://github.com/KoreanAI2022/2022-Korean-AI-Competition.git>
2. nsml CLI download : os 종류별(<https://ncp.nsml.navercorp.com/download>)
3. nsml login [github.com-id, email은 지원 안됨]
4. nsml run -d [dataset] -e [entryfile:main.py] : 실행 위치의 directory를 서버로 전송
5. nsml model ls [session-id] : 학습중 저장한 model checkpoint확인
6. nsml submit -t [session-id] [checkpoint] : checkpoint로 모의 submit 실행(stdout/stderr 확인 가능), submit -t 실행 후 종료(삭제) 필수(종료해야 쿼터 복원됨)
7. 6번의 모의 submit 세션 종료(삭제), nsml stop [session-id], nsml rm -f [session-id]
8. nsml submit [session-id] [checkpoint] : 실제 submit 진행, 실행 결과 leaderboard에 자동 반영
9. nsml rm -f [session-id] : session 및 session에 연동된 checkpoint도 함께 삭제됨(최고 score와 관련된 코드는 재현을 위해서 보존해야함)

# NSML 실행(예, CLI, mac 1/2)

```
ykkim@AL01978801 nia % git clone https://github.com/KoreanAI2022/2022-Korean-AI-Competition.git  
Cloning into '2022-Korean-AI-Competition'...
```

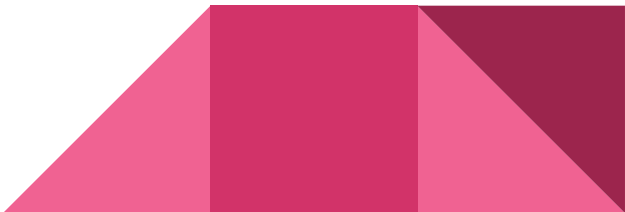
```
ykkim@AL01978801 nia % curl -O -L https://github.com/n-CLAIR/nsml_client-ncpnsml/raw/master/nsml_client.darwin.amd64.ncpnsml.tar.gz  
ykkim@AL01978801 nia % tar xvpf nsml_client.darwin.amd64.ncpnsml.tar.gz
```

```
ykkim@AL01978801 nia % export PATH=$PATH:`pwd`
```

```
ykkim@AL01978801 nia % cd 2022-Korean-AI-Competition/baseline
```

```
ykkim@AL01978801 baseline % nsml login  
INFO[2022/08/17 10:06:28.584] connecting to ncp-cli.nsml.navercorp.com:443  
GitHub Username: bluebrush  
GitHub Password: *****  
INFO[2022/08/17 10:06:34.456] Welcome to NSML!
```

```
ykkim@AL01978801 baseline % nsml run -d t1-cmd -e main.py  
INFO[2022/08/17 10:06:53.478] .nsmlignore check - start  
Building docker image. It may take a while ⇐ 긴 시간이 소요될 수 있음.  
.....  
Session ncp-admin/t1-cmd/11 is started ⇐ [session-id]가 생성됨.
```



# NSML 실행(예, CLI, mac 2/2)

```
ykkim@AL01978801 baseline % nsml ps
```

Name	Created	Args	Status	Summary	Description	# of Models	Size	Type
ncp-admin/t1-cmd/11	a minute ago		Running			0	116.72 KB	normal

```
ykkim@AL01978801 baseline % nsml model ls ncp-admin/t1-cmd/11
```

Checkpoint	Last Modified	Elapsed	Summary	Size
0	7 minutes ago	6.261	epoch=0, train_loss=2.873322969022712, train_cer=1.102525072732406, val_loss=1.1990902568655284, val_cer=0.9576686819830713, number_of_files=1	599.42 MB

```
ykkim@AL01978801 baseline % nsml submit -t ncp-admin/t1-cmd/11 0
```

```
.....  
Session ncp-admin/t1-cmd/12 is started  
.Infer test set takes 43.59376645088196 seconds  
.Done  
You can see session log using [nsml logs ncp-admin/t1-cmd/12]
```

```
ykkim@AL01978801 baseline % nsml submit ncp-admin/t1-cmd/11 0
```

```
.....  
Building docker image. It may take a while  
.....ncp-admin/t1-cmd/13: Session ready  
ncp-admin/t1-cmd/13: Load model  
load nsml model takes 5.551342248916626 seconds  
.Infer test set. The inference should be completed within 3600 seconds.
```

# NSML 실행 (예, web UI)

← → ↻ ncp.nsmml.navercorp.com

NSML admin... Datasets NSML Resources

Search Dataset (name)...

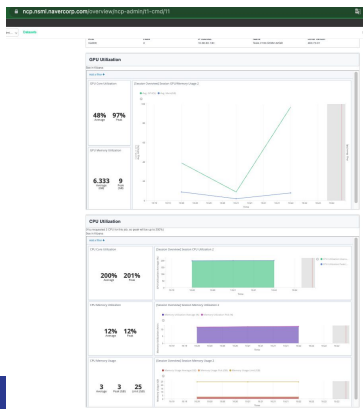
- > t1-cmd 17 hours ago · 22.67 GB
- > t2-car 17 hours ago · 23.51 GB
- > t2-conf 17 hours ago · 23.21 GB
- > t1-free 17 hours ago · 33.4 GB

<< Sessions [ admin group ]

Name Search ... Status: All

Select

<input type="checkbox"/>	t1-cmd	24.00	2 minutes ago
<input type="checkbox"/>	ncp-admin/t1-cmd/t1	2 minutes exec · Running	



← → ↻ ncp.nsmml.navercorp.com/terminal/ncp-admin/t1-cmd/t1

NSML admin... Datasets Sessions

Search Session (number)...

Common Args 0, Selected 1

> rev : 11 Running 4 minutes ago

> rev : 6 Exit 5 days ago

ncp-admin/t1-cmd/t1

NSML Log start!

[INFO] create\_script started...

[INFO] Preprocessing to Done

split dataset start !!

Applying Spec Augmentation...

Applying Spec Augmentation...

[INFO] Epoch 0 start

[INFO] train Start

[INFO] step: 0/2400, Loss: 41.389245, acc: 24.00, elapsed: 2.38s 0.20w 0.00s, lr: 0.000000

[INFO] step: 10/2400, Loss: 8.372887, acc: 32.40, elapsed: 1.40s 1.40w 0.00s, lr: 0.000000

[INFO] step: 100/2400, Loss: 5.840853, acc: 0.46, elapsed: 3.40s 1.00w 0.00s, lr: 0.000000

[INFO] step: 150/2400, Loss: 4.887236, acc: 0.56, elapsed: 1.51s 4.10w 0.00s, lr: 0.000005



# 자원 배분(각 조건 AND조건), 제약 조건

1. GPU: 팀당 V100 2장
2. CPU: 팀당 16장
3. RAM: 170GB (shared-memory포함)
4. 동시 실행: 2개 권장(초과시 임의 제제가 있을 수 있음)
5. 권장 실행 환경
  - a. 유형 1 (x 2개) 세션: `nsml run -d [dataset] -g 1 -c 8 --memory 75G --shm 10G`
  - b. “학습 1개 세션 + submit 1개 세션” 또는 “학습 2개 세션”
6. 모델 checkpoint 저장 공간 제한
  - a. 권장 : 450GB, 초과시 삭제 요청 및 임의 삭제, 추가 저장 불가 등의 조치가 될 수 있음
  - b. <https://ncp.nsml.navercorp.com/settings/profile> 에서 [Storage Management]탭에서 사용량 확인 및 삭제 가능 (삭제시 최고 score의 학습 세션은 삭제하지 않도록 주의)

## “submit -t” vs “submit”

`nsml submit -t` : `submit` 과정에 문법적 오류가 없는지 확인하는 과정, `stderr/stdout`을 볼 수 있으며, 일반 세션과 동일하게 “`nsml ps`”, “`nsml rm -f`”, “`nsml stop`”으로 제어 가능. 작은 셋으로 구성되서 완료 시간이 빠름

`nsml submit` : 리더보드의 제출 과정으로 `stderr/stdout`을 확인할 수 없음.



# submit 시 제약 사항

submit시 “자원 제약”, “시간 제약”, “간격 제약”

submit과정은 팀(자신)의 쿼터를 사용하여 진행하게 됩니다. 따라서 학습을 진행중일 때 쿼터 부족으로 submit을 진행하지 못할 수도 있습니다. (쿼터는 2GPU씩 지급되었음)

학습에서 2GPU를 사용하였다면, submit에서도 2GPU가 할당됩니다. 단 CPU core는 최대 4CPU만 할당됩니다. (submit과정에 CPU 연산이 많다면 영향을 받을 수 있습니다.)

submit는 각 데이터셋 별로(TRACK1-1,1-2,2-1,2-1 각 4개 따로 적용) 1시간에 1회만 진행가능합니다. submit을 성공한 시점 기준으로 1시간 제한이 적용됩니다.

submit 과정은 1시간이 이내에 끝나야하며, 1시간을 초과하는 경우 오류로 종료됩니다.오류에 의한 종료시 다시 1시간을 기다리지 않아도 됩니다.