AI Hub

aihubshell 가이드

본 문서는 aihubshell 가이드로서 aihubshell을 이용하여 데이터셋을 조회 및 다운로드, 활용하는데 도움을 드립니다.

[목차]

1. aihubshell 이용방법

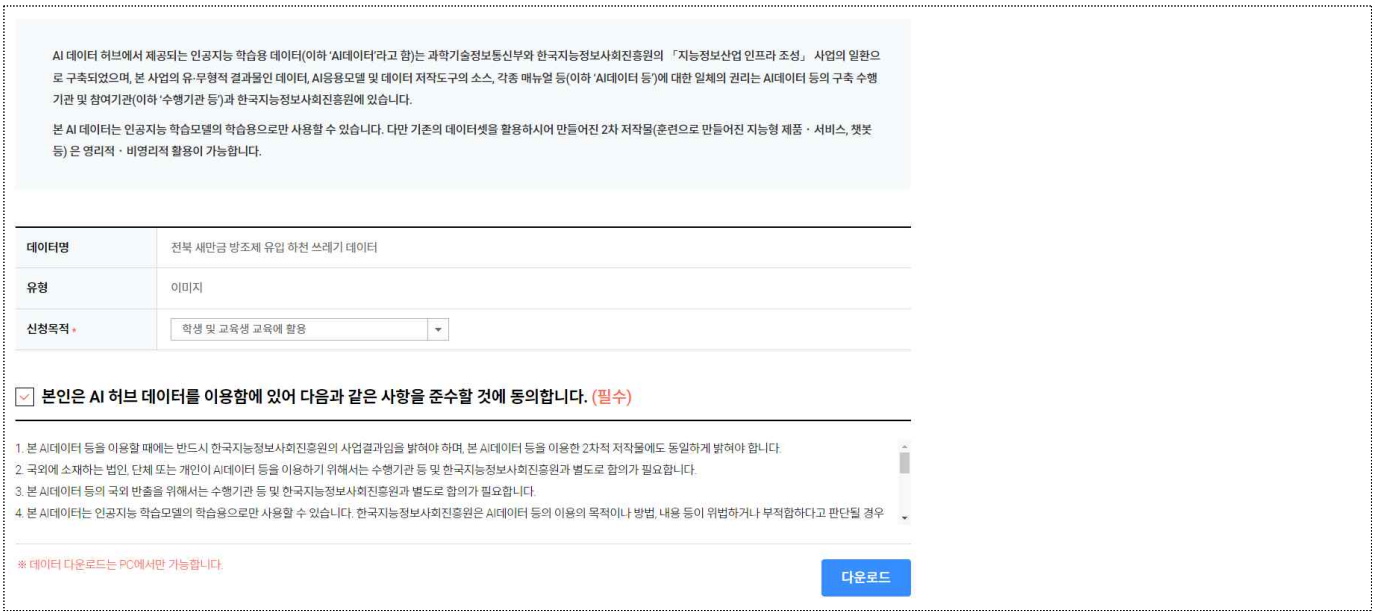
2. 데이터 통신 보안

3. 네트워크 속도

1. aihubshell 이용방법

1. aihubshell 을 다운로드 합니다.
1-1. WEB에서 다운로드 받는 방법
1-2. 서버에서 cli를 통한 다운로드 방법
2. 원하는 데이터셋에 대하여 조회하여 datasetkey값을 확인합니다.
2-1. 데이터셋 조회 및 grep 명령어를 활용한 조회
2-2. 데이터셋 파일구조 조회 및 filekey 획득방법
3. aihubshell 을 이용하여 원하는 데이터셋을 다운로드 합니다.
3-1. 데이터셋 전체 다운로드
3-2. 데이터셋 파일 개별 다운로드, 멀티 다운로드
4. 원하는 데이터패키지에 대하여 조회하여 datapckagekey값을 확인합니다.
4-1. 데이터패키지 조회 및 grep 명령어를 활용한 조회
4-2. 데이터패키지 파일구조 조회 및 filekey 획득방법
5. aihubshell 을 이용하여 원하는 데이터셋을 다운로드 합니다.
5-1. 데이터패키지 전체 다운로드
5-2. 데이터패키지 파일 개별 다운로드, 멀티 다운로드

※ 홈페이지를 통해 데이터셋에 대한 다운로드 사용 권한을 승인받습니다.



[그림1] 데이터셋 다운로드 신청

AI 허브 홈페이지 > AI 데이터 찾기 > AI 허브 오픈 API에서 확인이 가능합니다.

- **aihubshell 다운로드 및 환경준비**

- aihubshell 다운로드
aihubshell은 aihub의 학습용 데이터를 cli 환경에서 다운로드할 수 있는 도구입니다.
aihubshell은 cli 환경에서 curl 명령어로 다운로드 받을 수 있습니다.

```
복사 aihub@DESKTOP:~$ curl -o "aihubshell" https://api.aihub.or.kr/api/aihubshell.do
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  7704  100  7704    0     0   218k      0  --:--:--  --:--:--  --:--:--  221k
aihub@DESKTOP:~$
aihub@DESKTOP:~$ ls -al
-rw-r--r-- 1 aihub aihub 7704 Sep 7 00:00 aihubshell
```

- aihubshell 실행권한 부여
aihubshell은 실행권한을 반드시 부여하셔야 활용이 가능합니다.

```
복사 aihub@DESKTOP:~$ chmod +x aihubshell
     aihub@DESKTOP:~$ ls -al
     -rwxr-xr-x 1 aihub aihub 7704 Sep 7 00:00 aihubshell*
```

- 전역 실행 bin 등록
aihubshell을 /usr/bin/에 복사하여 전역명령어로 등록합니다.
전역명령어로 등록하시면 어느 위치에서든 aihubshell 명령어를 사용하실 수 있습니다.

```
복사 aihub@DESKTOP:~$ sudo cp aihubshell /usr/bin
     [sudo] password for aihub:
     aihub@DESKTOP:~$
복사 aihub@DESKTOP:~$ cd /usr/bin/
복사 aihub@DESKTOP:/usr/bin$ ls -al | grep aihubshell
     -rwxr-xr-x 1 aihub aihub 7704 Sep 7 00:00 aihubshell
```

• **aihubshell 명령어 활용**
- 메뉴얼 조회
aihubshell 명령어 및 옵션 상세 정보를 조회 합니다.

```
복사 aihub@DESKTOP: $ aihubshell -help
     aihubshell에 대한 명령어 및 옵션 상세
     ENGL_CMGG KOREAN_CMGG DETAIL_CN
     -aihubid {사용자계정} | https://aihub.or.kr 에서 회원가입 후 이용하실 수 있습니다.
     ..... 생 략 .....
```

- 데이터셋 정보 조회 [list 모드]
데이터셋 정보를 조회하며 형식은 [datasetkey, 데이터명] 입니다.
grep 명령어를 사용하여 필터링 할 수 있습니다.

```
복사 aihub@DESKTOP: $ aihubshell -mode l
     ==================DataSet 목록==================
     50, AR/VR 화면정확도 향상을 위한 플렌옵틱 카메라 이미지
     51, K-Fashion 이미지
     57, 멀티모달
     59, 반려동물 구분을 위한 동물 영상
     ..... 생 략 .....
     aihub@DESKTOP: $
복사 aihub@DESKTOP: $ aihubshell -mode l | grep 이미지
     50, AR/VR 화면정확도 향상을 위한 플렌옵틱 카메라 이미지
     51, K-Fashion 이미지
     63, 상표 이미지 및 텍스트
     ..... 생 략 .....
     aihub@DESKTOP: $
복사 aihub@DESKTOP: $ aihubshell -mode l | grep 가리비
     593, 지능형 스마트양식장 통합 데이터(가리비)
```

- 데이터셋 파일 정보 조회 [list 모드 datasetkey 옵션 활용]
데이터셋 [list 모드]에서 datasetkey 옵션을 추가하여 원하는 파일 정보를 조회 합니다.
파일 정보의 형식은 폴더 구조 및 [파일명 | 용량 | filekey]입니다.

```
복사 aihub@DESKTOP: $ aihubshell -mode l -datasetkey 593
=========================================
Fetching file tree structure...
The contents are encoded in UTF-8 including Korean characters.
If the following contents are not output normally,
Please modify the character information of the OS.
=========================================
 └──111.지능형 스마트양식장 통합 데이터(가리비)
 ├──01.데이터
 │  ├──1.Training
 │  │  ├──라벨링데이터
 │  │  │  └──TL1.zip | 108 MB | 51937
 │  │  └──원천데이터
 │  │  └──TS1.zip | 11 GB | 51938
 ..... 생 략 .....
```

- 데이터셋 다운로드 [download 모드]
[download 모드]에서 datasetkey를 입력하여 AI학습용 데이터 전체를 다운로드 합니다.
명령어를 실행한 장소에서 다운로드가 진행 됩니다.
다운로드가 완료되면 자동으로 아카이빙파일 병합, 압축해제, 압축파일 제거 순으로 작업이 진행 됩니다.
위의 사항을 고려하여 다운로드 받을 데이터의 2~3배 이상의 용량을 확보하시길 바랍니다.
-aihubapikey에 특수문자가 포함되어 있는 경우 아래의 예시와 같이 '(홑따움표)로 감싸주어 처리 하시길 바랍니다.

```
aihub@DESKTOP:~$ cd temp/
복사 aihub@DESKTOP:~/temp$ aihubshell -mode d -datasetkey 593 -aihubapikey '1234F316-E56E-7F89-A4C2-57450CEB26DD'
Authentication successful.
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:--  --:--:--  --:--:--     0
100 14.1G    0 14.1G    0     0  52.5M      0 --:--:--  0:04:35  --:--:-- 56.9M
Request successful with HTTP status 200.
Download successful.
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/원천데이터/VS1.zip.part0
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/원천데이터/VS1.zip.part1073741824
..... 생 략 .....
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/1.Training/원천데이터/TS1.zip.part11811160064
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/1.Training/라벨링데이터/TL1.zip.part0
잠시 기다려 주세요 병합중 입니다.
Merging TL1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/1.Training/라벨링데이터
Merging TS1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/1.Training/원천데이터
Merging VL1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/라벨링데이터
Merging VS1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/원천데이터
병합이 완료 되었습니다.
aihub@DESKTOP:~$
aihub@DESKTOP:~$ ls -al
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 .
drwxr-x--- 1 aihub aihub 4096 Sep 7 00:00 ..
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 '111.지능형 스마트양식장 통합 데이터(가리비)'
```

- 데이터셋 파일 선택 다운로드 [download 모드 filekey 옵션 활용]
데이터셋 [download 모드]에서 filekey 옵션을 추가하면 해당하는 파일을 다운로드 합니다.
filekey는 ,(콤마)로 구분해서 멀티 다운로드 기능을 제공합니다.

```
복사 aihub@DESKTOP:~/temp$ aihubshell -mode d -datasetkey 593 -filekey 51937 -aihubapikey '1234F316-E56E-7F89-A4C2-57450CEB26DD'
Authentication successful.
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:--  --:--:--  --:--:--     0
100  107M  0  107M    0     0  43.2M      0  --:--:--  0:00:02  --:--:--  60.1M
Request successful with HTTP status 200.
Download successful.
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/1.Training/라벨링데이터/TL1.zip.part0
잠시 기다려 주세요 병합중 입니다.
Merging TL1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/1.Training/라벨링데이터
병합이 완료 되었습니다.
aihub@DESKTOP:~$
aihub@DESKTOP:~$ ls -al
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 .
drwxr-x--- 1 aihub aihub 4096 Sep 7 00:00 ..
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 '111.지능형 스마트양식장 통합 데이터(가리비)'
aihub@DESKTOP:~$
aihub@DESKTOP:~$
복사 aihub@DESKTOP:~/temp$ aihubshell -mode d -datasetkey 593 -filekey 51939,51940 -aihubapikey '1234F316-E56E-7F89-A4C2-57450CEB26DD
Authentication successful.
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:--  --:--:--  --:--:--     0
100 1449M  0 1449M    0     0  46.9M      0  --:--:--  0:00:30  --:--:--  60.1M
Request successful with HTTP status 200.
Download successful.
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/원천데이터/VS1.zip.part0
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/원천데이터/VS1.zip.part1073741824
111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/라벨링데이터/VL1.zip.part0
잠시 기다려 주세요 병합중 입니다.
Merging VL1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/라벨링데이터
Merging VS1.zip in ./111.지능형 스마트양식장 통합 데이터(가리비)/01.데이터/2.Validation/원천데이터
병합이 완료 되었습니다.
aihub@DESKTOP:~$
aihub@DESKTOP:~$ ls -al
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 .
drwxr-x--- 1 aihub aihub 4096 Sep 7 00:00 ..
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 '111.지능형 스마트양식장 통합 데이터(가리비)'
```

- 데이터패키지 정보 조회 [list 모드]
데이터패키지 정보를 조회하며 형식은 [datapckagekey, 데이터패키지명] 입니다.
grep 명령어를 사용하여 필터링 할 수 있습니다.

```
복사 aihub@DESKTOP:~$ aihubshell -mode pl
==================데이터패키지 목록==================
1, 멀티세션, 대화 관련 말뭉치 데이터
2, 금융, 법률 등 전문분야 말뭉치
3, 요약, 기계독해를 위한 말뭉치 데이터
.......... 생 략 ..............
aihub@DESKTOP:~$
복사 aihub@DESKTOP:~$ aihubshell -mode pl | grep 데이터
1, 멀티세션, 대화 관련 말뭉치 데이터
2, 금융, 법률 등 전문분야 말뭉치
3, 요약, 기계독해를 위한 말뭉치 데이터
.......... 생 략 ..............
```

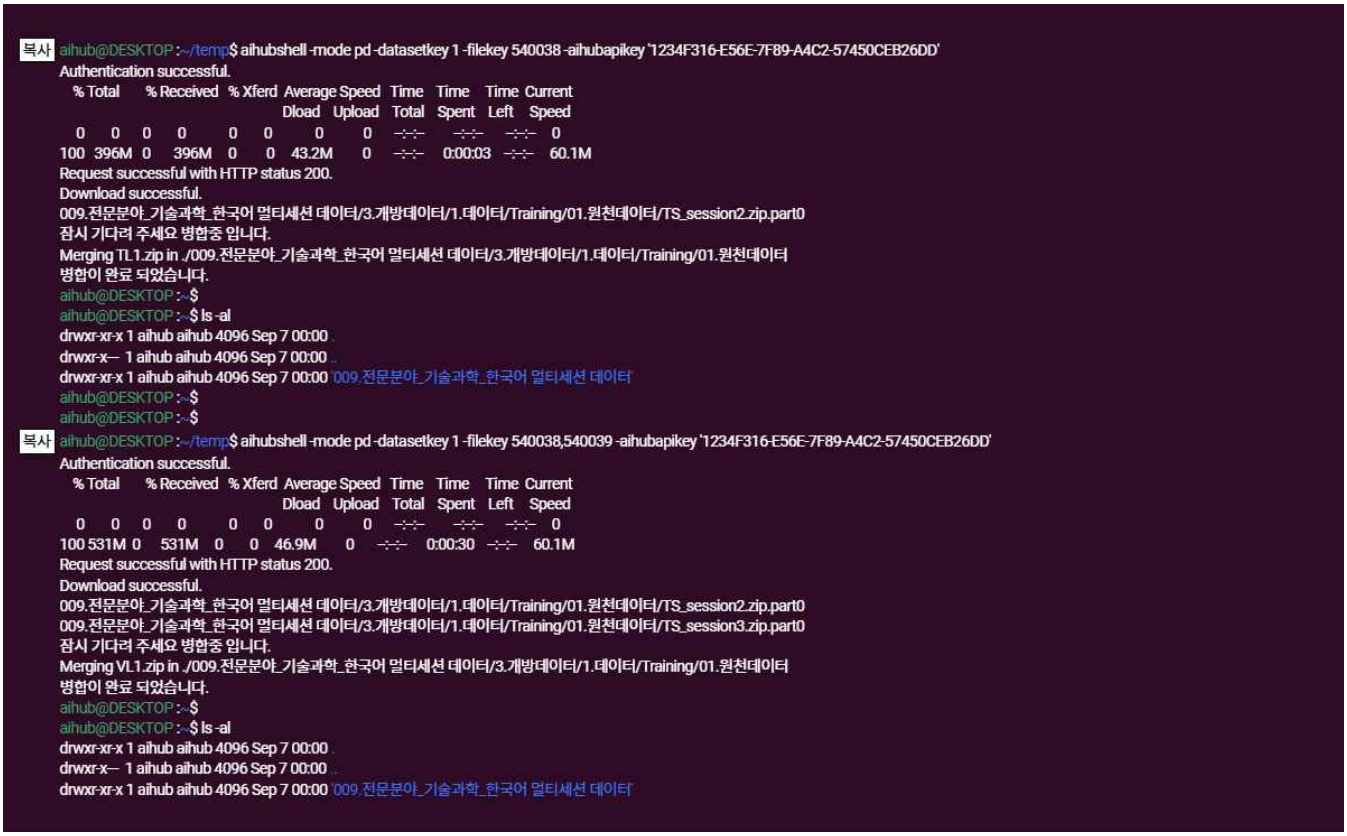
- 데이터패키지 파일 정보 조회 [list 모드 datapckagekey 옵션 활용]
데이터패키지 [list 모드]에서 datapckagekey 옵션을 추가하여 원하는 파일 정보를 조회 합니다.
파일 정보의 형식은 폴더 구조 및 [파일명 | 용량 | filekey]입니다.

```
복사 aihub@DESKTOP:~$ aihubshell -mode pl -datapckagekey 1
=========================================
Fetching file tree structure....
The contents are encoded in UTF-8 including Korean characters.
If the following contents are not output normally,
Please modify the character information of the OS.
=========================================
 └──009.전문분야_기술과학_한국어 멀티세션 데이터
 ├──3.개방데이터
 |  ├──1.데이터
..... 생 략 .....
 └──010.전문분야_사회과학_한국어 멀티세션 데이터
..... 생 략 .....
```

- 데이터패키지 다운로드 [download 모드]
데이터패키지 [download 모드]에서 datapckagekey를 입력하여 AI학습용 데이터 전체를 다운로드 합니다.
명령어를 실행한 장소에서 다운로드가 진행 됩니다.
다운로드가 완료되면 자동으로 아카이빙파일 병합, 압축해제, 압축파일 제거 순으로 작업이 진행 됩니다.
위의 사항을 고려하여 다운로드 받을 데이터의 2~3배 이상의 용량을 확보하시길 바랍니다.
-aihubapikey에 특수문자가 포함되어 있는 경우 아래의 예시와 같이 '(홑따움표)로 감싸주어 처리 하시길 바랍니다.

```
aihub@DESKTOP:~$ cd temp/
복사 aihub@DESKTOP:~/temp$ aihubshell -mode pd -datapckagekey 1 -aihubapikey '1234F316-E56E-7F89-A4C2-57450CEB26DD'
Authentication successful.
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:--  --:--:--  --:--:--     0
100 20.12G  0 20.12G    0     0  52.5M      0  --:--:--  0:10:35  --:--:--  56.9M
Request successful with HTTP status 200.
Download successful.
009.전문분야_기술과학_한국어 멀티세션 데이터/3.개방데이터/1.데이터/Training/01.원천데이터/TS_session2.zip.part0
009.전문분야_기술과학_한국어 멀티세션 데이터/3.개방데이터/1.데이터/Training/01.원천데이터/TS_session3.zip.part0
..... 생 략 .....
잠시 기다려 주세요 병합중 입니다.
Merging VL1.zip in ./009.전문분야_기술과학_한국어 멀티세션 데이터/3.개방데이터/1.데이터/Training/01.원천데이터
..... 생 략 .....
병합이 완료 되었습니다.
aihub@DESKTOP:~$
aihub@DESKTOP:~$ ls -al
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 .
drwxr-x--- 1 aihub aihub 4096 Sep 7 00:00 ..
drwxr-xr-x 1 aihub aihub 4096 Sep 7 00:00 '009.전문분야_기술과학_한국어 멀티세션 데이터'
```

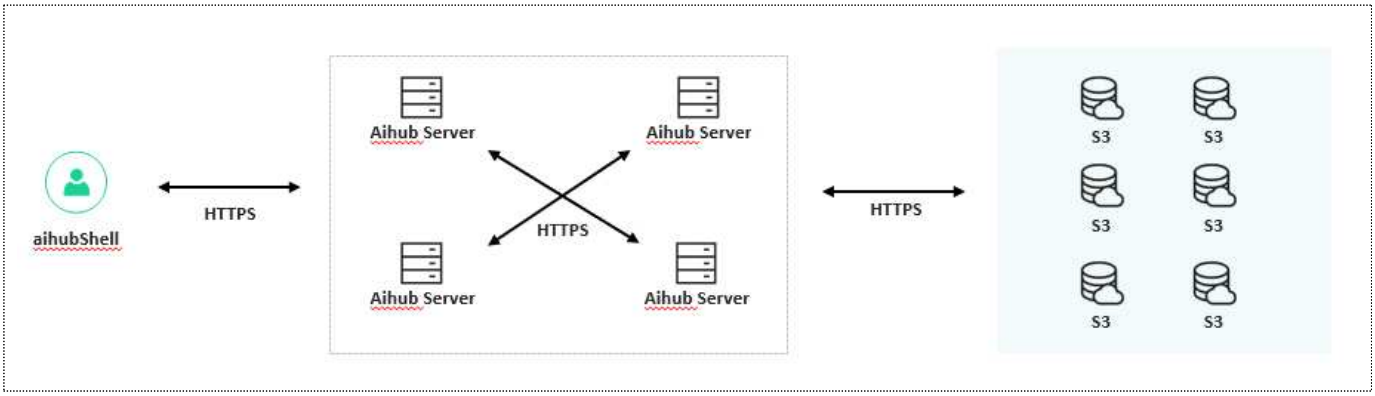
- 데이터패키지 파일 선택 다운로드 [download 모드 filekey 옵션 활용]
데이터패키지 [download 모드]에서 filekey 옵션을 추가하면 해당하는 파일을 다운로드 합니다.
filekey는 ,(콤마)로 구분해서 멀티 다운로드 기능을 제공합니다.



[그림2] cli 사용화면

2. 데이터 통신 보안

오픈API 서비스의 모든 통신은 https 를 기반하여 암호화된 통신으로 이루어 집니다.
aihubshell 을 이용하여 중계 서비스와 통신하고 사용승인이 유효한 요청에 대하여 데이터셋을 스트리밍 서비스로 제공합니다.



[그림3] 사용자-중계서버-스트리밍서버-데이터셋 S3

3. 네트워크 속도

속도관련 테스트는 AWS에서 제공하는 S3 cli 및 대용량 파일 다운로드 웹솔루션 이노릭스를 속도의 기준으로 합니다.
https를 기반으로 하는 서비스임으로 이용자의 네트워크 상태에 따라 속도의 차이가 발생합니다.
네트워크 속도는 NIA에서 제공하는 속도측정 플렛폼을 이용하시길 바랍니다.

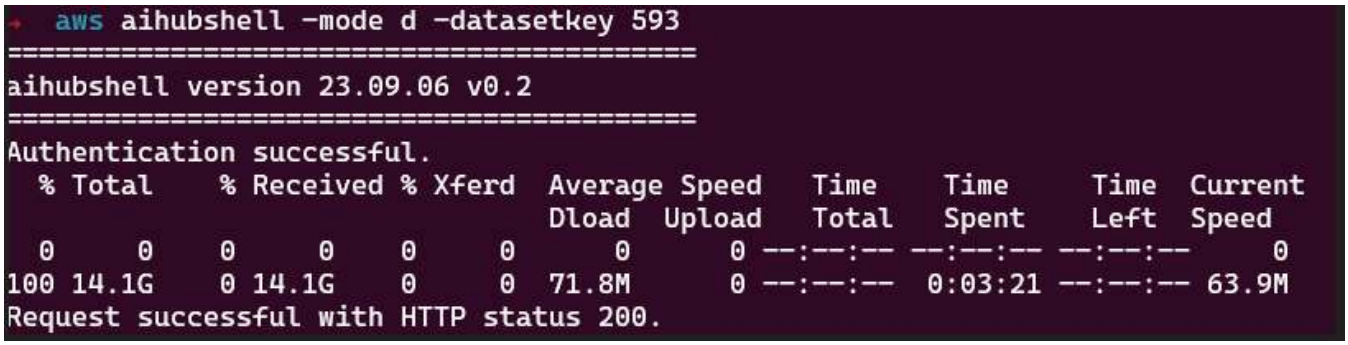
- 체감속도 네트워크 6~700Mbps일 경우 100GB 파일 다운로드 30분 소요 예상
- S3 cli / 이노릭스 / 오픈API 소요 확인

NIA 속도측정 플렛폼
https://speed.nia.or.kr/index.asp

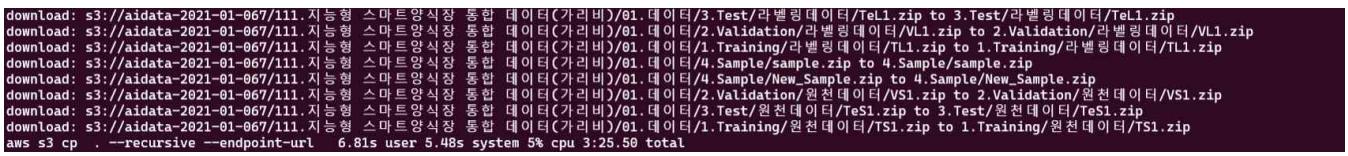
[그림4] NIA 속도측정 플렛폼

이노릭스 다운로드 – AI 허브 홈페이지

aihubshell 다운로드



S3 cli 다운로드



[그림5] 이노릭스 다운로드, aihubshell 다운로드, S3 cli 다운로드