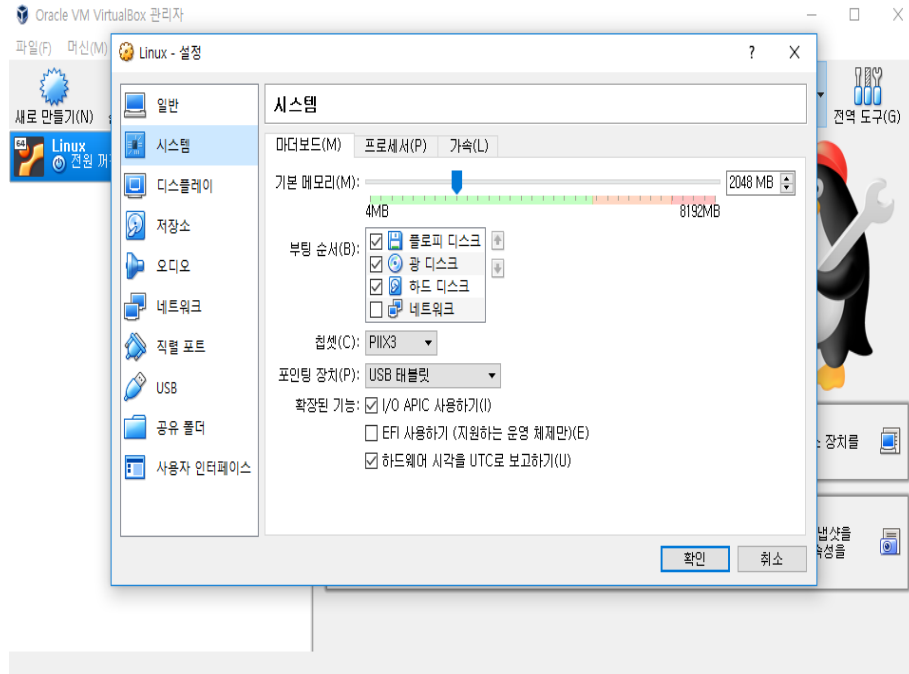


수집 시스템 구축하기			
작성자	송무섭	작성일	18.12.06

1. 수집 시스템 구축하기

● 가상머신 만들기



● Scrapy 설치하기

1. Python-pip 설치

```
l@l-VirtualBox:~$ sudo apt-get install python-pip
```

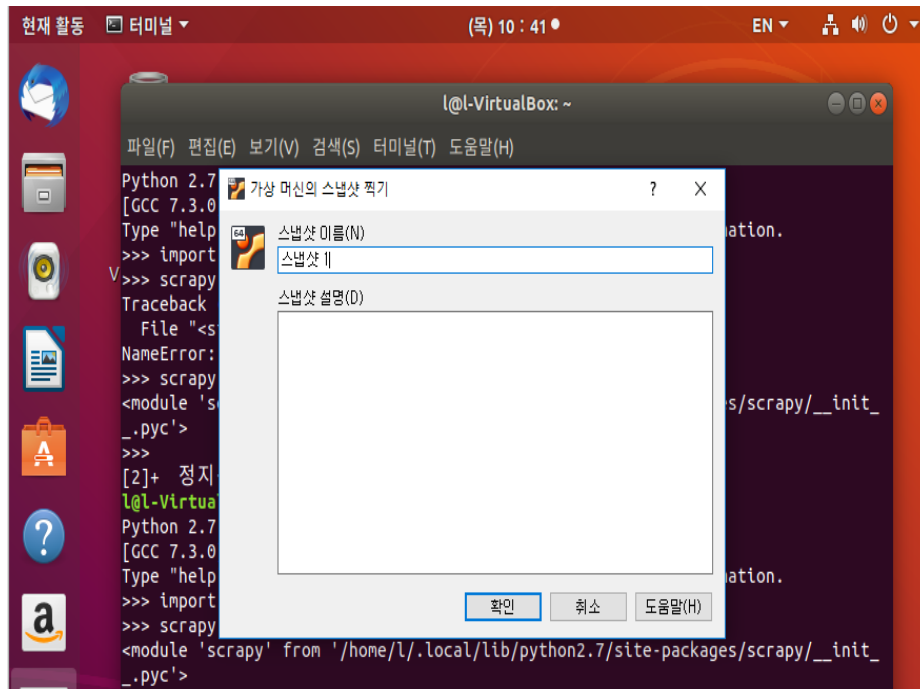
2. Scrapy 설치

```
l@l-VirtualBox:~$ pip install scrapy
```

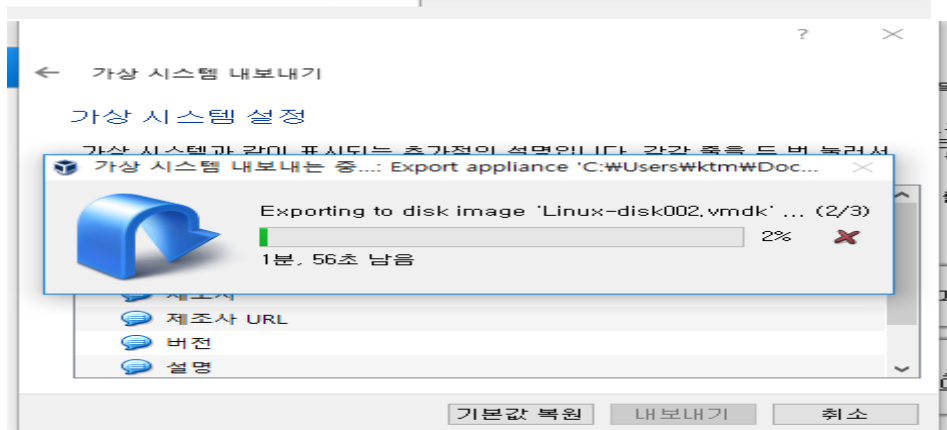
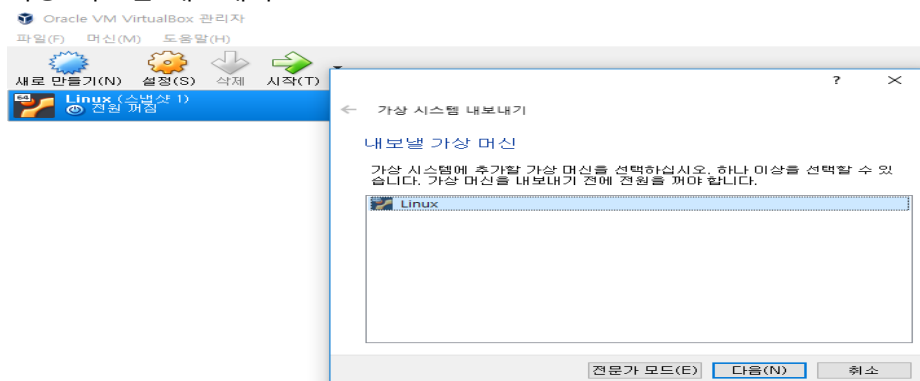
3. 확인

```
l@l-VirtualBox:~$ python
Python 2.7.15rc1 (default, Nov 12 2018, 14:31:15)
[GCC 7.3.0] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import scrapy
>>> scrapy
<module 'scrapy' from '/home/l/.local/lib/python2.7/site-packages/scrapy/__init_
.pyc'>
```

- 스냅샷 만들기



- 가상 시스템 내보내기



2. MySQL에서 데이터 가져오기

- Python에서 MySQL 연결하기

```
In [1]: import mysql.connector
...:
...: mydb = mysql.connector.connect(
...:     host='localhost',
...:     user = 'root',
...:     passwd='qwer1234',
...:     database = 'mydatabase'
...: )
...: print(mydb)
<mysql.connector.connection.MySQLConnection object at 0x00000188EEE7BA90>
```

- Table 확인하기

```
In [2]: mycursor = mydb.cursor()
...:
...: mycursor.execute("SHOW TABLES")
...: for x in mycursor:
...:     print(x)
...:
...:
('customers',)
('customers1',)
('tableb',)
```

- 데이터 수집

```
mycursor.execute("SELECT * FROM tableb")
myresult = mycursor.fetchall()
for x in myresult:
    print(x)

('LKBP-065 무지백팩 + 파우치 학생 가방 캐주얼 백팩 ', '25,800', '', '더블유팝')
('겨울시즌 캐리어 여행가방 세트+파우치 증정 ', '32,900', '109,600', '트래블메이트')
('샤오미 하이엔드 백팩 ', '65,610', '69,800', '가온미')
('뉴엔 N73 테리 캐리어 노트북 백팩 ', '26,500', '26,600', '월코')
('BLMG 프리미엄 하이스픽 여행가방 캐리어 ', '32,900', '74,000', '최우수판매자')
('여행가방 여행용가방 보스턴 보스턴 더스위츠 ', '24,900', '', '더스위츠가방')
```

- 데이터 확인하기

CSV 파일

805	자전거 짐받이 가방 안장가방 캐리어 렉짐받이 거치대	22,900		이세이브
806	천연가죽 클러치백/가죽 가방/리얼 레더 클러치	35,400	35,500	미스안나
807	여성 가죽백팩 여자가방 대학생 직장인백팩 미니백팩	21,800		미틈몰
808	여행가방 여행용보스턴백 기내용캐리어 캐리어 여행용	11,800		레브REVE
809	트레빌 스파덱스 트레빌 캐리어커버 S~L (18형~28형)	14,900		트레빌

테이블 값

```
('자전거 짐받이 가방 안장가방 캐리어 렉짐받이 거치대 ', '22,900', '', '이세이브')
('천연가죽 클러치백/가죽 가방/리얼 레더 클러치 ', '35,400', '35,500', '미스안나')
('여성 가죽백팩 여자가방 대학생 직장인백팩 미니백팩 ', '21,800', '', '미틈몰')
('여행가방 여행용보스턴백 기내용캐리어 캐리어 여행용 ', '11,800', '', '레브REVE')
('트레빌 스파덱스 트레빌 캐리어커버 S~L (18형~28형) ', '14,900', '', '트레빌')
```

값이 같은 것을 알 수 있다.

- Crontab 사용하기(10분마다 사용)

```
GNU nano 2.9.3 /tmp/crontab.5FARjA/crontab 변경됨.
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h dom mon dow   command

*/10 * * * * /usr/bin/python3 /home/l/hello.py >>echo.log

저장할 파일 이름: /tmp/crontab.5FARjA/crontab
l@l-VirtualBox:~$ cat echo.log
v
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
Hello
```

로그에 10분마다 사용한 흔적을 볼 수 있다.