This is a readme file introducing how to deploy this Scrapy web spider to a machine (in our project, we deploy this spider to an AWS server).

1. Python 3.5.2 on Ubuntu system;
2. pip install virtualenv;
3. workon [virtual environment name: scrapy];
4. pip install scrapy;
5. pip install openpyxl;
6. after installing the above two modules, our pip list should be:

asn1crypto (0.24.0)	attrs (17.4.0)	Automat (0.6.0)	beautifulsoup4 (4.6.3)
certifi (2018.11.29)	cffi (1.11.5)	chardet (3.0.4)	constantly (15.1.0)
cryptography (2.2.1)	cssselect (1.0.3)	et-xmlfile (1.0.1)	hyperlink (18.0.0)
idna (2.6)	incremental (17.5.0)	jdcal (1.3)	lxml (4.2.1)
openpyxl (2.5.1)	parsel (1.4.0)	pip (9.0.3)	pyasn1 (0.4.2)
pyasn1-modules (0.2.1)	pycparser (2.18)	PyDispatcher (2.0.5)	pyOpenSSL (17.5.0)
queuelib (1.5.0)	requests (2.20.1)	Scrapy (1.5.0)	service-identity (17.0.0)
setuptools (39.0.1)	six (1.11.0)	Twisted (17.9.0)	urllib3 (1.24.1)
w3lib (1.19.0)	wheel (0.30.0)	zope.interface (4.4.3)	

7. scrapy crawl twitchdb.

The step 7 will start a session of crawling and insert the new information to the Excel files in ".//database".

To make a crontab schedule on Ubuntu, we need to edit the crontab files to run the script "python .//main.py".

Ref. https://askubuntu.com/questions/2368/how-do-i-set-up-a-cron-job