

## P3 Kafka-python on CentOS

### How to install PIP and kafka-python on CentOS VM

<http://sharadchhetri.com/2014/05/30/install-pip-centos-rhel-ubuntu-debian/>

```
[cloudera@localhost ~]$ sudo rpm -ivh http://dl.fedoraproject.org/pub/epel/6/x86\_64/epel-release-6-8.noarch.rpm
Retrieving http://dl.fedoraproject.org/pub/epel/6/x86\_64/epel-release-6-8.noarch.rpm
warning: /var/tmp/rpm-tmp.OPsdAy: Header V3 RSA/SHA256 Signature, key ID 0608b895: NOKEY
Preparing... ##### [100%]
 1:epel-release ##### [100%]
[cloudera@localhost ~]$
```

```
[cloudera@localhost ~]$ sudo yum install -y python-pip
Loaded plugins: fastestmirror, refresh-packagekit, security
Setting up Install Process
Loading mirror speeds from cached hostfile
epel/metalink | 12 kB 00:00
* base: centos.gol.com
* epel: mirror.symnds.com
* extras: mirror.raystedman.net
* updates: mirrors.greenmountainaccess.net
epel | 4.3 kB 00:00
epel/primary_db | 5.8 MB 00:00
Resolving Dependencies
--> Running transaction check
--> Package python-pip.noarch 0:7.1.0-1.el6 will be installed
--> Processing Dependency: python-setuptools for package: python-pip-7.1.0-1.el6.noarch
--> Running transaction check
--> Package python-setuptools.noarch 0:0.6.10-3.el6 will be installed
--> Finished Dependency Resolution
Dependencies Resolved
```

```
=====
Package      Arch      Version      Repository      Size
=====
Installing:
python-pip    noarch    7.1.0-1.el6   epel            1.5 M
Installing for dependencies:
python-setuptools noarch    0.6.10-3.el6   base           336 k
Transaction Summary
```

```
=====
Install      2 Package(s)
Total download size: 1.9 M
Installed size: 8.1 M
Downloading Packages:
(1/2): python-pip-7.1.0-1.el6.noarch.rpm | 1.5 MB 00:00
(2/2): python-setuptools-0.6.10-3.el6.noarch.rpm | 336 kB 00:00
=====
```

```
Total      3.3 MB/s | 1.9 MB 00:00
warning: rpmts_HdrFromFdno: Header V3 RSA/SHA256 Signature, key ID 0608b895: NOKEY
Retrieving key from file:///etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-6
Importing GPG key 0x0608B895:
 Userid : EPEL (6) <epel@fedoraproject.org>
 Package: epel-release-6-8.noarch (installed)
 From    : /etc/pki/rpm-gpg/RPM-GPG-KEY-EPEL-6
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
Warning: RPMDB altered outside of yum.
Installing : python-setuptools-0.6.10-3.el6.noarch      1/2
Installing : python-pip-7.1.0-1.el6.noarch             2/2
Verifying  : python-pip-7.1.0-1.el6.noarch             1/2
Verifying  : python-setuptools-0.6.10-3.el6.noarch     2/2
Installed:
python-pip.noarch 0:7.1.0-1.el6
Dependency Installed:
python-setuptools.noarch 0:0.6.10-3.el6
Complete!
```

```
[cloudera@localhost ~]$
```

```
[cloudera@localhost ~]$ pip -V
pip 7.1.0 from /usr/lib/python2.6/site-packages (python 2.6)
```

```
[cloudera@localhost ~]$ pip list
/usr/lib/python2.6/site-packages/pip/_vendor/requests/packages/urllib3/util/ssl_.py:90: InsecurePlatformWarning: A true SSLContext object is not available. This prevents urllib3,
InsecurePlatformWarning
You are using pip version 7.1.0, however version 8.1.0 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
Beaker (1.3.1)

cups (1.0)
...
[cloudera@localhost ~]$
```

Lets upgrade PIP as we are asked to:

```
[cloudera@localhost ~]$ sudo pip install --upgrade pip
/usr/lib/python2.6/site-packages/pip/_vendor/requests/packages/urllib3/util/ssl_.py:90: InsecurePlatformWarning: A true SSLContext object is not available. This prevents urllib3,
InsecurePlatformWarning
You are using pip version 7.1.0, however version 8.1.0 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
/usr/lib/python2.6/site-packages/pip/_vendor/requests/packages/urllib3/util/ssl_.py:90: InsecurePlatformWarning: A true SSLContext object is not available. This prevents urllib3,
InsecurePlatformWarning
Collecting pip
  Downloading pip-8.1.0-py2.py3-none-any.whl (1.2MB)
    100% |#####| 1.2MB 383kB/s
Installing collected packages: pip
  Found existing installation: pip 7.1.0
    Uninstalling pip-7.1.0:
      Successfully uninstalled pip-7.1.0
Successfully installed pip-8.1.0
```

```
[cloudera@localhost ~]$ pip -V
pip 8.1.0 from /usr/lib/python2.6/site-packages (python 2.6)
[cloudera@localhost ~]$
```

However, 'pip install pykafka' fails - with many errors

**Looking closely at the Docs for PyKafka - it says it works on Python 2.7+ only...**  
**Have to upgrade Python to 2.7 OR - use 'kafka-python' lib which is supported for Python 2.6:**

Let use kafka-python on this VM:

<https://github.com/dpkp/kafka-python>

```
[cloudera@localhost ~]$ sudo pip install kafka-python
DEPRECATION: Python 2.6 is no longer supported by the Python core team, please upgrade your Python. A future version of pip will drop support for Python 2.6
Collecting kafka-python
/usr/lib/python2.6/site-
packages/pip/_vendor/requests/packages/urllib3/util/ssl_.py:315: SNIMissingWarning: An HTTPS request has been made, but the SNI (Subject Name Indication) extension to TLS
...
Successfully installed kafka-python-1.0.1 six-1.10.0
[cloudera@localhost ~]$
```

**Run a simple Python producer with Kafka console consumer:**

kafka-python producer:

```
[cloudera@localhost ~]$ python
Python 2.6.6 (r266:84292, Jul 23 2015, 15:22:56)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-11)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from kafka import KafkaProducer
>>> producer = KafkaProducer(bootstrap_servers='localhost:9092')
>>> producer.send('spark_topic', b'message 1')
<kafka.producer.future.FutureRecordMetadata object at 0xc19450>
>>> producer.send('spark_topic', b'message 2')
<kafka.producer.future.FutureRecordMetadata object at 0xc19710>
>>>
```

```
[cloudera@localhost ~]$ /home/cloudera/Marina/kafka_2.11-0.8.2.2/bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic spark_topic
message 1
message 2
```

The screenshot shows a terminal window with a dark background and a light blue title bar. The window title is "kafka-python producer". The terminal content shows the following steps:

```
File Edit View Search Terminal Help
Downloading kafka_python-1.0.1-py2-none-any.whl (137kB)
100% |████████████████████████████████████████| 143kB 2.8MB/s
Collecting six (from kafka-python)
Downloading six-1.10.0-py2.py3-none-any.whl
Installing collected packages: six, kafka-python
Successfully installed kafka-python-1.0.1 six-1.10.0
[cloadera@localhost ~]$ python
Python 2.6.6 (r266:84292, Jul 23 2015, 15:22:56)
[GCC 4.4.7 20120313 (Red Hat 4.4.7.1)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from kafka import KafkaProducer
>>> producer = KafkaProducer(bootstrap_servers='localhost:9092')
>>> producer.send('spark topic', b'message 1')
<kafka.producer.future.FutureRecordMetadata object at 0xc19458>
>>> producer.send('spark topic', b'message 2')
<kafka.producer.future.FutureRecordMetadata object at 0xc19710>
>>>
```

Below the terminal window, there is a "Kafka Console tools" window with a similar title bar. It shows the output of the `kafka-topics.sh` command:

```
File Edit View Search Terminal Help
[cloadera@localhost ~]$ /home/cloadera/Marina/kafka_2.11-0.8.2/bin/kafka-topics.sh --describe -zookeeper localhost:2181
Topic:spark_topic PartitionCount:4 ReplicationFactor:1 Configs:
sr: 0 Topic: spark_topic Partition: 0 Leader: 0 Replicas: 0 I
sr: 0 Topic: spark_topic Partition: 1 Leader: 0 Replicas: 0 I
sr: 0 Topic: spark_topic Partition: 2 Leader: 0 Replicas: 0 I
sr: 0 Topic: spark_topic Partition: 3 Leader: 0 Replicas: 0 I
sr: 0
[cloadera@localhost ~]$ /home/cloadera/Marina/kafka_2.11-0.8.2/bin/kafka-conso
te-consumer.sh --zookeeper localhost:2181 --topic spark_topic
message 1
message 2
```