

Installation guide

Supported Python versions

Scrapy requires Python 3.6+, either the CPython implementation (default) or the PyPy 7.2.0+ implementation (see [ref`pythonimplementations`](#)).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 12); [backlink](#)

Unknown interpreted text role "ref".

Installing Scrapy

If you're using [Anaconda](#) or [Miniconda](#), you can install the package from the [conda-forge](#) channel, which has up-to-date packages for Linux, Windows and macOS.

To install Scrapy using `conda`, run:

```
conda install -c conda-forge scrapy
```

Alternatively, if you're already familiar with installation of Python packages, you can install Scrapy and its dependencies from PyPI with:

```
pip install Scrapy
```

We strongly recommend that you install Scrapy in [ref`a dedicated virtualenv <intro-using-virtualenv>](#), to avoid conflicting with your system packages.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 33); [backlink](#)

Unknown interpreted text role "ref".

Note that sometimes this may require solving compilation issues for some Scrapy dependencies depending on your operating system, so be sure to check the [ref`intro-install-platform-notes`](#).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 36); [backlink](#)

Unknown interpreted text role "ref".

For more detailed and platform specifics instructions, as well as troubleshooting information, read on.

Things that are good to know

Scrapy is written in pure Python and depends on a few key Python packages (among others):

- [lxml](#), an efficient XML and HTML parser
- [parsel](#), an HTML/XML data extraction library written on top of `lxml`,
- [w3lib](#), a multi-purpose helper for dealing with URLs and web page encodings
- [twisted](#), an asynchronous networking framework
- [cryptography](#) and [pyOpenSSL](#), to deal with various network-level security needs

The minimal versions which Scrapy is tested against are:

- Twisted 14.0
- `lxml` 3.4
- `pyOpenSSL` 0.14

Scrapy may work with older versions of these packages but it is not guaranteed it will continue working because it's not being tested against them.

Some of these packages themselves depends on non-Python packages that might require additional installation steps depending on your platform. Please check [ref`platform-specific guides below <intro-install-platform-notes>](#).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 65); [backlink](#)

Unknown interpreted text role "ref".

In case of any trouble related to these dependencies, please refer to their respective installation instructions:

- [lxml installation](#)

- `:doc: cryptography installation <cryptography:installation>`

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 73); [backlink](#)

Unknown interpreted text role "doc".

Using a virtual environment (recommended)

TL;DR: We recommend installing Scrapy inside a virtual environment on all platforms.

Python packages can be installed either globally (a.k.a system wide), or in user-space. We do not recommend installing Scrapy system wide.

Instead, we recommend that you install Scrapy within a so-called "virtual environment" (`:mod:venv`). Virtual environments allow you to not conflict with already-installed Python system packages (which could break some of your system tools and scripts), and still install packages normally with `pip` (without `sudo` and the likes).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 89); [backlink](#)

Unknown interpreted text role "mod".

See `ref:tut-venv` on how to create your virtual environment.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 95); [backlink](#)

Unknown interpreted text role "ref".

Once you have created a virtual environment, you can install Scrapy inside it with `pip`, just like any other Python package. (See `ref:platform-specific guides <intro-install-platform-notes>` below for non-Python dependencies that you may need to install beforehand).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 97); [backlink](#)

Unknown interpreted text role "ref".

Platform specific installation notes

Windows

Though it's possible to install Scrapy on Windows using `pip`, we recommend you to install [Anaconda](#) or [Miniconda](#) and use the package from the [conda-forge](#) channel, which will avoid most installation issues.

Once you've installed [Anaconda](#) or [Miniconda](#), install Scrapy with:

```
conda install -c conda-forge scrapy
```

To install Scrapy on Windows using `pip`:

Warning

This installation method requires "Microsoft Visual C++" for installing some Scrapy dependencies, which demands significantly more disk space than Anaconda.

1. Download and execute [Microsoft C++ Build Tools](#) to install the Visual Studio Installer.
2. Run the Visual Studio Installer.
3. Under the Workloads section, select **C++ build tools**.
4. Check the installation details and make sure following packages are selected as optional components:
 - **MSVC** (e.g MSVC v142 - VS 2019 C++ x64/x86 build tools (v14.23))
 - **Windows SDK** (e.g Windows 10 SDK (10.0.18362.0))
5. Install the Visual Studio Build Tools.

Now, you should be able to `ref:install Scrapy <intro-install-scrapy>` using `pip`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 141); [backlink](#)

Unknown interpreted text role "ref".

Ubuntu 14.04 or above

Scrapy is currently tested with recent-enough versions of lxml, twisted and pyOpenSSL, and is compatible with recent Ubuntu distributions. But it should support older versions of Ubuntu too, like Ubuntu 14.04, albeit with potential issues with TLS connections.

Don't use the `python-scrapy` package provided by Ubuntu, they are typically too old and slow to catch up with latest Scrapy.

To install Scrapy on Ubuntu (or Ubuntu-based) systems, you need to install these dependencies:

```
sudo apt-get install python3 python3-dev python3-pip libxml2-dev libxslt1-dev zlib1g-dev libffi-dev libssl-
```

- `python3-dev`, `zlib1g-dev`, `libxml2-dev` and `libxslt1-dev` are required for `lxml`
- `libssl-dev` and `libffi-dev` are required for cryptography

Inside a `ref`virtualenv <intro-using-virtualenv>``, you can install Scrapy with `pip` after that:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 166); [backlink](#)

Unknown interpreted text role "ref".

```
pip install scrapy
```

Note

The same non-Python dependencies can be used to install Scrapy in Debian Jessie (8.0) and above.

macOS

Building Scrapy's dependencies requires the presence of a C compiler and development headers. On macOS this is typically provided by Apple's Xcode development tools. To install the Xcode command line tools open a terminal window and run:

```
xcode-select --install
```

There's a [known issue](#) that prevents `pip` from updating system packages. This has to be addressed to successfully install Scrapy and its dependencies. Here are some proposed solutions:

- *(Recommended)* **Don't** use system Python. Install a new, updated version that doesn't conflict with the rest of your system. Here's how to do it using the [homebrew](#) package manager:

- Install [homebrew](#) following the instructions in <https://brew.sh/>
- Update your `PATH` variable to state that homebrew packages should be used before system packages (Change `.bashrc` to `.zshrc` accordingly if you're using [zsh](#) as default shell):

```
echo "export PATH=/usr/local/bin:/usr/local/sbin:$PATH" >> ~/.bashrc
```

- Reload `.bashrc` to ensure the changes have taken place:

```
source ~/.bashrc
```

- Install python:

```
brew install python
```

- Latest versions of python have `pip` bundled with them so you won't need to install it separately. If this is not the case, upgrade python:

```
brew update; brew upgrade python
```

- *(Optional)* `ref`Install Scrapy inside a Python virtual environment <intro-using-virtualenv>``.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scrapy-master\docs\intro\[scrapy-master] [docs] [intro]install.rst, line 218); [backlink](#)

Unknown interpreted text role "ref".

This method is a workaround for the above macOS issue, but it's an overall good practice for managing dependencies and can complement the first method.

After any of these workarounds you should be able to install Scrapy:

```
pip install Scrapy
```

PyPy

We recommend using the latest PyPy version. The version tested is 5.9.0. For PyPy3, only Linux installation was tested.

Most Scrapy dependencies now have binary wheels for CPython, but not for PyPy. This means that these dependencies will be built during installation. On macOS, you are likely to face an issue with building the Cryptography dependency. The solution to this problem is described [here](#), that is to `brew install openssl` and then export the flags that this command recommends (only needed when installing Scrapy). Installing on Linux has no special issues besides installing build dependencies. Installing Scrapy with PyPy on Windows is not tested.

You can check that Scrapy is installed correctly by running `scrapy bench`. If this command gives errors such as `TypeError: ... got 2 unexpected keyword arguments`, this means that `setuptools` was unable to pick up one PyPy-specific dependency. To fix this issue, run `pip install 'PyPyDispatcher>=2.1.0'`.

Troubleshooting

AttributeError: 'module' object has no attribute 'OP_NO_TLSv1_1'

After you install or upgrade Scrapy, Twisted or pyOpenSSL, you may get an exception with the following traceback:

```
[...]
File "[...]/site-packages/twisted/protocols/tls.py", line 63, in <module>
    from twisted.internet._sslverify import _setAcceptableProtocols
File "[...]/site-packages/twisted/internet/_sslverify.py", line 38, in <module>
    TLSVersion.TLSv1_1: SSL.OP_NO_TLSv1_1,
AttributeError: 'module' object has no attribute 'OP_NO_TLSv1_1'
```

The reason you get this exception is that your system or virtual environment has a version of pyOpenSSL that your version of Twisted does not support.

To install a version of pyOpenSSL that your version of Twisted supports, reinstall Twisted with the `tls` extra option:

```
pip install twisted[tls]
```

For details, see [Issue #2473](#).