

Computation

Depen- dency	Minimum Ver- sion	Notes
SciPy	1.4.1	Miscellaneous statistical functions
numba	0.50.1	Alternative execution engine for rolling operations (see <i>Enhancing Performance</i>)
xarray	0.15.1	pandas-like API for N-dimensional data

Excel files

Dependency	Minimum Version	Notes
xlrd	2.0.1	Reading Excel
xlwt	1.3.0	Writing Excel
xlsxwriter	1.2.2	Writing Excel
openpyxl	3.0.3	Reading / writing for xlsx files
pyxlsb	1.0.6	Reading for xlsb files

HTML

Dependency	Minimum Version	Notes
BeautifulSoup4	4.8.2	HTML parser for read_html
html5lib	1.1	HTML parser for read_html
lxml	4.5.0	HTML parser for read_html

One of the following combinations of libraries is needed to use the top-level `read_html()` function:

- BeautifulSoup4 and html5lib
- BeautifulSoup4 and lxml
- BeautifulSoup4 and html5lib and lxml
- Only lxml, although see *HTML Table Parsing* for reasons as to why you should probably **not** take this approach.

Warning:

- if you install BeautifulSoup4 you must install either lxml or html5lib or both. `read_html()` will **not** work with *only* BeautifulSoup4 installed.
- You are highly encouraged to read *HTML Table Parsing gotchas*. It explains issues surrounding the installation and usage of the above three libraries.

XML

Dependency	Minimum Version	Notes
lxml	4.5.0	XML parser for read_xml and tree builder for to_xml

SQL databases

Dependency	Minimum Version	Notes
SQLAlchemy	1.4.0	SQL support for databases other than sqlite
psycopg2	2.8.4	PostgreSQL engine for sqlalchemy
pymysql	0.10.1	MySQL engine for sqlalchemy

Other data sources

Dependency	Minimum Version	Notes
PyTables	3.6.1	HDF5-based reading / writing
blosc	1.20.1	Compression for HDF5
zlib		Compression for HDF5
fastparquet	0.4.0	Parquet reading / writing
pyarrow	1.0.1	Parquet, ORC, and feather reading / writing
pyreadstat	1.1.0	SPSS files (.sav) reading

Warning:

- If you want to use `read_orc()`, it is highly recommended to install pyarrow using conda. The following is a summary of the environment in which `read_orc()` can work.

System	Conda	PyPI
Linux	Successful	Failed(pyarrow==3.0 Successful)
macOS	Successful	Failed
Windows	Failed	Failed

Access data in the cloud

Dependency	Minimum Version	Notes
fsspec	0.7.4	Handling files aside from simple local and HTTP
gcsfs	0.6.0	Google Cloud Storage access
pandas-gbq	0.14.0	Google Big Query access
s3fs	0.4.0	Amazon S3 access