Dask Installation

Contents

Dask currently supports Linux, macOS, and Windows. See the Changelog for comprehensive release notes for each Dask version.

How to Install Dask

Dask installation can happen in a few different ways. You can install Dask with `conda`, with `pip`, or install from source.

Conda

If you use the Anaconda distribution, Dask installation will occur by default. You can also install or upgrade Dask using the conda install command:

```
conda install dask
```

This installs Dask and **all** common dependencies, including pandas and NumPy. Dask packages are maintained both on the default channel and on conda-forge. You can select the channel with the `-c` flag:

```
conda install dask -c conda-forge
```

Optionally, you can obtain a minimal Dask installation using the following command:

```
conda install dask-core
```

This will install a minimal set of dependencies required to run Dask similar to (but not exactly the same as) `python -m pip install dask` below.

Pip

You can use pip to install everything required for most common uses of Dask (e.g. Dask Array, Dask DataFrame, etc.). This installs both Dask and dependencies, like NumPy and pandas, that are necessary for different workloads. This is often the right choice for Dask users:

```
python -m pip install "dask[complete]"    # Install everything
```

You can also install only the Dask library. Modules like `dask.array`, `dask.dataframe`, or `dask.distributed` won't work until you also install NumPy, pandas, or Tornado, respectively. This is common for downstream library maintainers:

```
python -m pip install dask              # Install only core parts of dask
```

We also maintain other dependency sets for different subsets of functionality:

```
python -m pip install "dask[distributed]" # Install requirements for distributed dask
```

We have these options so that users of the lightweight core Dask scheduler aren't required to download the more exotic dependencies of the collections (Numpy, pandas, Tornado, etc.).

Install from Source

To install Dask from source, clone the repository from github:

```
git clone https://github.com/dask/dask.git
cd dask
python -m pip install .
```

You can also install all dependencies as well:

```
python -m pip install ".[complete]"
```

You can view the list of all dependencies within the `extras_require` field of `setup.py`.

Or do a developer install by using the `-e` flag (see the Install section in the Development Guidelines):

```
python -m pip install -e .
```

Anaconda

Dask is included by default in the Anaconda distribution.

Optional dependencies

Specific functionality in Dask may require additional optional dependencies. For example, reading from Amazon S3 requires s3fs. These optional dependencies and their minimum supported versions are listed below.

bokeh	>=2.4.2	Visualizing dask diagnostics
cityhash		Faster hashing of arrays
distributed	>=2.0	Distributed computing in Python
fastparquet		Storing and reading data from parquet files
gcsfs	>=0.4.0	File-system interface to Google Cloud Storage
graphviz		Graph visualization using the graphviz engine
ipycytoscape		Graph visualization using the cytoscape engine
murmurhash		Faster hashing of arrays
numpy	>=1.21	Required for dask.array
pandas	>=1.3	Required for dask.dataframe
psutil		Enables a more accurate CPU count
pyarrow	>=7.0	Python library for Apache Arrow
s3fs	>=0.4.0	Reading from Amazon S3
scipy		Required for dask.array.stats