

Collections

- `dask.array` : Parallelized version of numpy arrays
- `dask.bag` : Parallelized collections
- `dask.dataframe` : Parallelized version of pandas dataframes

Schedulers

- `synchronous` : Not parallelized; for debugging
- `threaded` : thread-based
- `multiprocessing` : process (fork/spawn) based
- `dask.distributed` : Parallelize across multiple nodes!

Custom parallelization:

- `dask.delayed` : Build the task graph before you run it
- **Futures API** : Build the task graph on the fly

Traditional

Compile

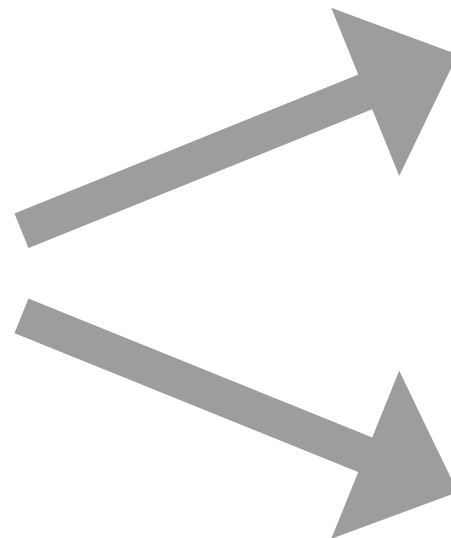
Run

Task-Based

Compile

Generate
Task Graph

Execute Tasks



(Just my advice!)