Parallel Data Download with NLDASAgent

The *NLDASAgent* module has been enhanced with three new configuration options to improve download performance and reliability:

- parallel – Enables concurrent downloading of data instead of processing requests sequentially. When enabled (True), multiple download tasks are executed simultaneously, significantly reducing the overall execution time.
- max_workers – Specifies the maximum number of worker threads used during parallel execution. Increasing this value allows more download tasks to run concurrently, which can further improve performance depending on the available system resources.
- file_size_check – Verifies the size of each downloaded file to ensure that the download completed successfully. This prevents corrupted or incomplete files from being used in subsequent workflow steps and avoids unnecessary reprocessing.

These options provide users with greater control over download performance while maintaining data integrity. Figure 1 illustrates the new configuration options available in the NLDASAgent module.

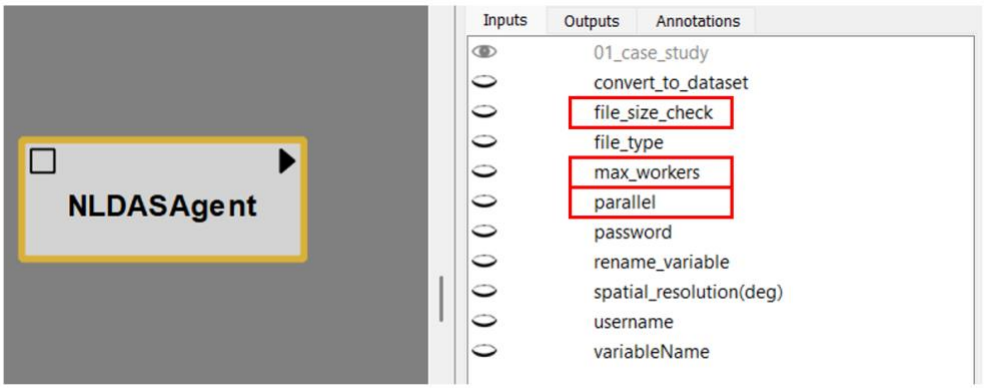

Figure 1. New configuration options of the NLDASAgent module, including parallel, max_workers, and file_size_check.

The following example demonstrates how to configure and use the new parallel download capability.

- Add a *TimeRange* module and configure it with:
  - timeini: 2015/01/01 00:00:00
  - timeend: 2015/01/10 00:00:00
- Add a *SpaceRange* module and specify the spatial extent:
  - x_max: -75
  - x_min: -76
  - y_max: 45
  - y_min: 44
- Connect the timerange output of the *TimeRange* module to the subrange input of the *SpaceRange* module.
- Add an *NLDASAgent* module and enter your NASA EarthData credentials.

    - Configure the *NLDASAgent* to download the following variables:
        - u-wind [m/s]
        - v-wind [m/s]
    - Enable parallel processing by setting: parallel = True
    - Mark the “Boolean” choice in the ”convert_to_dataset” option.
- Add an *msmShowChart* module and connect the desired output variables from the *NLDASAgent* to visualize the downloaded data.

To evaluate the effectiveness of the parallel execution feature, duplicate the workflow described above (*TimeRange*, *SpaceRange*, *NLDASAgent*, and *msmShowChart*) and modify the duplicated workflow as follows:

- Change the simulation period to:
    - timeini: 2016/01/01 00:00:00
    - timeend: 2016/01/10 00:00:00
- Set parallel = False in the duplicated *NLDASAgent* module while keeping all other settings unchanged.

Running both workflows under the same conditions demonstrates the performance improvement achieved through parallel processing. In this example, the workflow with parallel enabled completes in approximately 2 minutes, whereas the workflow with parallel disabled requires approximately 10 minutes. This represents an execution time reduction of about 80%, highlighting the efficiency of concurrent downloading for large datasets.

Figure 2 presents the complete workflow used for the performance comparison. Figure 3 compares the execution times of the parallel and sequential configurations, and Figure 4 shows the Temperature and Pressure outputs produced by the NLDASAgent module.

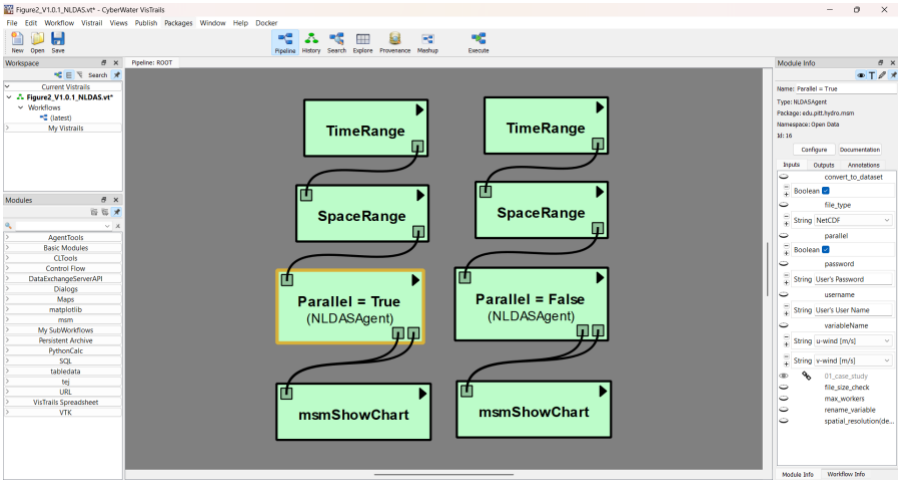

Figure 2. Complete CyberWater workflow demonstrating parallel and sequential NLDASAgent configurations for performance comparison.

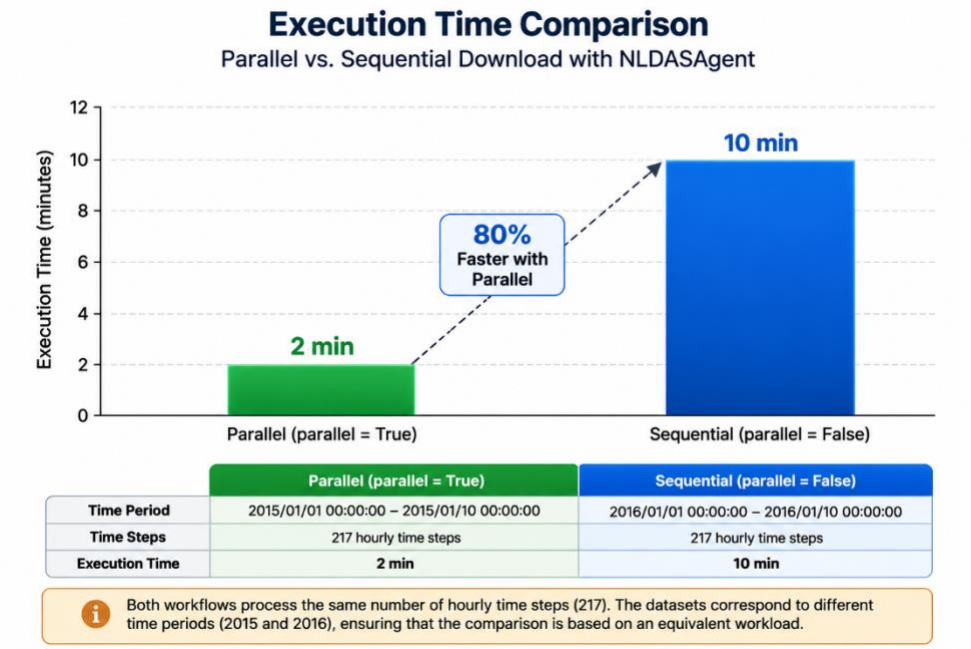

Figure 3. Execution time comparison between the parallel (parallel = True) and sequential (parallel = False) NLDASAgent configurations.

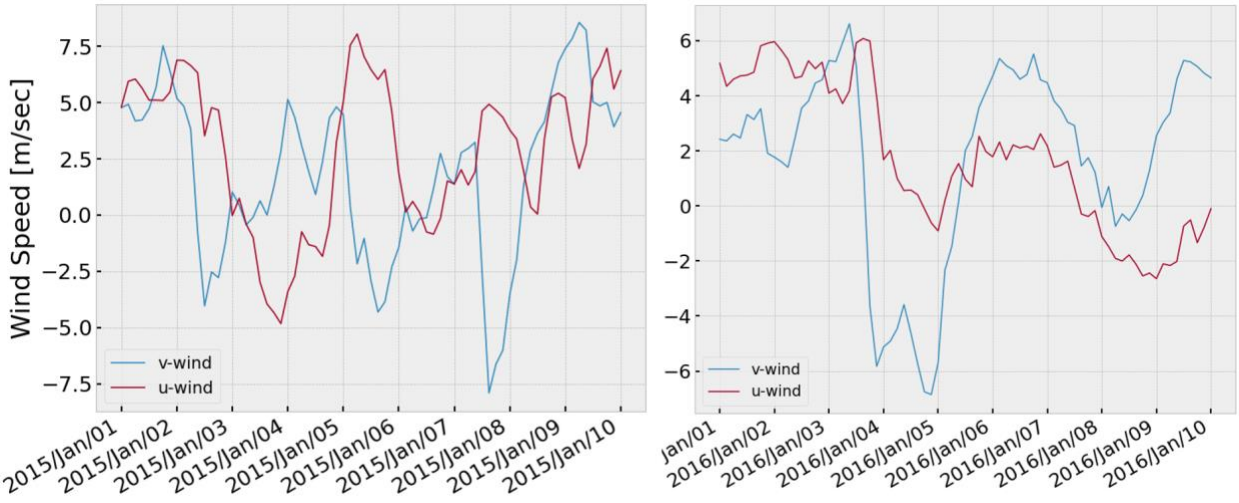

Figure 4. Wind Speed outputs generated by the NLDASAgent module and displayed using msmShowChart.

The workflow in Figure 2 can be downloaded from the CyberWater website <Parallel vs. Sequential Download with NLDASAgent>.