Performative Benchmarking of Unidata MetPy with ASV

Jaye Norman 7/23/2025

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Why should we benchmark?

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

C++ification

- Linfeng has created C++ to help MetPy run faster through bottlenecks like CAPE calculation
- We needed a way to quantify his and future changes

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Other Benefits

- Benchmarking also identifies bottlenecks and shows us when merges change the performance of the code

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

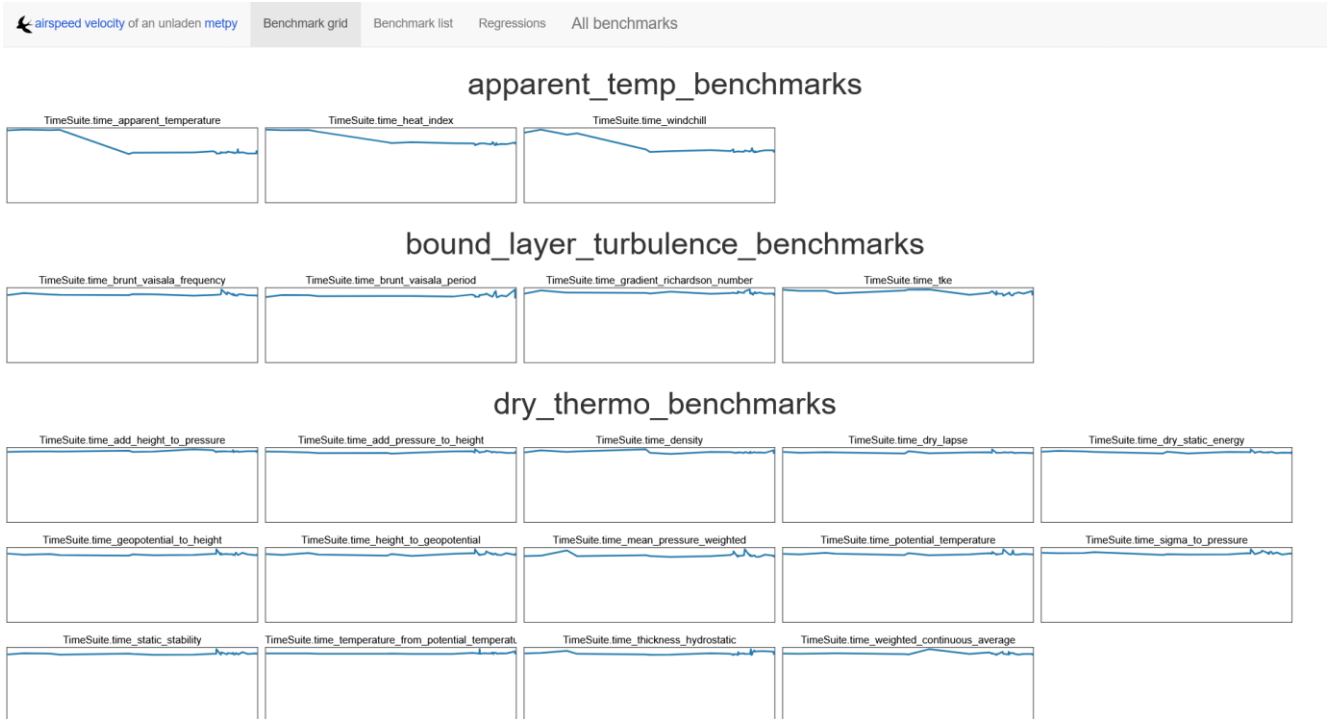
Where can I see the results?

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

unidata.github.io/MetPy-benchmark

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

ASV Webpage



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

What software did we use?

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

GitHub

- Provides a home for the MetPy source code
- The MetPy repo holds the benchmarking functions and configs for CI/CD
- The MetPy-benchmark repo holds the results of the benchmark runs

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Airspeed Velocity

- ASV is an open-source python benchmarking package
- Creates environments based on historical states of a repo, runs the benchmark functions, and returns them in a pretty html format

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins

- Unidata's Jenkins instance is used for CI/CD workflows
- Runs on a machine owned by UCAR and ensures that the machine specs are the same between each benchmark run

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Docker

- Within Jenkins, the benchmarks run in a Docker container from a Dockerfile
- Improves consistency between runs and is portable to many devices for local benchmarking

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

GitHub Actions

- The MetPy-benchmark repo has a GHA that uses ASV to generate the html from the results
- Action also deploys files to a static page

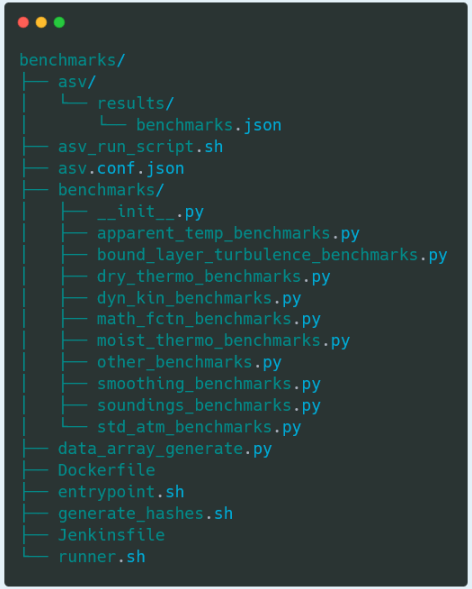
This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

How does benchmarking with ASV work?

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

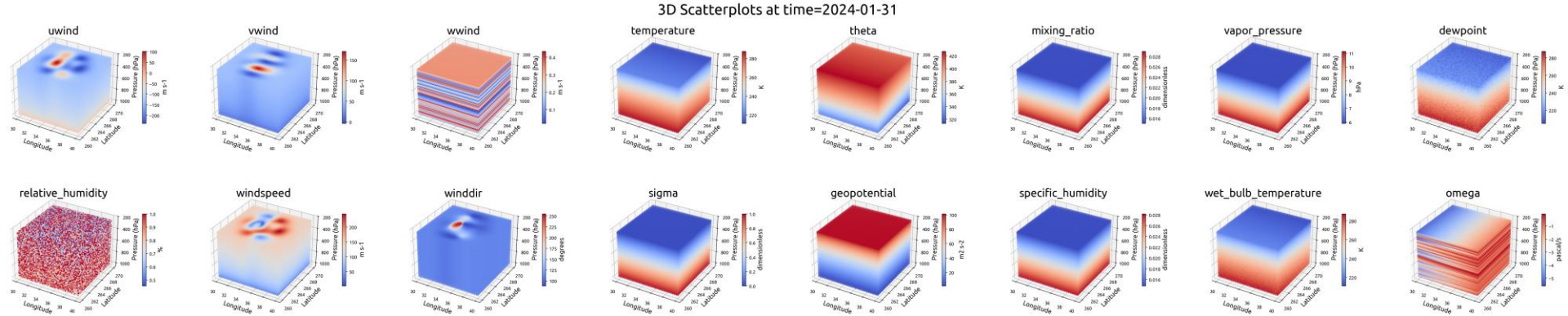
File Tree

- asv.conf.json: configuration file for ASV
- asv/results: where the results are stored
- benchmarks/.py files: benchmarking snippets



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Benchmark Dataset



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Benchmark setup_cache

```
    def setup_cache(self):
        """Collect the sample dataset from the filepath and opens it as an
xarray.

        Returns
        -------
        ds
            Dataset with artificial meteorology data for testing
        """
        base_path = os.path.dirname(__file__)  # path to current file
        file_path = os.path.join(base_path, '..', 'data_array_compressed.nc')
        file_path = os.path.abspath(file_path)
        ds = xr.open_dataset(file_path)
        return ds
```

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Benchmark setup

```python
def setup(self, ds):
    """Set up the appropriate slices from the sample dataset for testing.

    Parameters
    ----------
    ds : dataset
        The dataset made in setup_cache which contains the testing data
    """
    self.timeslice = ds.isel(time=0)
    self.pressureslice = ds.isel(time=0, pressure=0)
```

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Example Benchmark

```python
def time_lcl(self, timeslice):
    """Benchmarks the LCL function over a 3d cube of data"""
  mpcalc.lcl(self.timeslice.pressure, self.timeslice.temperature,
             self.timeslice.dewpoint)
```

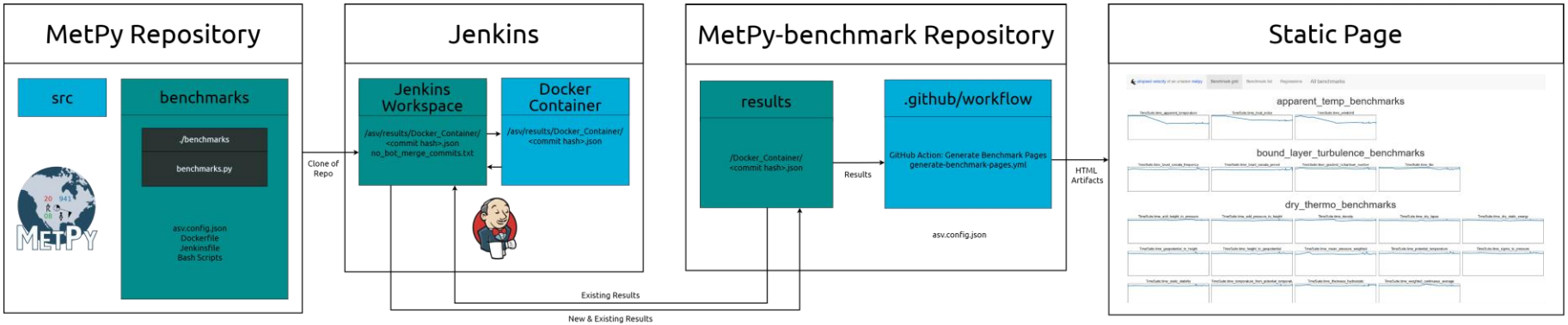
This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

How does the workflow run?

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins Trigger

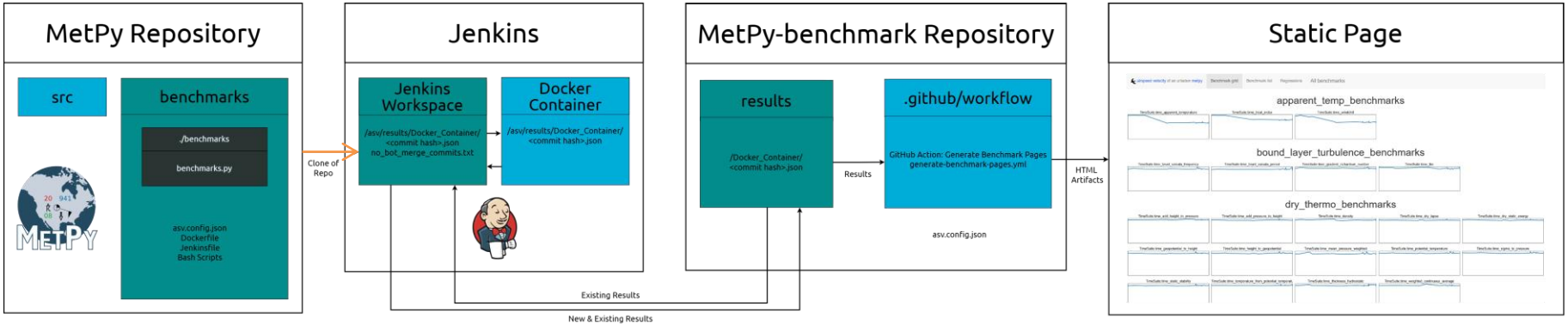
1) Jenkins is triggered Saturday morning



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins Setup

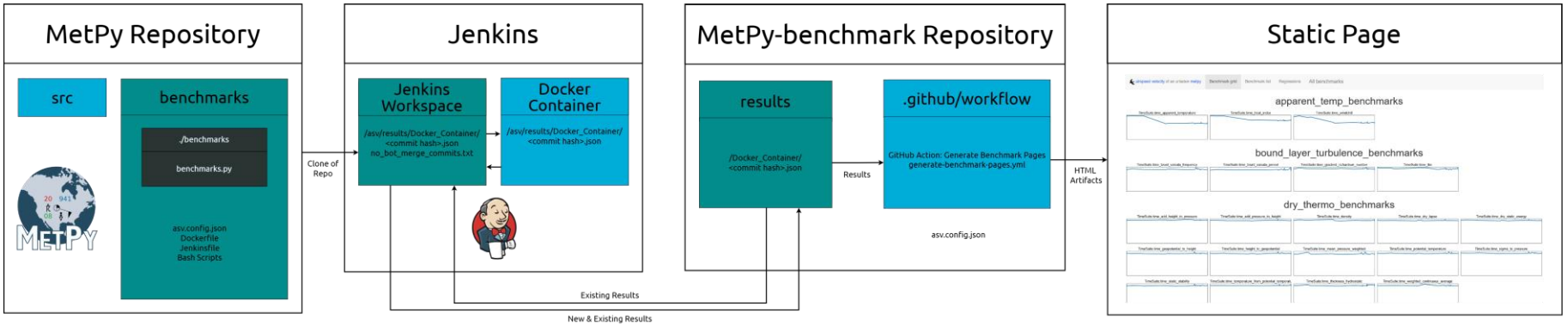
2) Jenkins clones MetPy



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins Setup

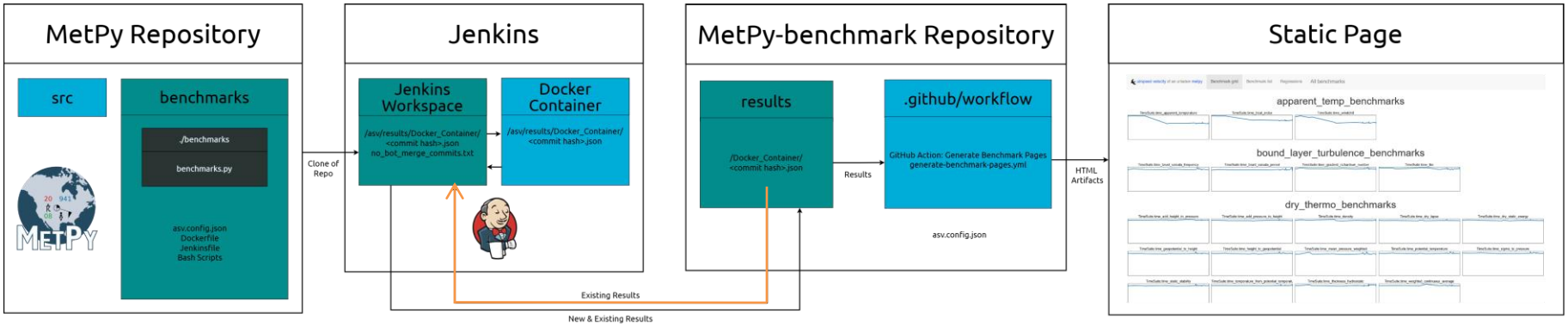
3) Jenkins searches MetPy's history for minor version commits and recent merges



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins Setup

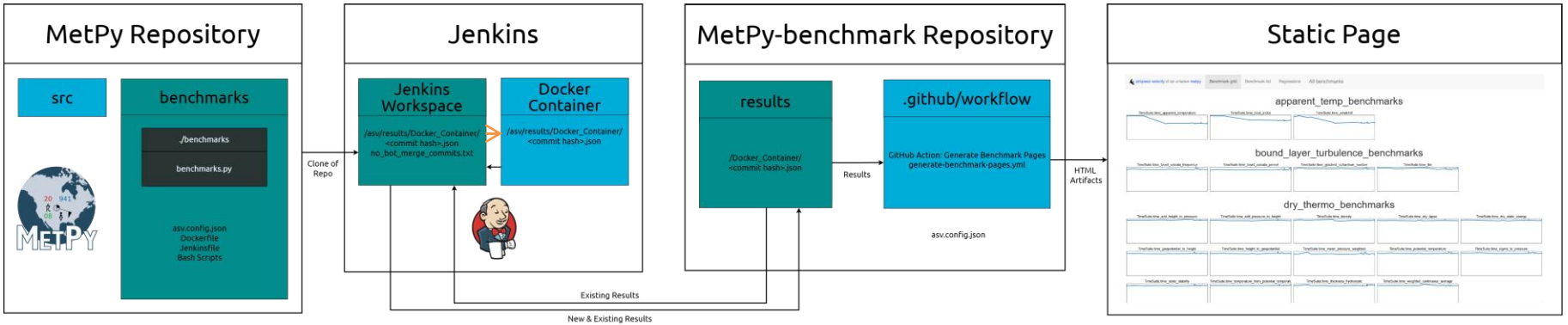
4) From MetPy-benchmark, Jenkins copies the existing results



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins Setup

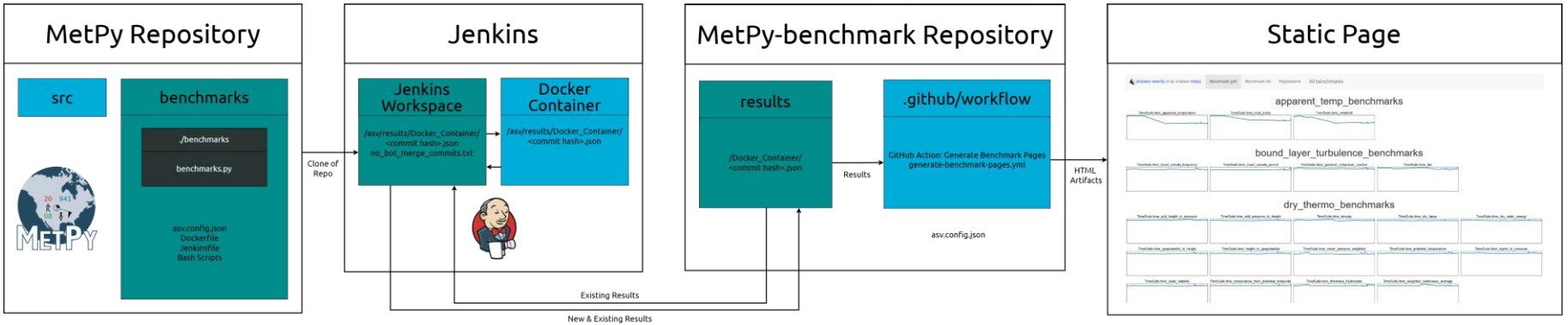
5) Jenkins builds a docker container from the Dockerfile



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Docker runs Benchmarks

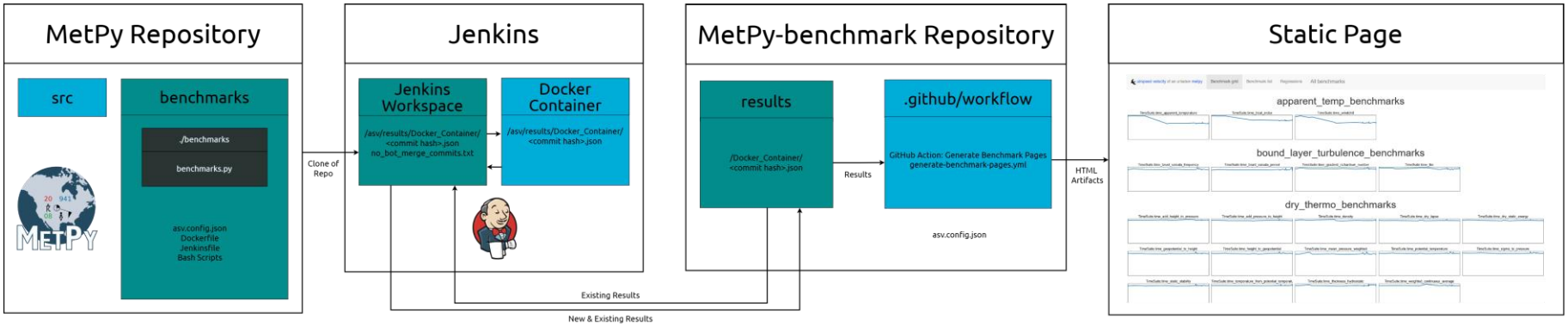
1) With a docker run command, the docker file runs benchmarks on the commit file



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Docker runs Benchmarks

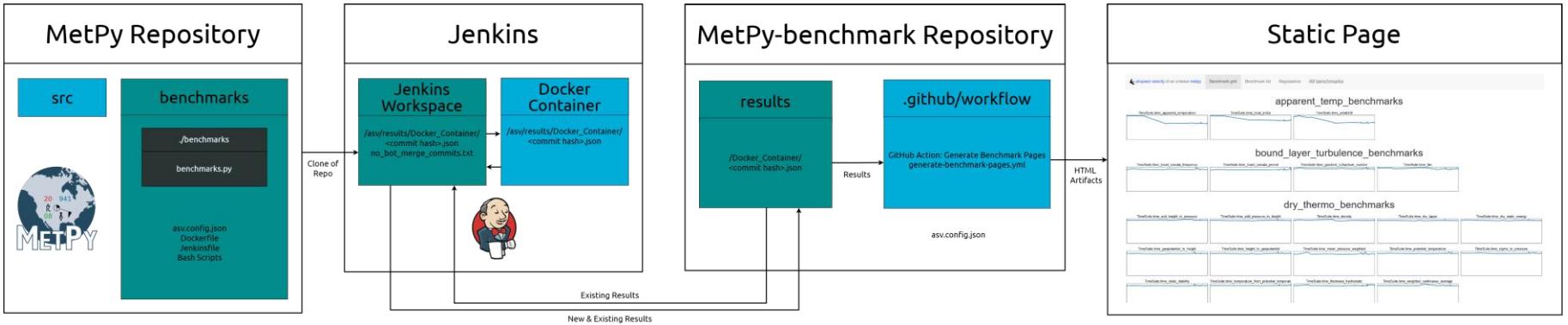
2) If there are already successful results for a certain commit, the benchmarks are skipped



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Docker runs Benchmarks

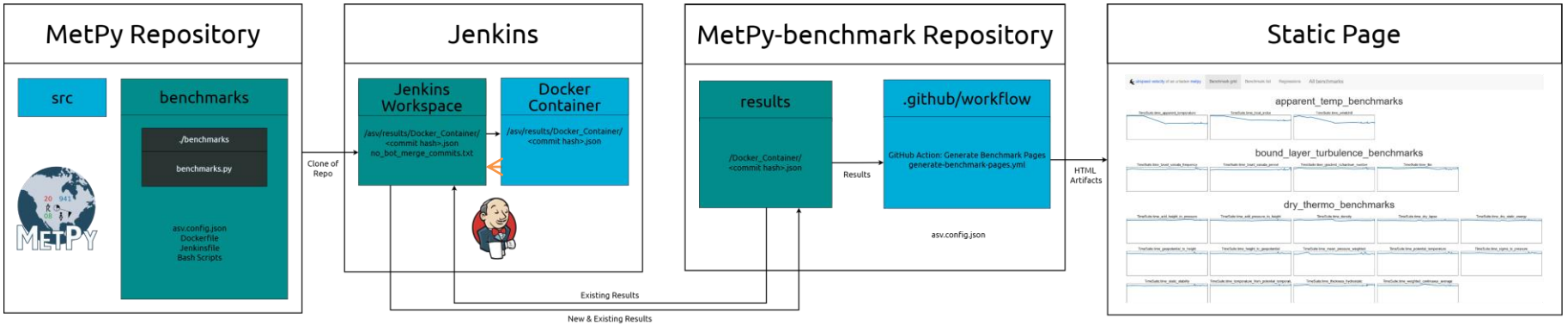
3) If there are not, like for a new commit, the benchmarks are run for this commit



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins saves results

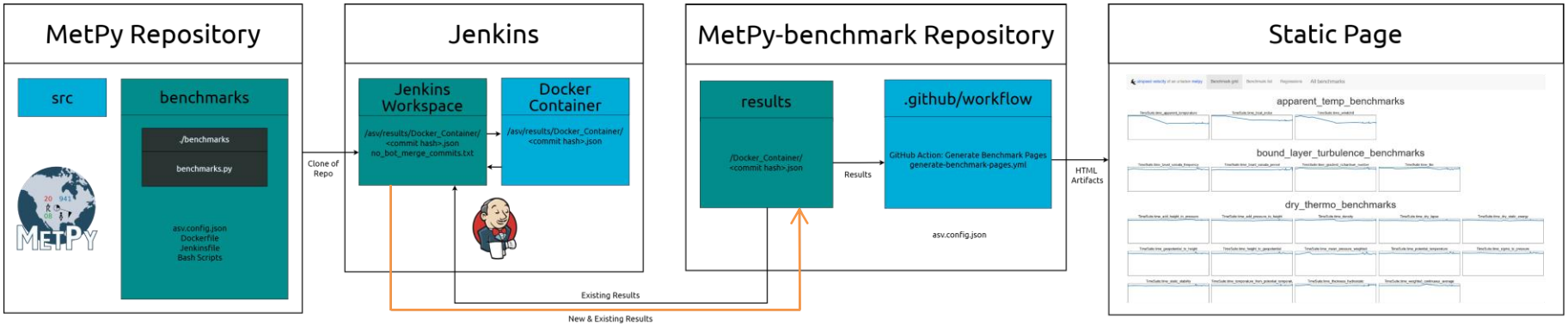
1) The docker container terminates when it's finished



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Jenkins saves results

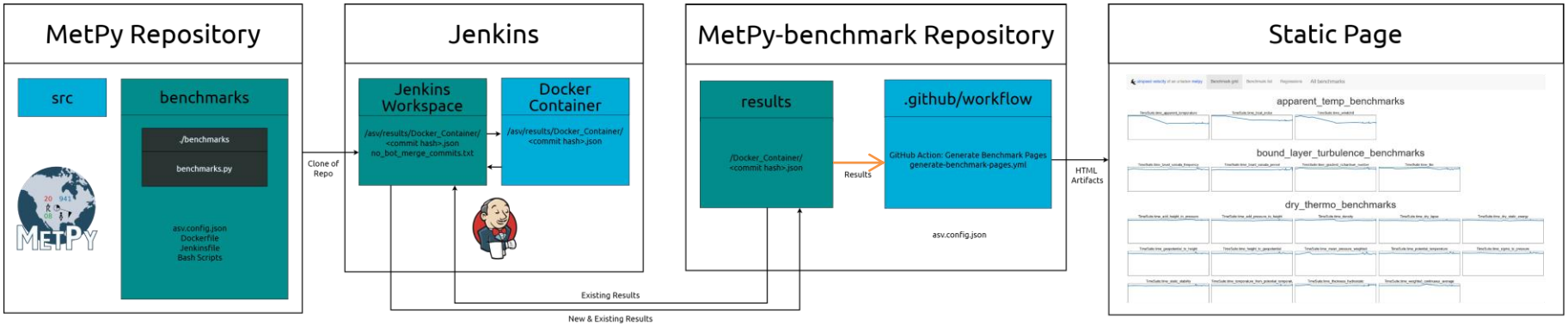
2) Jenkins pushes the results, old and new, to the Metpy-benchmark repository



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Metpy-benchmark Deployment

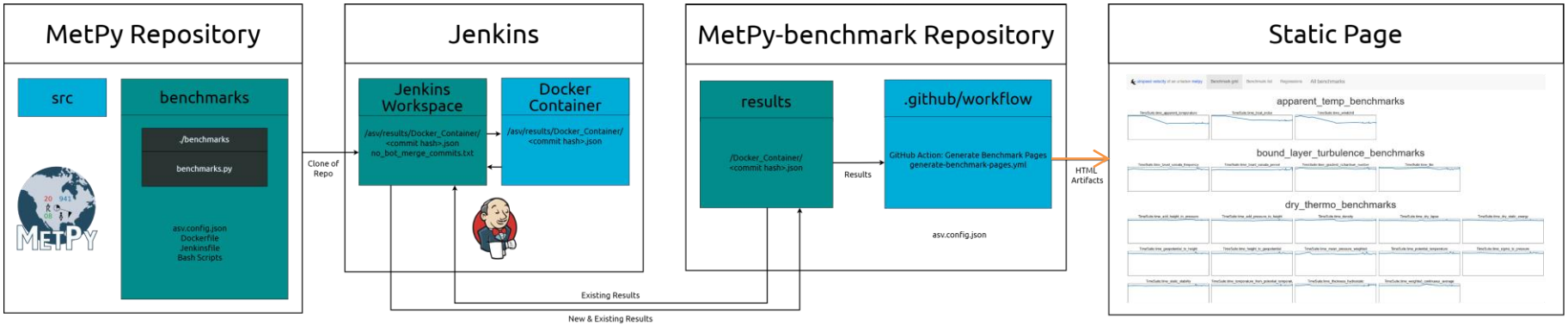
1) Upon push to Metpy-benchmark, a GitHub Action creates the html for the static page



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Metpy-benchmark Deployment

2) The GitHub action deploys the html to the static page



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Can we see performance changes before we merge commits?

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Comparative Benchmarking

- Comparative benchmarking is when you compare the performance of two branches
- ASV has a built-in function for this, and when combined with GHA, can do it automatically on pull requests

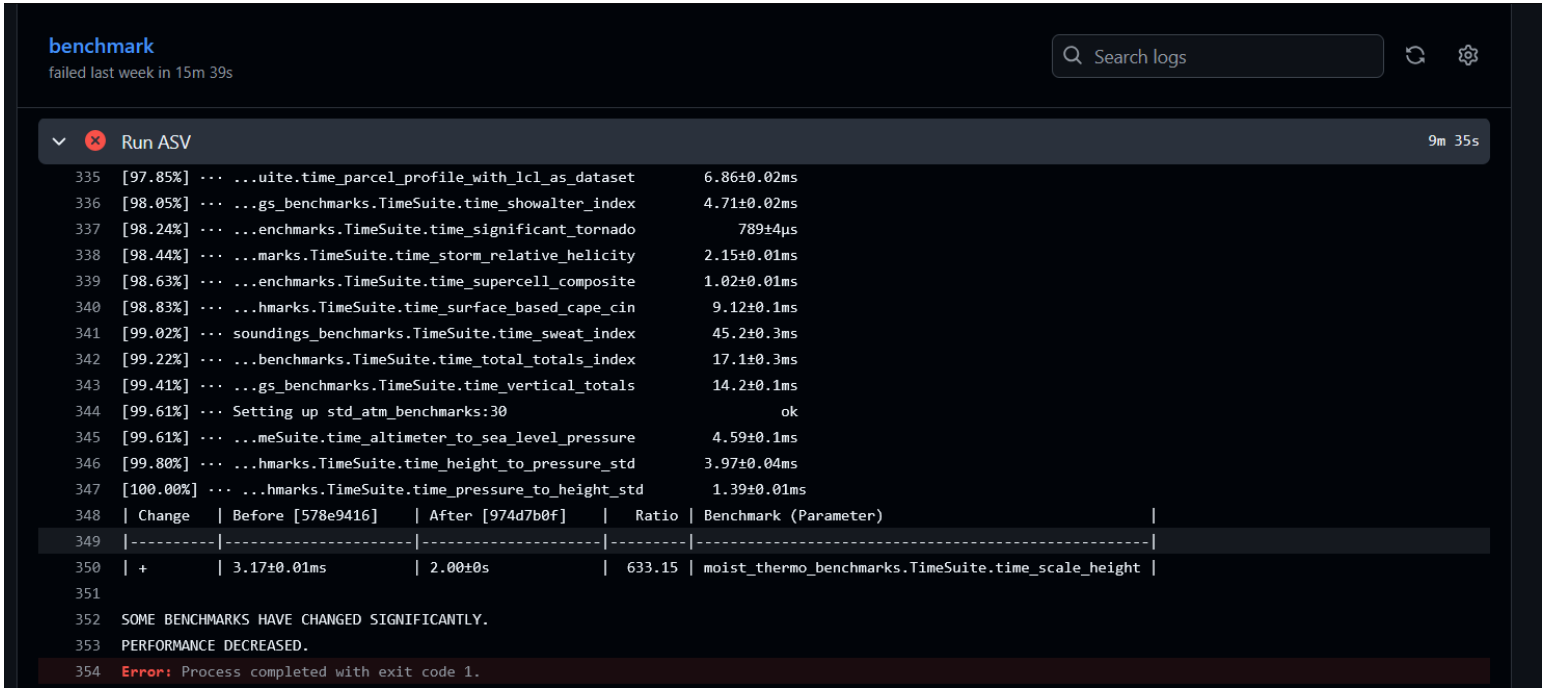
This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Comparative GitHub Action

- We can compare the current main branch to the pull request branch using their commit hashes if the PR is labelled.
- Currently a failure occurs when a benchmark is 10% slower, but this is customizable

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Example GitHub Action



This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

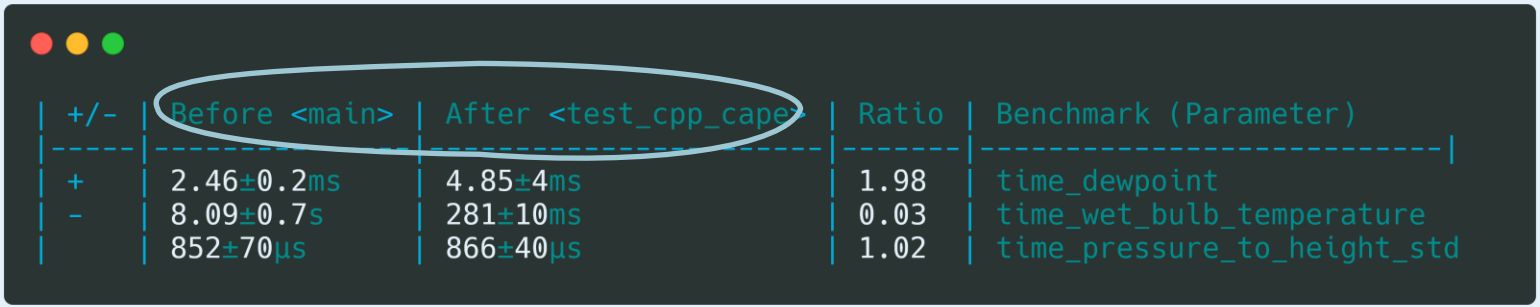
Local Comparisons

- You can also locally run a comparison assuming you have an untouched local main branch and have ASV installed
- This allows you to see if your changes are working as you anticipate

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Example Local Comparison

- This compares main and test_cpp_cape



+/-	Before <main>	After <test_cpp_cape>	Ratio	Benchmark (Parameter)
+	2.46±0.2ms	4.85±4ms	1.98	time_dewpoint
-	8.09±0.7s	281±10ms	0.03	time_wet_bulb_temperature
	852±70µs	866±40µs	1.02	time_pressure_to_height_std

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Example Local Comparison

- The dewpoint function is 98% slower

+/-	Before <main>	After <test_cpp_cape>	Ratio	Benchmark (Parameter)
+	2.46±0.2ms	4.85±4ms	1.98	time_dewpoint
-	8.09±0.7s	281±10ms	0.03	time_wet_bulb_temperature
	852±70µs	866±40µs	1.02	time_pressure_to_height_std

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Example Local Comparison

- The wet_bulb function is 97% faster

+/-	Before <main>	After <test_cpp_cape>	Ratio	Benchmark (Parameter)
+	2.46±0.2ms	4.85±4ms	1.98	time_dewpoint
-	8.09±0.7s	281±10ms	0.03	time_wet_bulb_temperature
	852±70µs	866±40µs	1.02	time_pressure_to_height_std

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

Example Local Comparison

- The p_to_h function hasn't changed much

+/-	Before <main>	After <test_cpp_cape>	Ratio	Benchmark (Parameter)
+	2.46±0.2ms	4.85±4ms	1.98	time_dewpoint
-	8.09±0.7s	281±10ms	0.03	time_wet_bulb_temperature
	852±70µs	866±40µs	1.02	time_pressure_to_height_std

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.

NSF Unidata is one of the University Corporation for Atmospheric Research (UCAR)'s Community Programs (UCP), and is funded primarily by the U. S. National Science Foundation (Grant AGS-1901712).

This material is based upon work supported by NSF Unidata under award #2403649 from the U. S. National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of the NSF.