Caching software and building docker images: PR #222

Philip Keicher for the CMS Collaboration

columnflow framework meeting, HH May 17, 2024

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SPONSORED BY THE
Federal Ministry
of Education
and Research

Motivation

Main goal: Cache software stack

- Extremely useful for (workflow) unit tests
- First attempt last year with github caching
- Now switched to docker images → in line with request from CAT conveners

"Problems" with current setup

1. Requirements distributed
    - user needs to know where which requirement file is
    - version of venv fixed on first requirement file
2. Not installable on multiple platforms (→ accessibility of conda environment)
3. **New** (and not understood by me): Software setup sometimes fails because venv initialization removes `CF_CONDA_BASE` from `PATH`

Updates- Problem 1

- Requirements now centrally managed via pyproject.toml

```
dependencies={file = ["sandboxes/cf.txt"]}
readme = {file = ["README.md"], content-type="text/markdown"}
optional-dependencies={
        "dev" = {file = ["sandboxes/dev.txt"]},
        "columnar"= {file = ["sandboxes/columnar.txt"]},
        "ml_tf"= {file = ["sandboxes/ml_tf.txt", "sandboxes/columnar.txt"]
}
```

- User can then install any columnflow version via `python -m pip install columnflow[PACKAGE]`
- Requirements for more complex venvs compiled with pip-compile

Updates- Problem 2

- conda environment is now compiled using unidep package
- Input: pyproject.toml
- Enables more transparent package management, depending on platform

Updates- Problem 3

- Install columnflow in venv and copy executables to `bin` folder

→ guarantees that all our custom commands/scripts are always accessible

Conclusions

- Talked about PR #222
- Makes changes to software setup (organization of conda/python packages)

→ More transparency and easier access

- No changes for users (except probably reinstalling the software stack once)
- New docker images for unit tests
- Current setup probably not perfect, feedback appreciated!

DEMO