



Reproducible computational workflows with **signac**

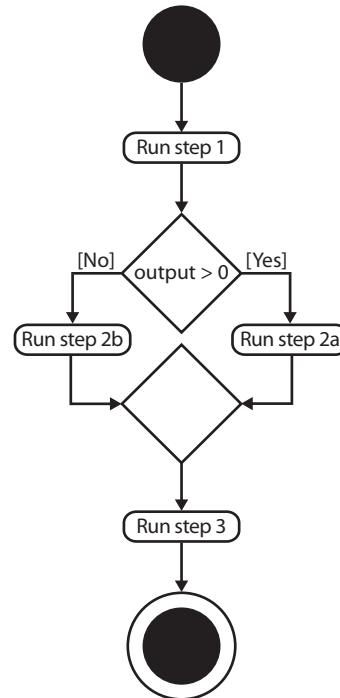
Bradley D. Dice, Vyas Ramasubramani, Carl S. Adorf, Sharon C. Glotzer

SC² High Performance Computing & Data Science Workshop

November 8, 2018, Ann Arbor, MI

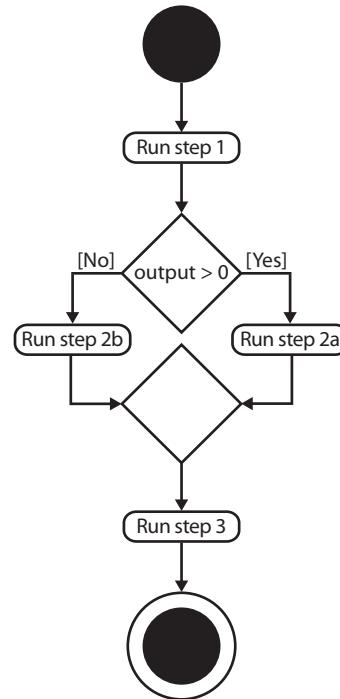
Planning a Parameter Study

concentration_A_0.25/
concentration_A_0.50/
concentration_A_0.75/



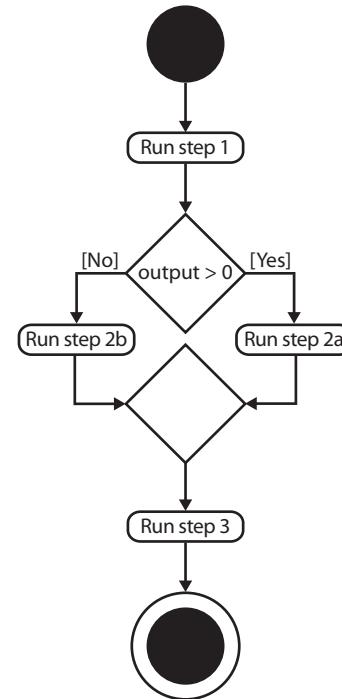
Planning a Parameter Study

concentration_A/0.25
concentration_A/0.50
concentration_A/0.75



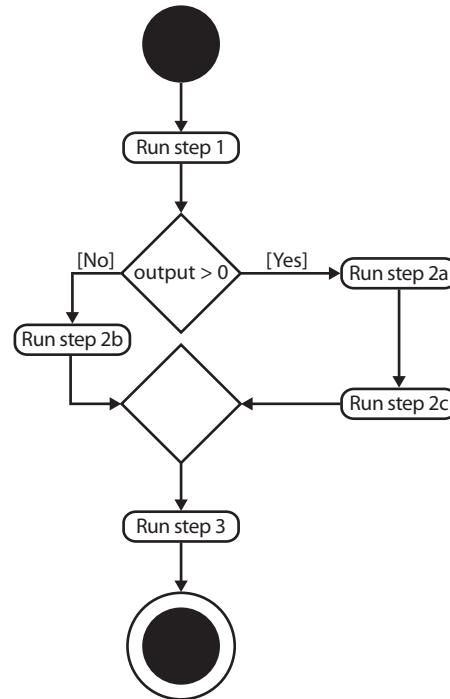
Planning a Parameter Study

conc_A/0.25
conc_A/0.50
conc_A/0.75



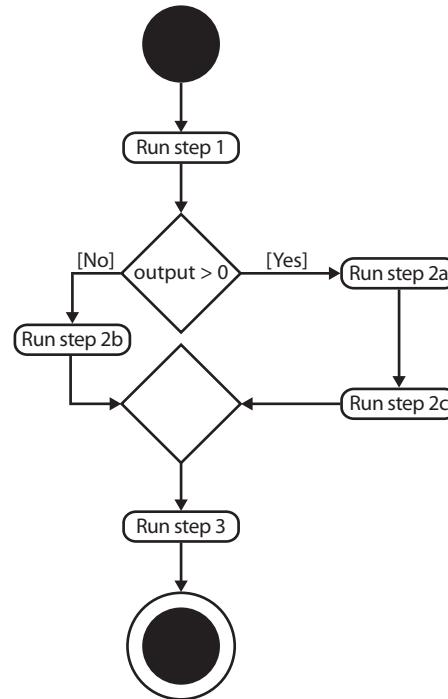
Planning a Parameter Study

conc_A/0.25/temp/0.8
conc_A/0.25/temp/1.0
conc_A/0.50
conc_A/0.75



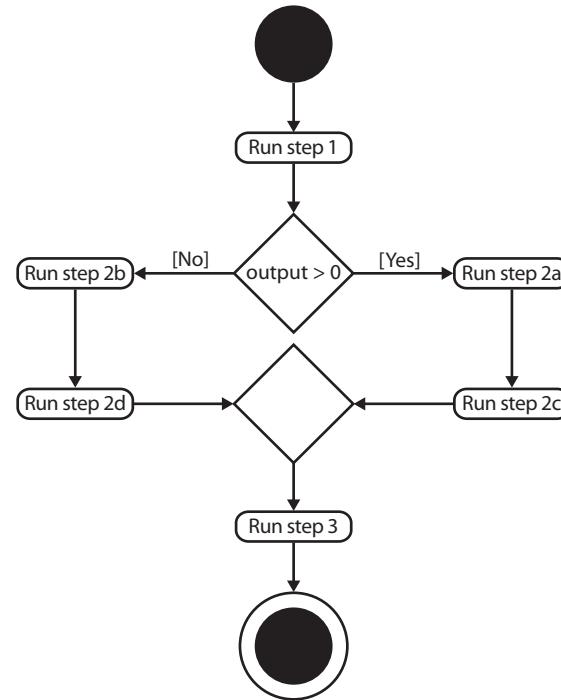
Planning a Parameter Study

temp/0.8/conc_A/0.25
temp/0.8/conc_A/0.50
temp/0.8/conc_A/0.75
temp/1.0/conc_A/0.25



Planning a Parameter Study

temp/0.8/conc_A/25/conc_B/05
temp/0.8/conc_A/50
temp/0.8/conc_A/75
temp/1.0/conc_A/25

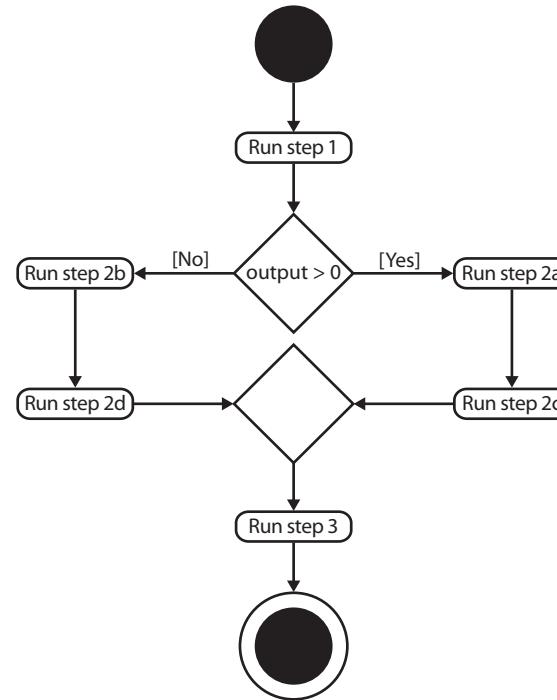


Planning a Parameter Study

```
#!/bin/bash
#SBATCH -J myproject
#SBATCH -N 16
#SBATCH -A ${MYACCOUNT}
#SBATCH -p ${QUEUE}
#SBATCH --ntasks-per-node 8
#SBATCH -t 12:00:00

cd ${WORKING_DIRECTORY}

mpirun -n 16 ./myscript.sh
```



The signac Framework



A **lightweight, application-agnostic software framework** that **unobtrusively helps users manage and scale file-based workflows, facilitating data reuse, sharing, and reproducibility.**

Python 2/3 | open source BSD-3 | install with pip or conda

<https://signac.io>

- Simple data and workflow management with the signac framework, JCMS 146, 220-229, (2018)

Topic Overview

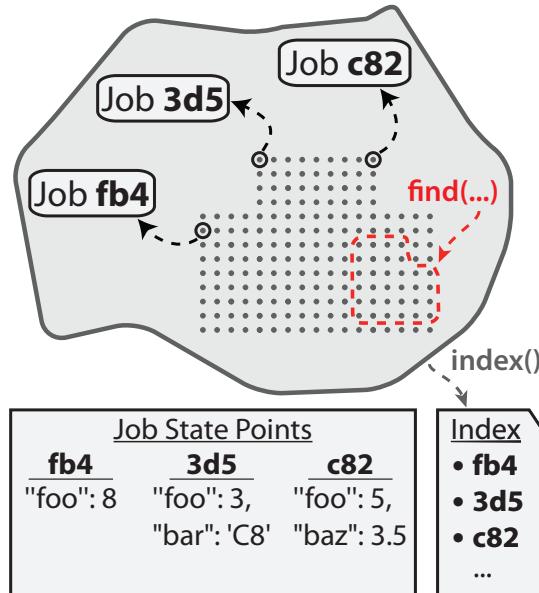
1. Introduction
2. Projectile Demo
3. Recent Development and Future Goals



Data (**signac**) and Workflow (**signac-flow**) Management

signac

(a) Active Workspace



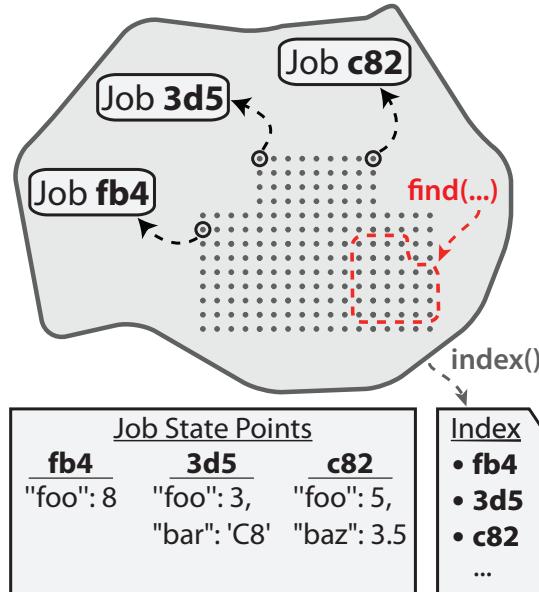
- signac: A Python framework for data and workflow management. In Proceedings of the 17th Python in Science Conference, 2018.

Data (**signac**) and Workflow (**signac-flow**) Management

signac

(a) Active Workspace

The workspace consists of **jobs**, data containers associated with distinct metadata mappings (called **state points**).



- signac: A Python framework for data and workflow management. In Proceedings of the 17th Python in Science Conference, 2018.

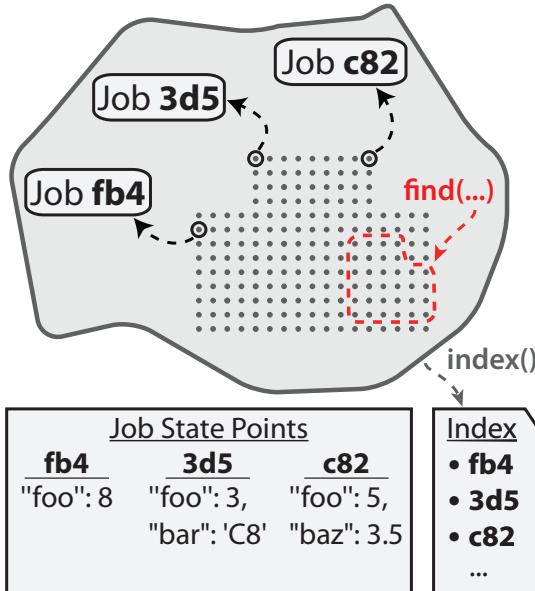
Data (**signac**) and Workflow (**signac-flow**) Management

signac

(a)

Active Workspace

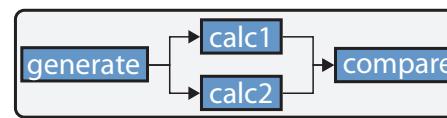
The workspace consists of **jobs**, data containers associated with distinct metadata mappings (called **state points**).



signac-flow

(b)

Project Workflow



(c)

Status Tracking

operation(job)	Status
generate(fb4)	✓
calc1(fb4)	✓
calc2(fb4)	✓
compare(fb4)	→
generate(3d5)	✓
calc1(3d5)	✓
calc2(3d5)	→
compare(3d5)	✗
...	...

- signac: A Python framework for data and workflow management. In Proceedings of the 17th Python in Science Conference, 2018.

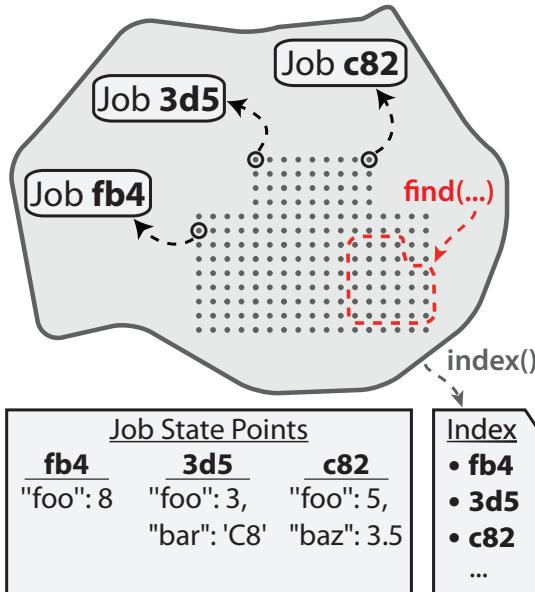
Data (**signac**) and Workflow (**signac-flow**) Management

signac

(a)

Active Workspace

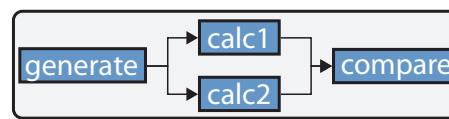
The workspace consists of **jobs**, data containers associated with distinct metadata mappings (called **state points**).



signac-flow

(b)

Project Workflow



(c)

Status Tracking

operation(job)	Status
generate(fb4)	✓
calc1(fb4)	✓
calc2(fb4)	✓
compare(fb4)	→
generate(3d5)	✓
calc1(3d5)	✓
calc2(3d5)	→
compare(3d5)	✗
...	...

The workflow consists of **operations**, linked through condition functions. This forms a directed graph.

- signac: A Python framework for data and workflow management. In Proceedings of the 17th Python in Science Conference, 2018.

The signac Team



Carl S. Adorf
Lead developer &
co-maintainer,
signac and signac-flow



Vyas Ramasubramani
Developer &
co-maintainer,
signac and signac-flow



Bradley Dice
Lead developer &
maintainer,
signac-dashboard

Top 6 Features Released In Last 12 Months

Advanced Searching and Aggregation



```
find({"T.$gt": 298})  
.groupby("P")
```

One-script Projects



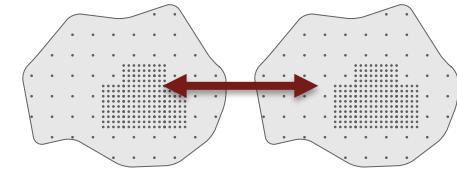
Implement workflows in < 10 lines.

Schema Detection

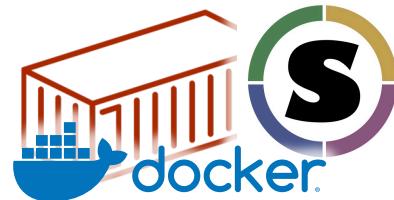


```
'T': int([298, 300, 302])  
'P': float([0.1, 1.0])
```

Synchronization



Improved Container Support



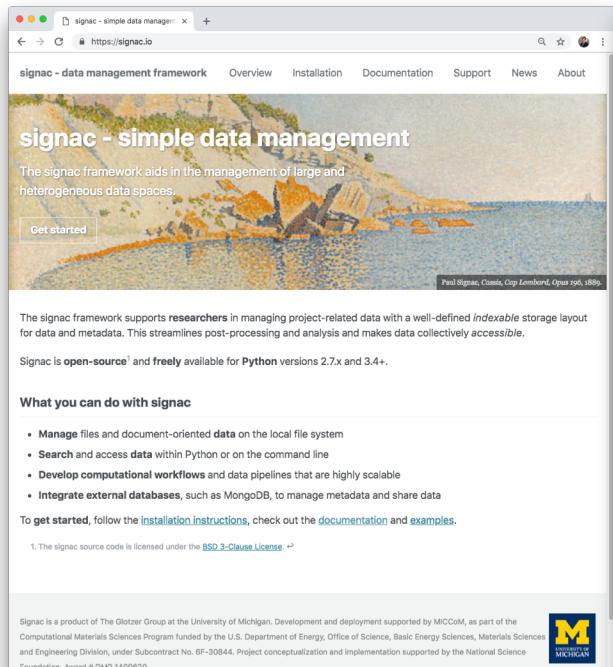
Templating



- container by DPIcons from the Noun Project
- Relational Schema by Bebris from the Noun Project
- search data by SBTS from the Noun Project
- Workflow by ProSymbols from the Noun Project

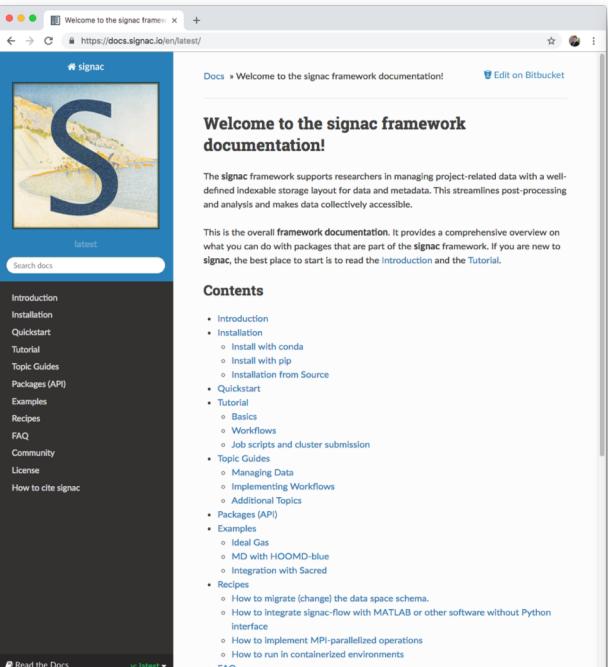
News, Documentation & Chat Room Support

<https://signac.io>



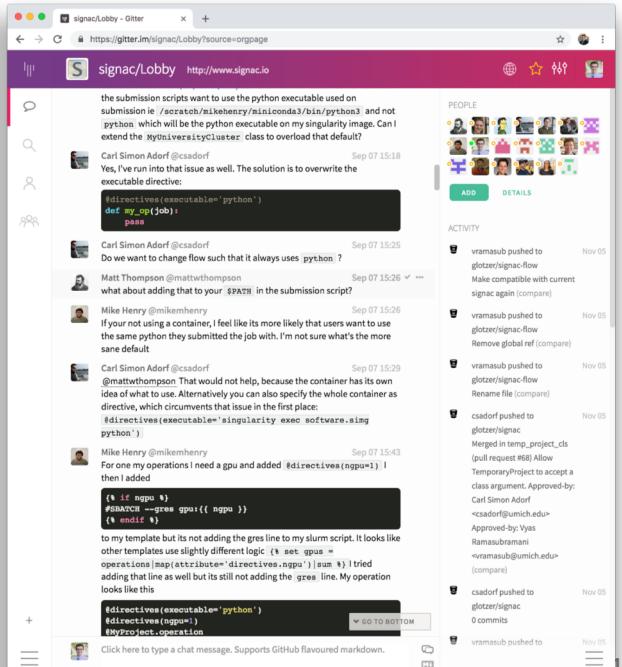
The screenshot shows the homepage of the signac simple data manager. It features a large image of a painting by Paul Signac, 'Cassis, Cap Lombard, Opus 196, 1889'. Below the image, the title 'signac - simple data management' is displayed. A subheader states: 'The signac framework aids in the management of large and heterogeneous data spaces.' A 'Get started' button is visible. The main content area includes a paragraph about the framework's purpose and a note that it is open-source and available for Python 2.7.x and 3.4+. A sidebar on the right lists various documentation sections such as Introduction, Installation, Quickstart, Tutorial, Topic Guides, Packages (API), Examples, Recipes, FAQ, Community, License, and How to cite signac. At the bottom, there is a note about the BSD 3-Clause License and a Michigan Wolverines logo.

<https://docs.signac.io>



The screenshot shows the welcome page of the signac framework documentation. It features a large blue letter 'S' on a background image of a painting. The title 'Welcome to the signac framework documentation!' is centered. Below the title, a paragraph explains the framework's purpose: 'The signac framework supports researchers in managing project-related data with a well-defined indexable storage layout for data and metadata. This streamlines post-processing and analysis and makes data collectively accessible.' A 'Contents' section is listed on the left, containing links to Introduction, Installation, Quickstart, Tutorial, Topic Guides, Packages (API), Examples, Recipes, FAQ, Community, License, and How to cite signac. The main content area also contains a paragraph about the overall framework documentation and its purpose.

<https://gitter.im/signac/Lobby>



The screenshot shows a Gitter chat room for the signac framework. The interface includes a sidebar with user icons and activity logs. The main area displays a conversation between several users. One user, vramasub, has pushed changes to glotzer/signac-flow, mentioning a fix for a singularity image issue. Another user, Carl Simon Adorf (@csadorf), responds with a code snippet showing how to extend the MyInLivercycleCluster class. The conversation continues with users discussing submission scripts and their execution environments. A large portion of the screen is occupied by a code editor showing a Python script with directives like '#SBATCH --gres gpu:1' and '#MyProject.operation'.

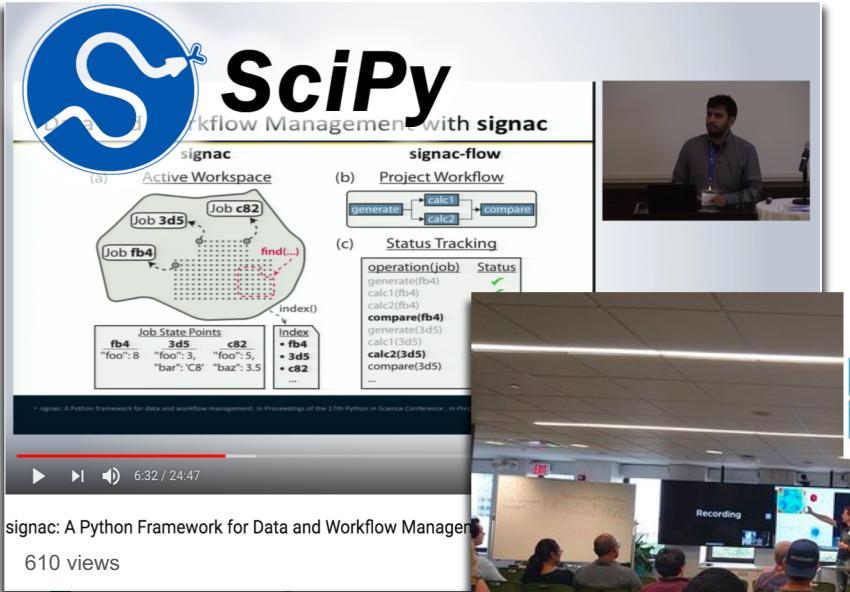


Impact of signac

- In use by several research groups across the country, especially in materials science and molecular simulations
 - Princeton, Vanderbilt, Boise State, Air Force Research Laboratory, ...
- 10,000+ downloads of signac & signac-flow conda packages
- Actively used on projects with thousands of jobs and terabytes of data
- signac-flow has built-in support for Flux and XSEDE clusters, and adaptable templates for any SLURM/TORQUE scheduler



Community Building



SciPy
Workflow Management with `signac`

The slide shows a presentation titled "Workflow Management with signac". It includes:

- (a) Active Workspace: A diagram showing a central workspace containing multiple jobs (fb4, 3d5, c82) and their dependencies.
- (b) Project Workflow: A flowchart illustrating the workflow steps: generate → calc1 → calc2 → compare.
- (c) Status Tracking: A table showing the status of various operations. Most are marked as green (successful), except for calc1(fb4) which is red (failed).

Signac logo: A blue circle with a white stylized 'S' and a blue ribbon-like tail.

610 views



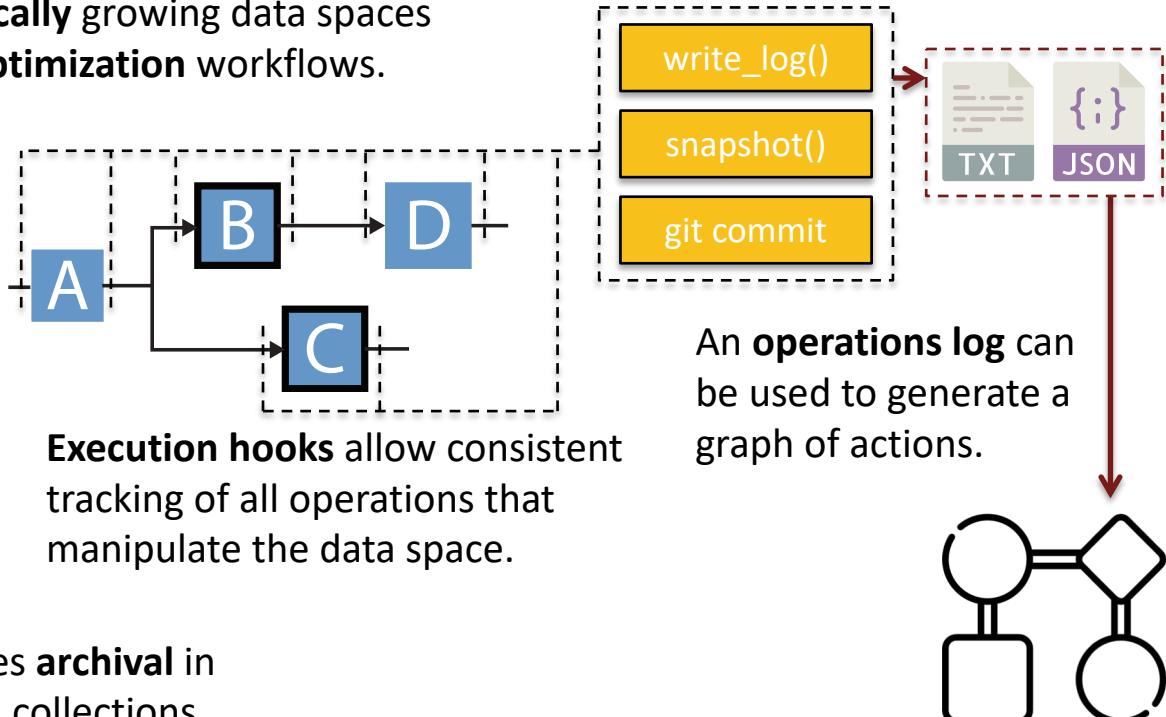
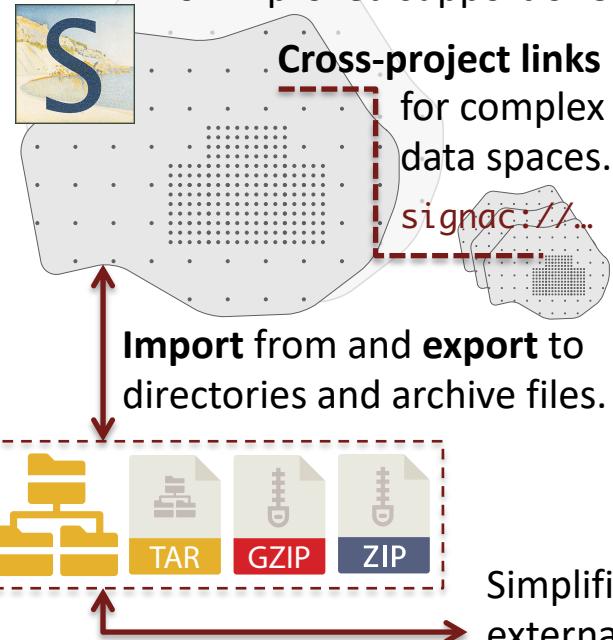
[signac Projectile Demo]

<https://github.com/UM-SC2/HPC-DS-workshop-2018/tree/master/session06>



Latest Development

Better handling of **dynamically** growing data spaces
for improved support of **optimization** workflows.

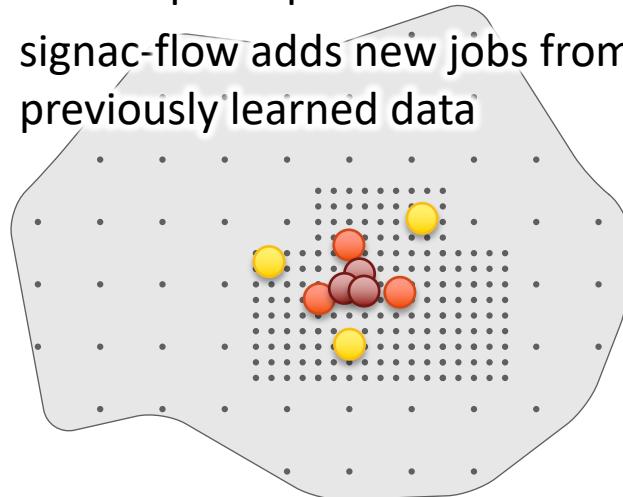


Future goal:

Automated Machine Learning Workflows

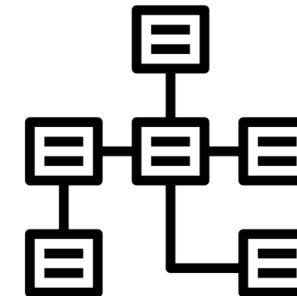
From **dynamic** workflows...

- signac data spaces are suited for complex optimization tasks
- signac-flow adds new jobs from previously learned data



... to distributed, iterative **exploration**

- prioritized exploration in high dimensional spaces
- distribution across systems



Thank you!

signac is a product of The Glotzer Group at the University of Michigan. Development and deployment supported by MICCoM, as part of the Computational Materials Sciences Program funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under Subcontract No. 6F-30844. Project conceptualization and implementation supported by the National Science Foundation, Award # DMR 1409620.

<https://signac.io>

