

Pachyderm at DIANA/HEP

Daniel Whitenack, [@dwhitena](https://twitter.com/dwhitena)

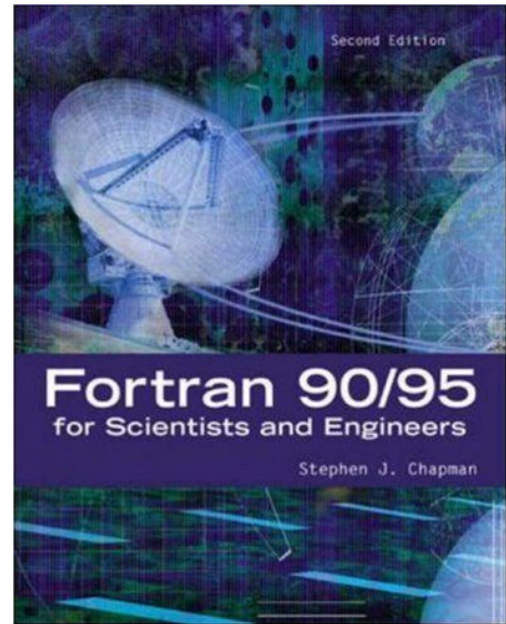
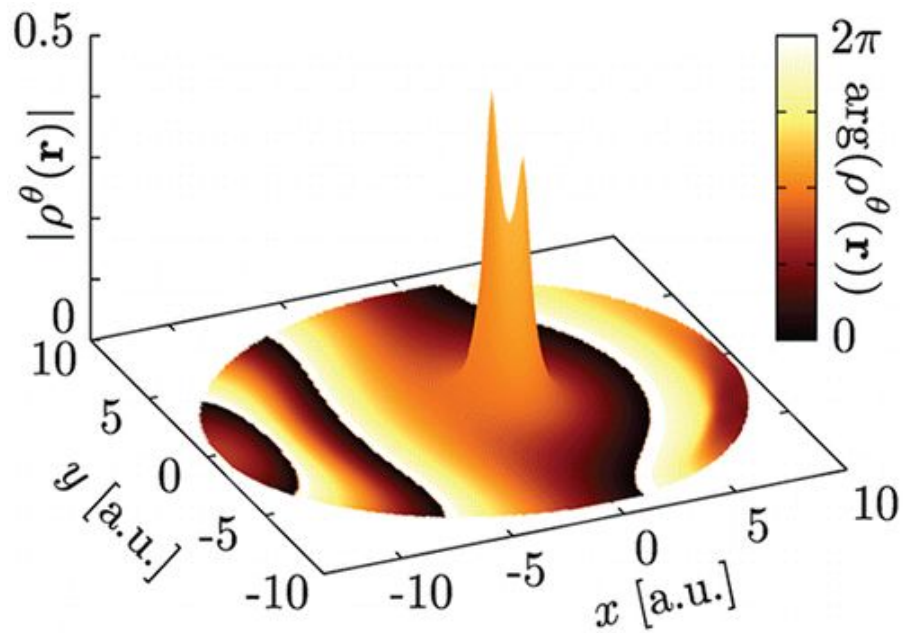
Data scientist, developer advocate, [@pachydermIO](https://twitter.com/pachydermIO)

Data analysis instructor, [@ardanlabs](https://twitter.com/ardanlabs)

Outline

1. A physics story
2. Intro to Pachyderm
3. Example use, details
4. Research use case
5. Questions

A Physics Story



Reproducibility and Collaboration?

generate_PRL_figs_good.m

second_try_H2.py

first_try_H2_with_low_conf3.m

H2_after_fix.conf

geo_opt_funcs.m

complex_scaling_patch.f90

Scaling and Parallelization?

“Both the bound 1-electron and metastable 2-electron complex densities are found exactly with finite differences...” -me, arXiv:1111.1934

Scaling and Parallelization?

“Both the bound 1-electron and metastable 2-electron complex densities are found exactly with finite differences...” -me, arXiv:1111.1934

Translation: “I’m not going to spend time making this scalable (as if I could). I just want to demonstrate the physics for this paper with a simple solution.”

(teaser)

Pachyderm solves these issues...

Intro to Pachyderm

A project which helps organizations/teams better manage data infra and distributed processing.

A project which helps organizations/teams better manage data infra and distributed processing.

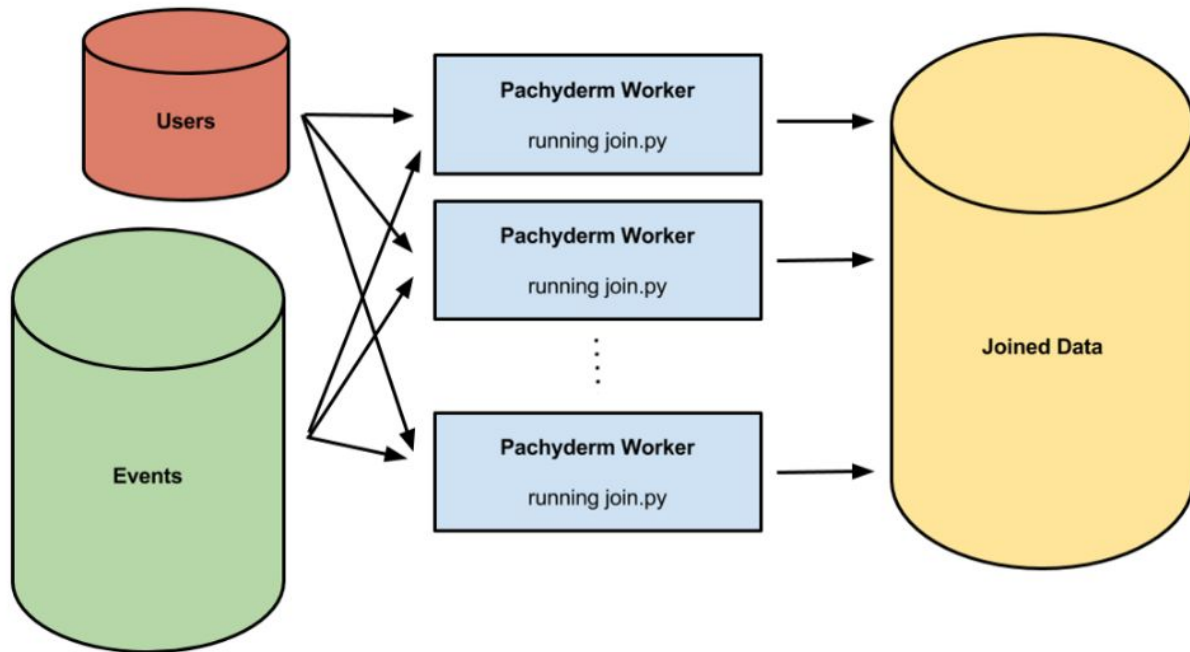
An open source, distributed processing and data versioning framework built on containers.

An open source, distributed processing and data versioning framework built on containers.

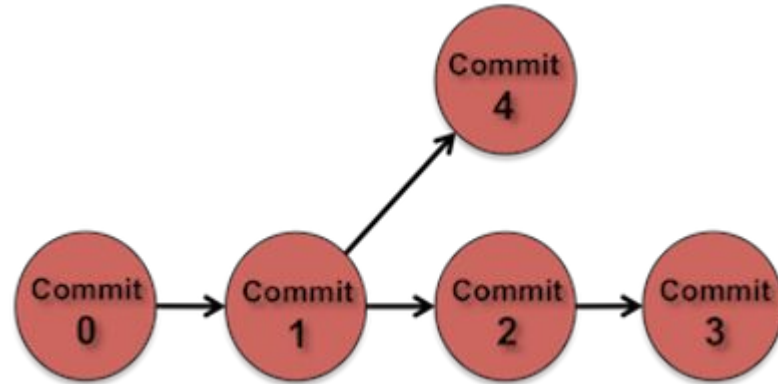
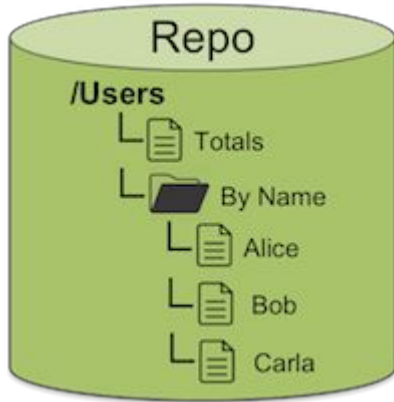
The screenshot shows the GitHub repository page for pachyderm/pachyderm. At the top, there is a search bar and navigation links for Pull requests, Issues, Gist, and ToDo. Below this, the repository name 'pachyderm / pachyderm' is displayed along with statistics: 97 Unwatch, 1,502 Unstar, and 161 Fork. A navigation bar includes links for Code, Issues (246), Pull requests (14), Boards, Reports, Projects (0), and Wiki. The repository description is 'Containerized Data Analytics' with the URL 'http://pachyderm.io'. Below the description are tags for analytics, big-data, containers, data-analysis, data-science, distributed-systems, docker, go, kubernetes, and pachyderm. A summary bar shows 6,266 commits, 338 branches, 605 releases, and 42 contributors. Below this are buttons for 'New pull request', 'Create new file', 'Upload files', 'Find file', and 'Clone or download'. The commit history is shown as a table with columns for the commit author, description, and time since the commit.

Commit	Description	Time
JoeyZwicker committed on GitHub Merge pull request #1294 from svanharmelen/b-docs		Latest commit 2b144f8 21 hours ago
contributing	Update setup.md	4 months ago
design	Fix persist.proto link	24 days ago
doc	Fix links to examples in documentation	2 days ago
etc	etc: update proto container	23 days ago
migration	fix	2 months ago
src	Add support for minio deploy.	6 days ago
.dockerignore	Docker ignore generate.	7 months ago

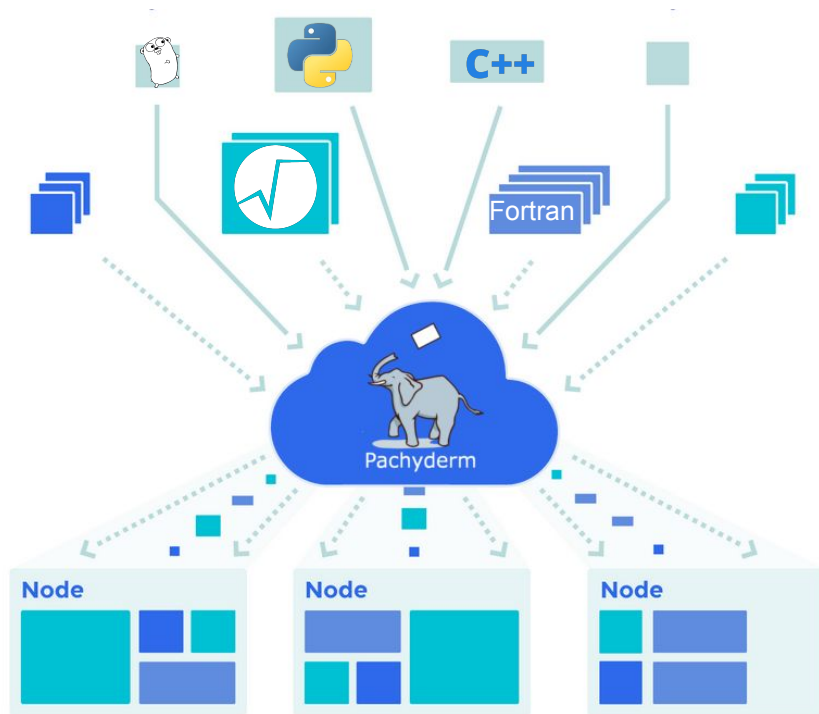
An open source, distributed processing and data versioning framework built on containers.



An open source, distributed processing and data versioning framework built on containers.



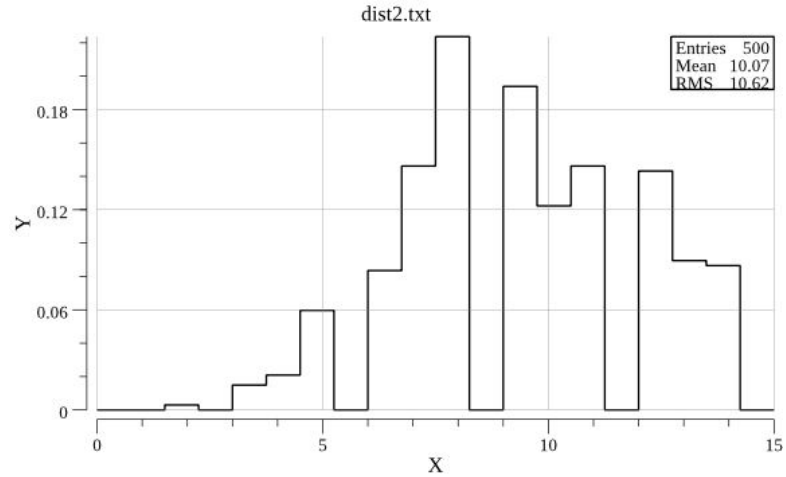
An open source, distributed processing and data versioning framework built on containers.



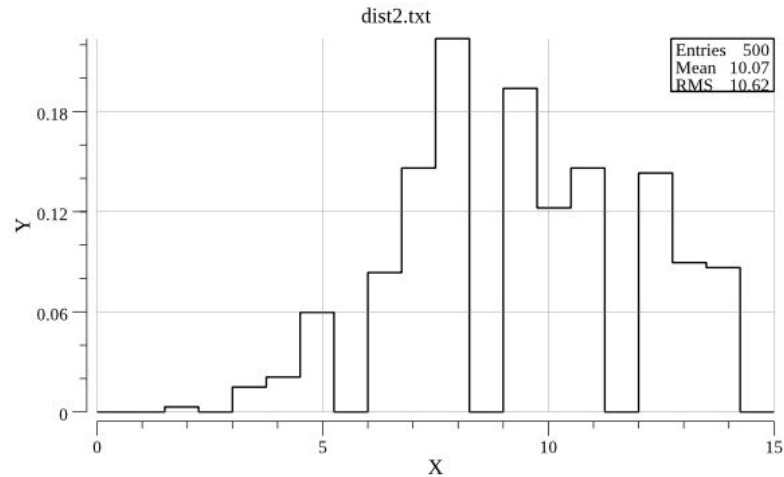
Example use, details

Let's say we want:

Let's say we want:



Let's say we want:

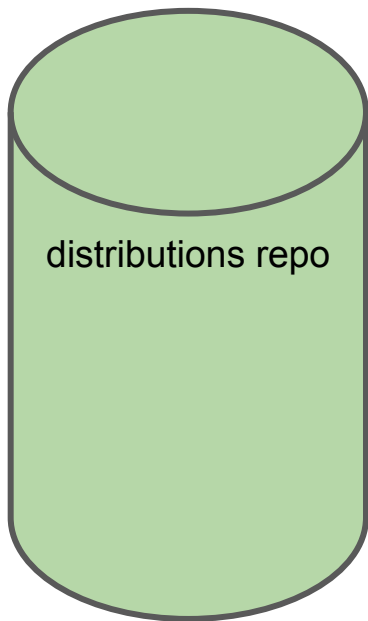


Lot's of them, for many different experiments, runs, etc...

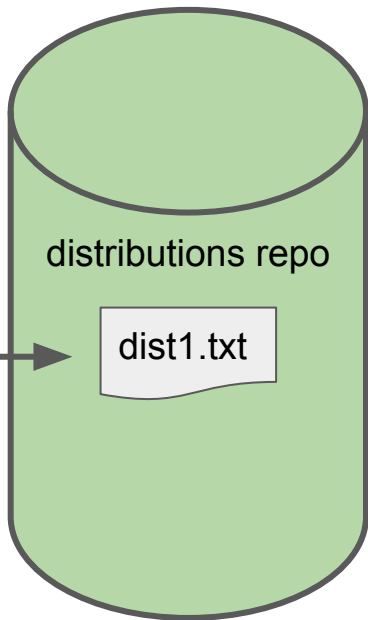


Pachyderm

Pachyderm

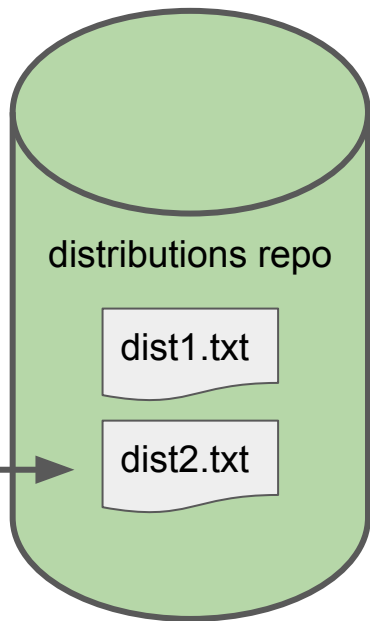


Pachyderm



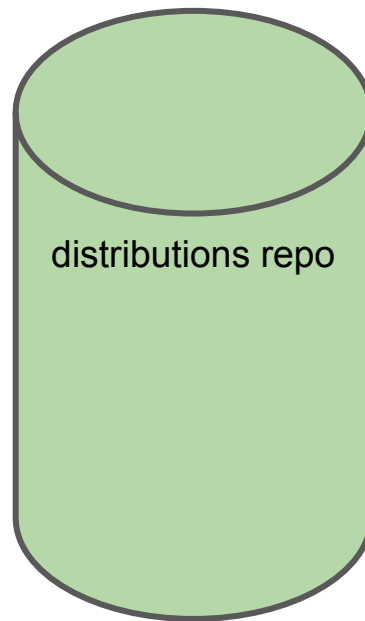
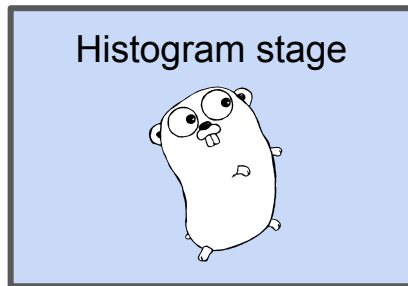
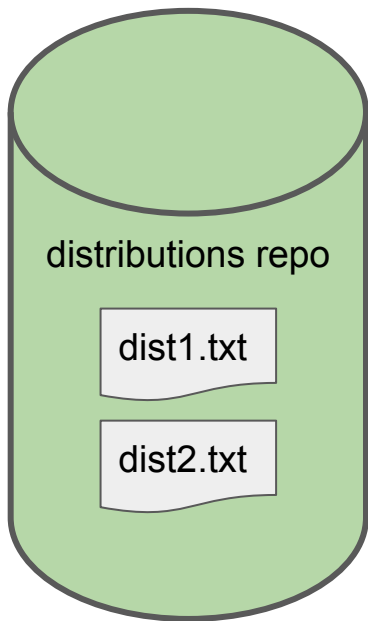
"Commit" 0

Pachyderm

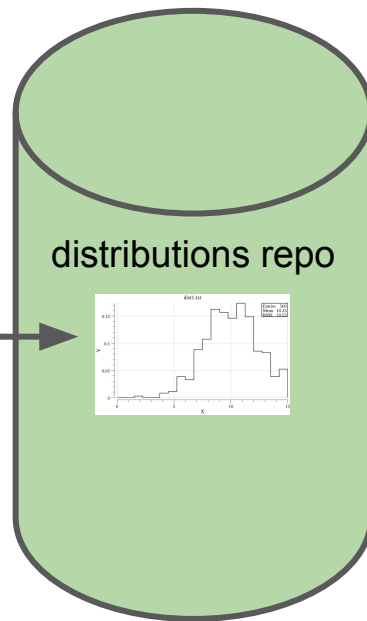
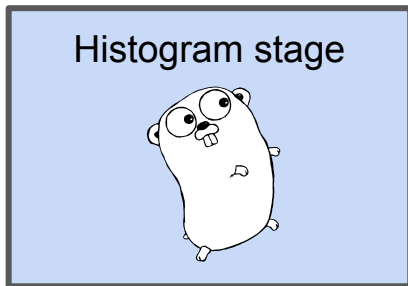
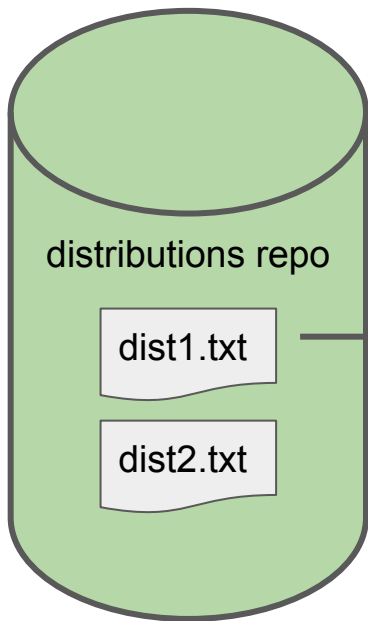


“Commit” 1

Pachyderm

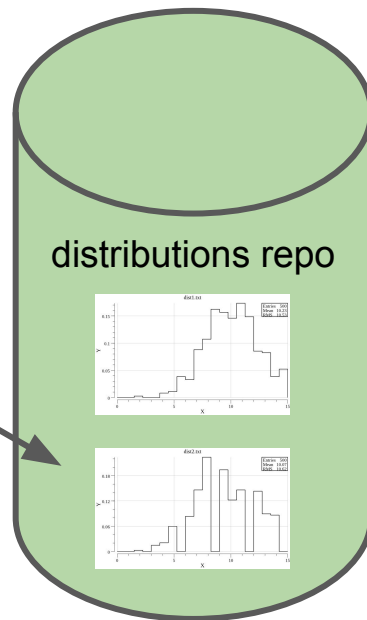
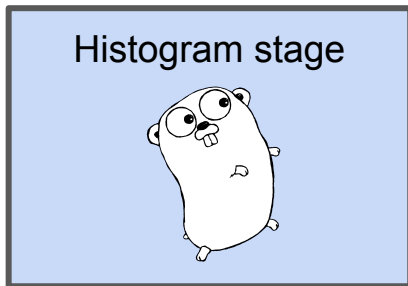
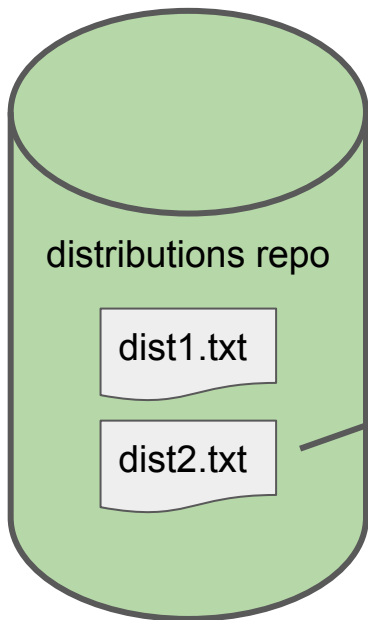


Pachyderm



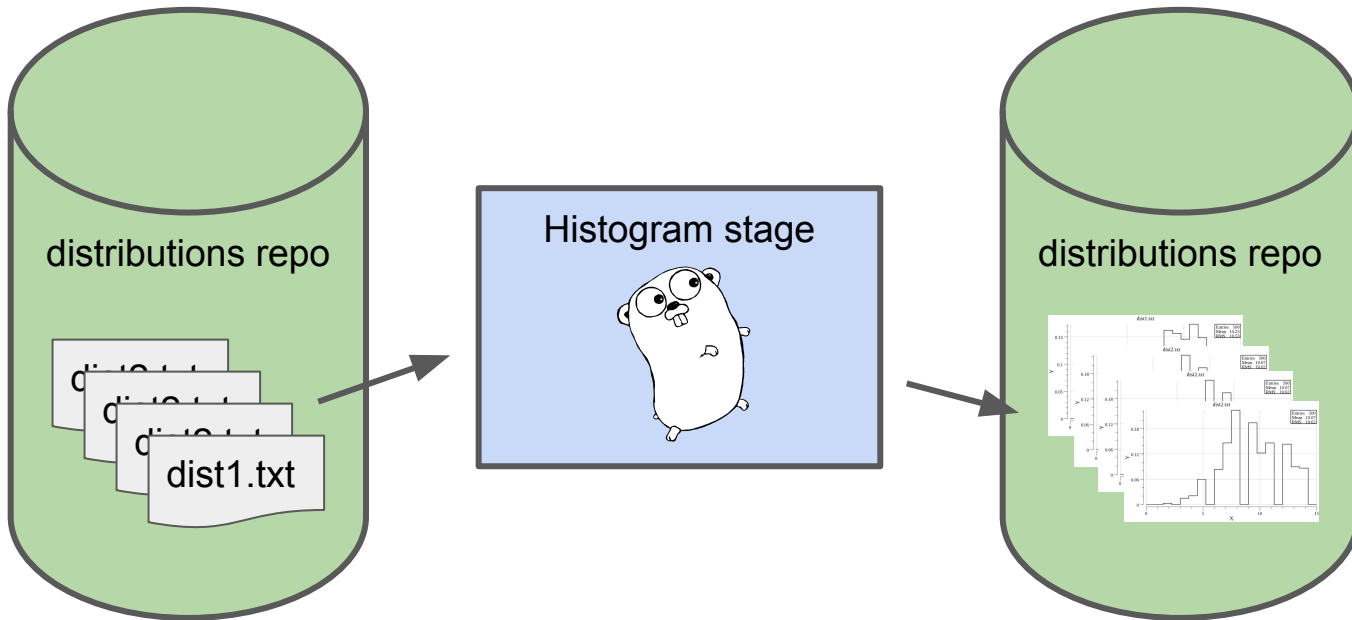
"Commit" 0

Pachyderm

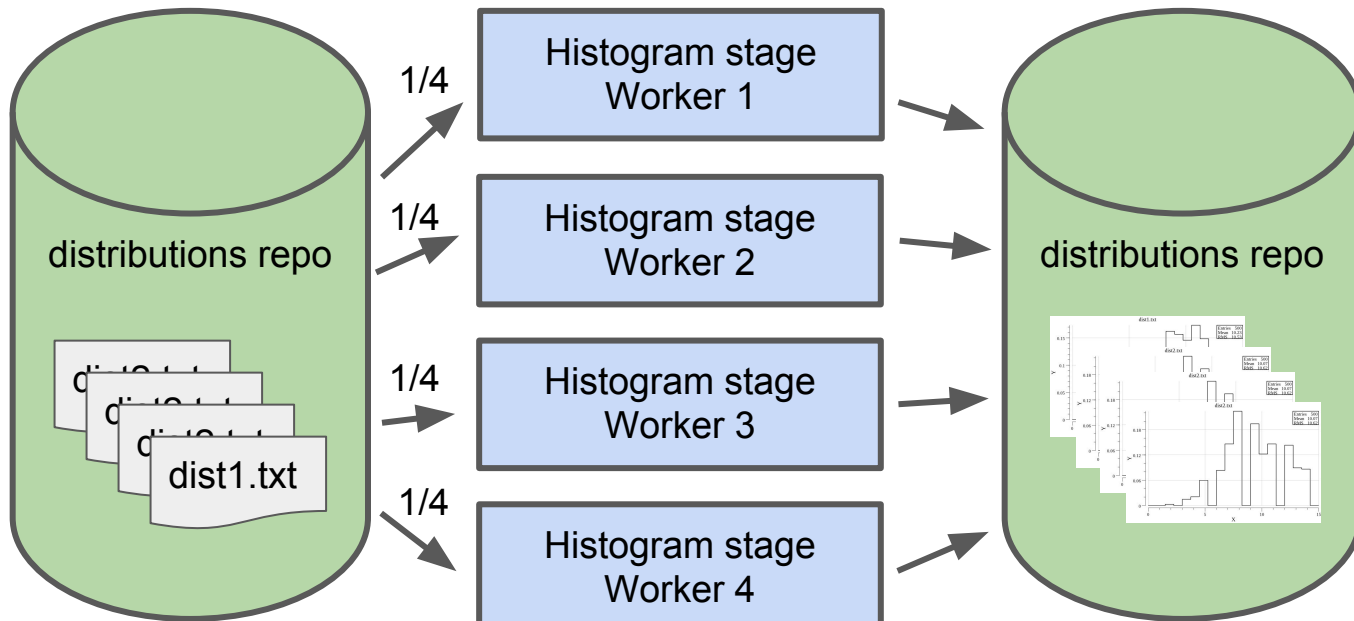


"Commit" 1

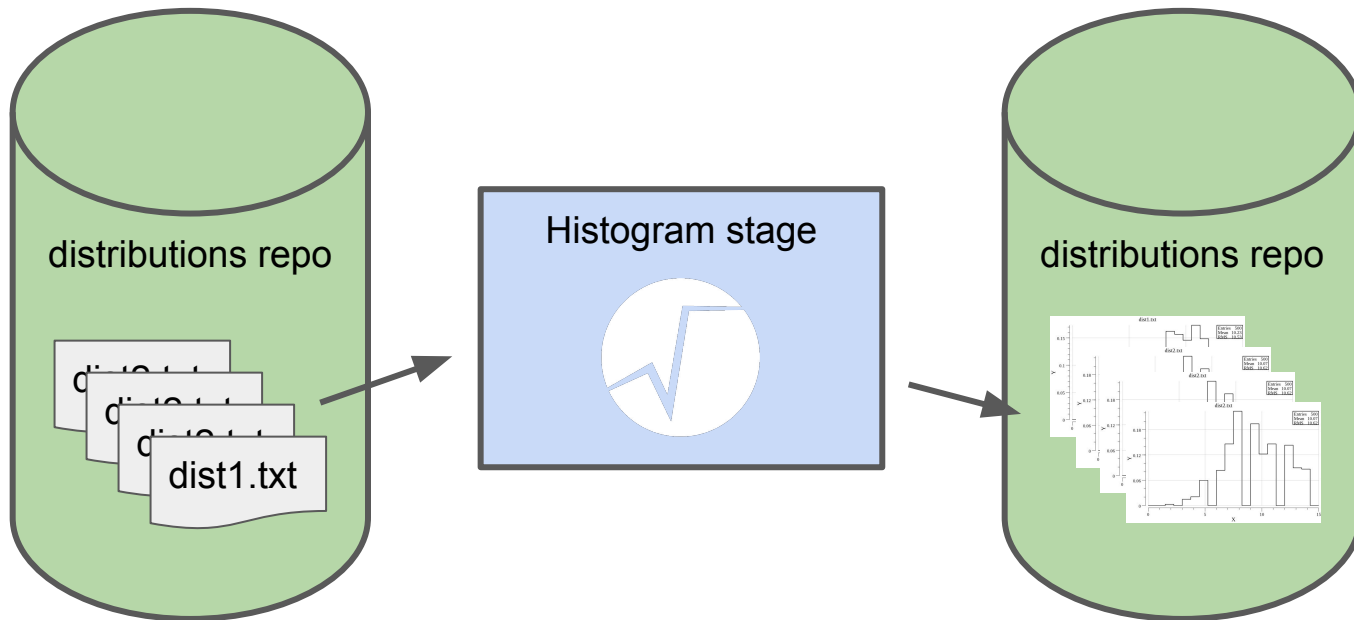
Pachyderm



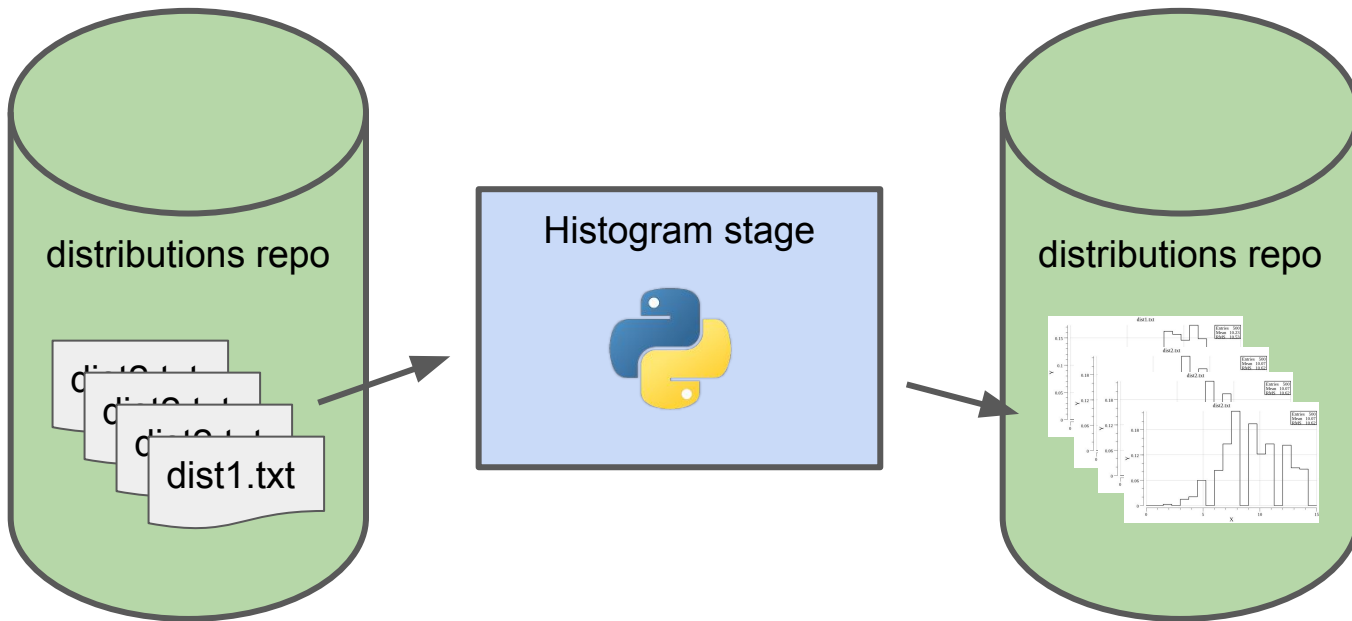
Pachyderm



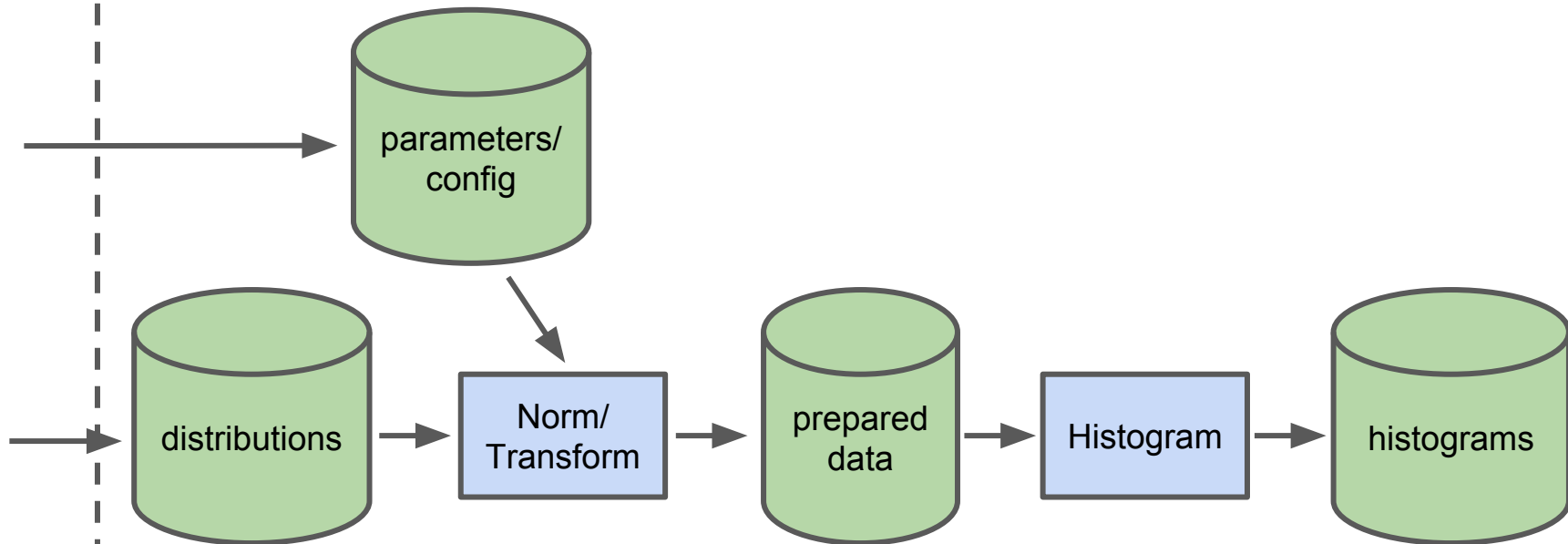
Pachyderm



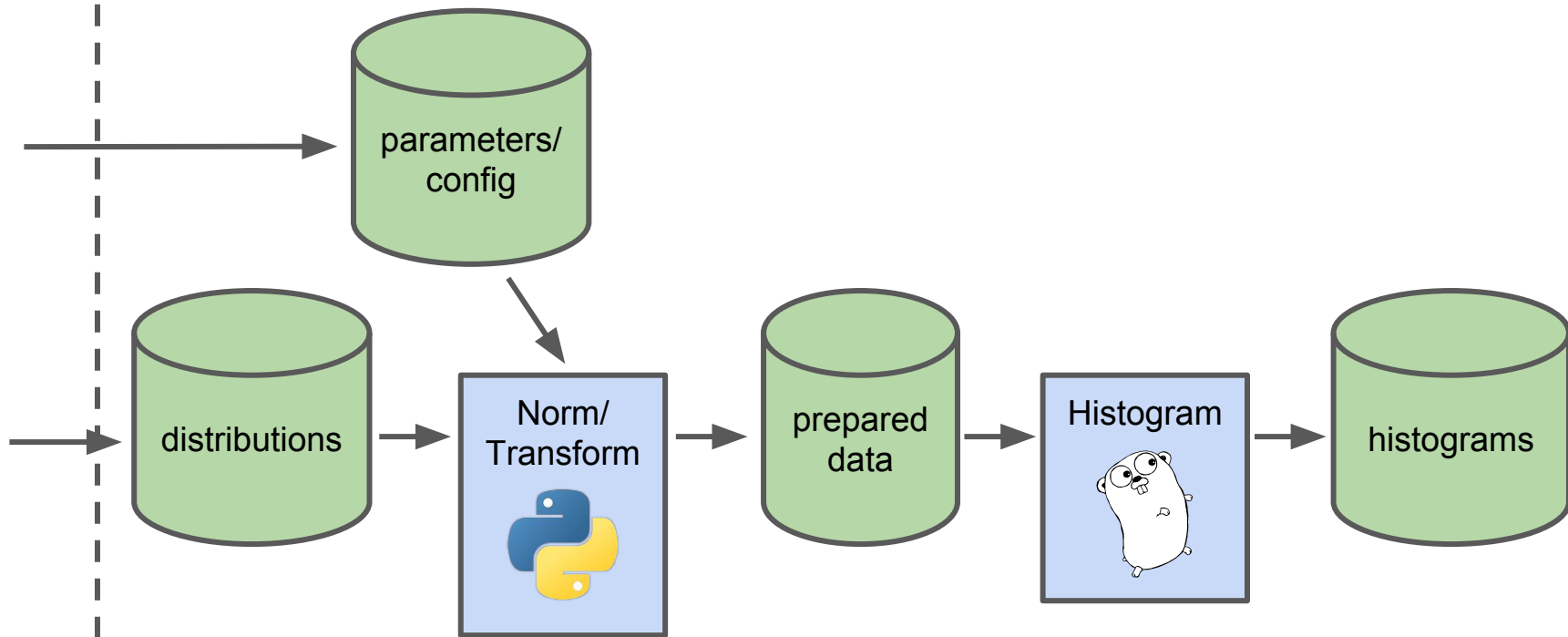
Pachyderm



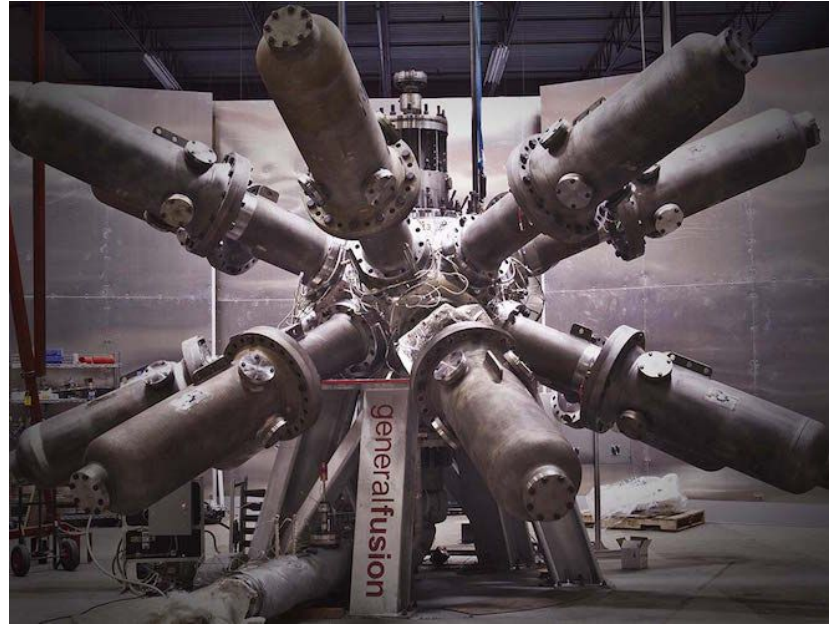
Pachyderm



Pachyderm



Research Use Case



generalfusion

Questions?

Resources

- Join the [Pachyderm Slack channel](#)
- Check out the [Pachyderm docs](#)
- Chat with me [@dwhitena](#)
- Try out [the simple example](#)