Scalable, Distributed, and Reproducible Machine Learning

Daniel Whitenack, @dwhitena
Data Scientist and Advocate, @pachydermIO

Outline

1. ML/AI challenges
2. Demo with Python and Pachyderm
3. Resources

Reproducibility

generate_figs_good.py

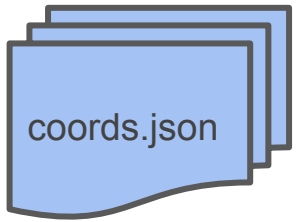
second_try_user_events.py

first_try_with_low_conf3.ipynb

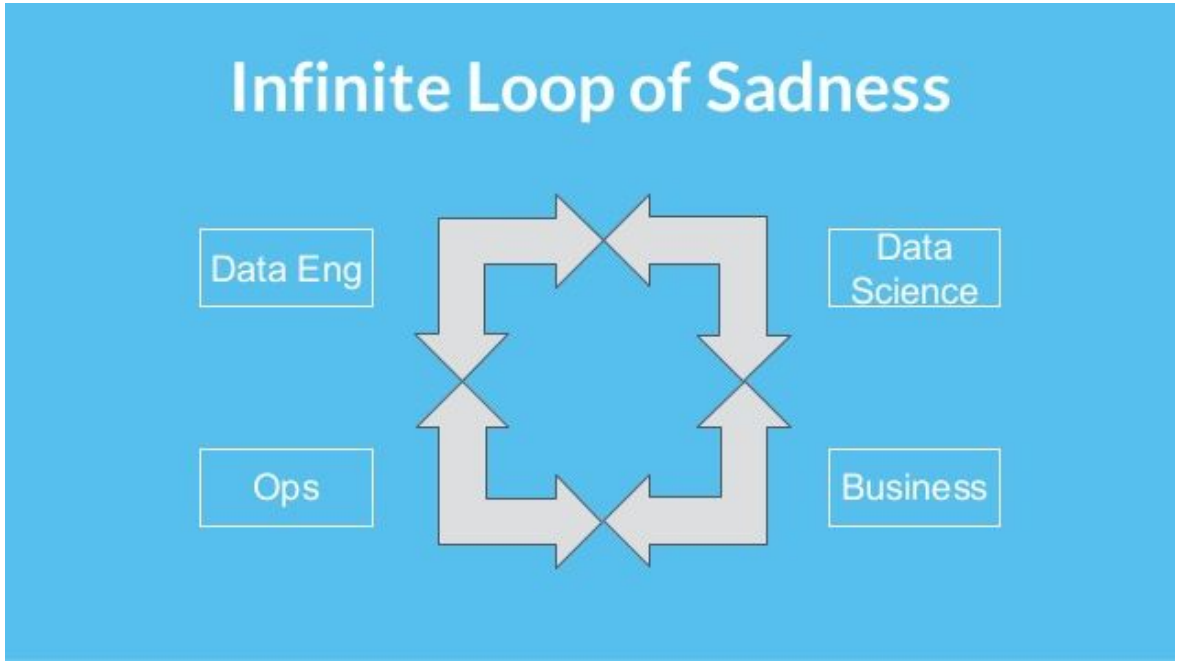
join_after_fix.json

opt_funcs_2016-11-12.m

scaling_patch.cpp

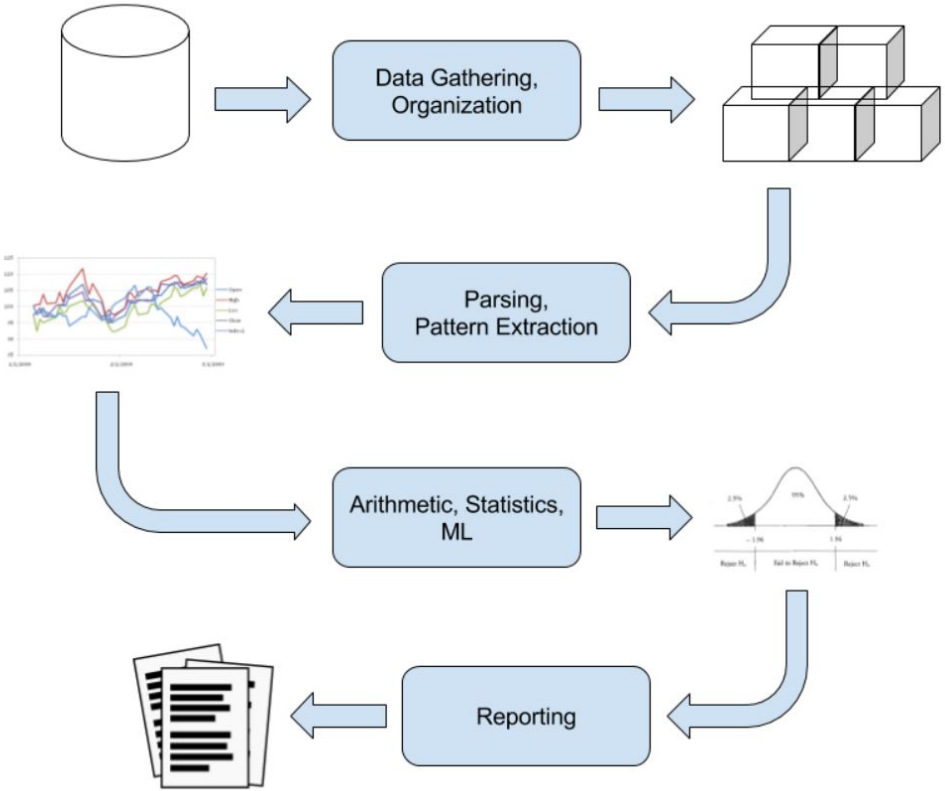


Workflow Management

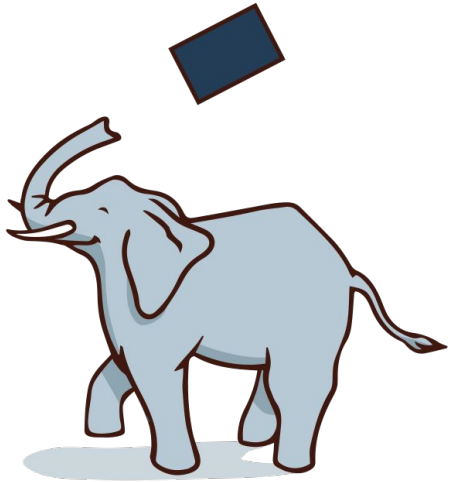


From Josh Wills

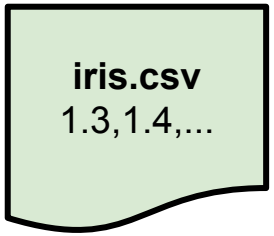
Audit Trails, Debugging, Maintenance



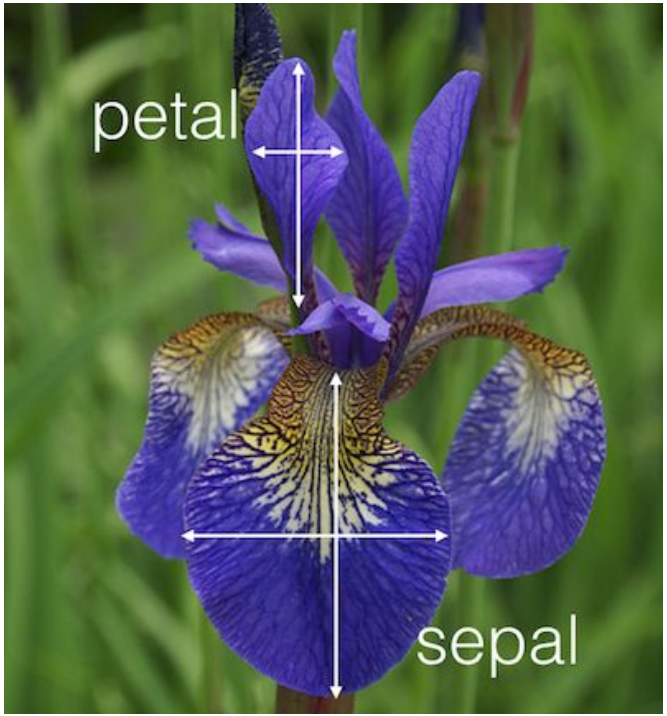
@dwhitena



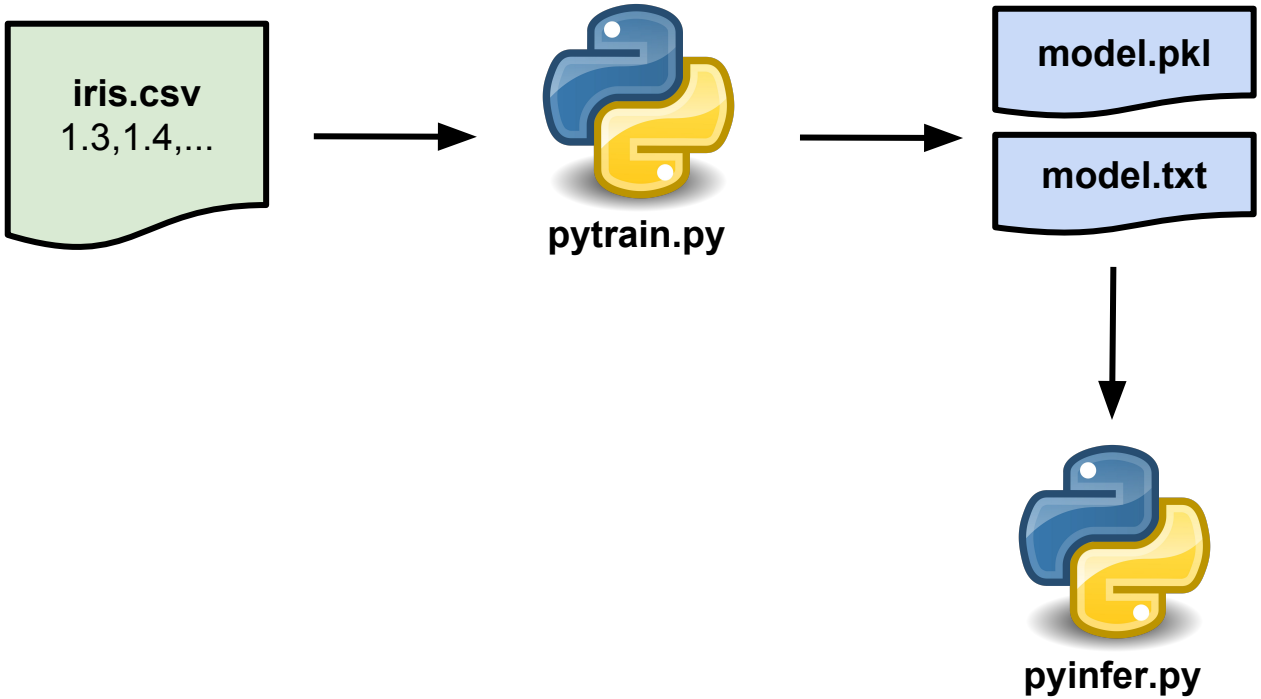
Demo Time

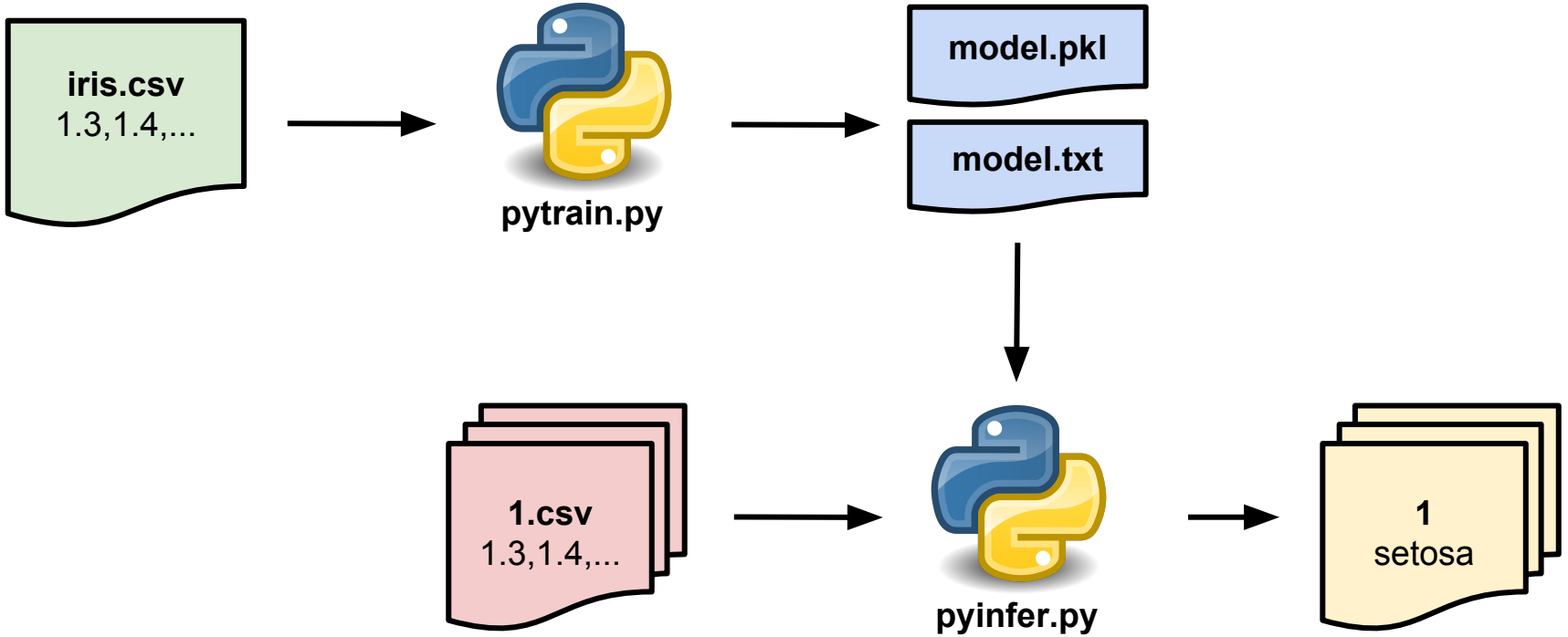

iris.csv
1.3,1.4,...

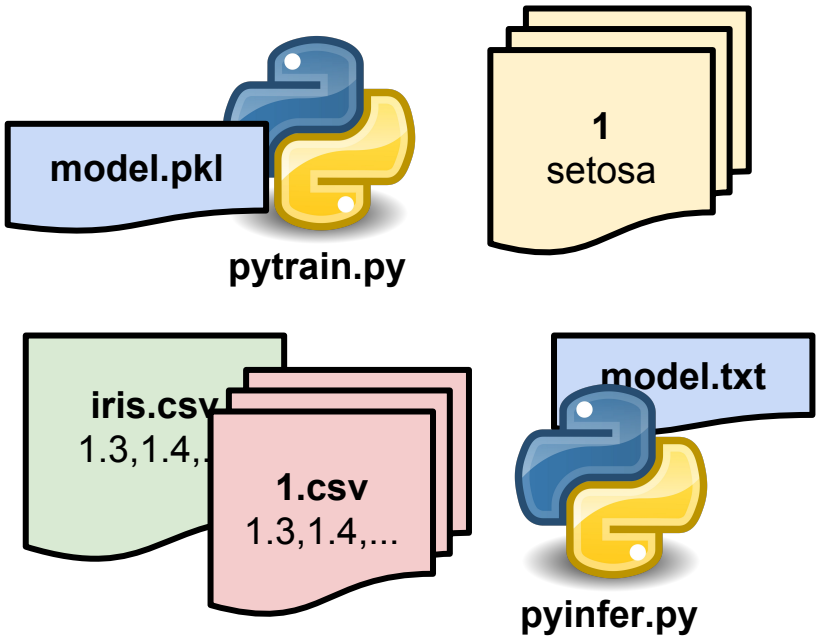
iris.csv
1.3,1.4,...

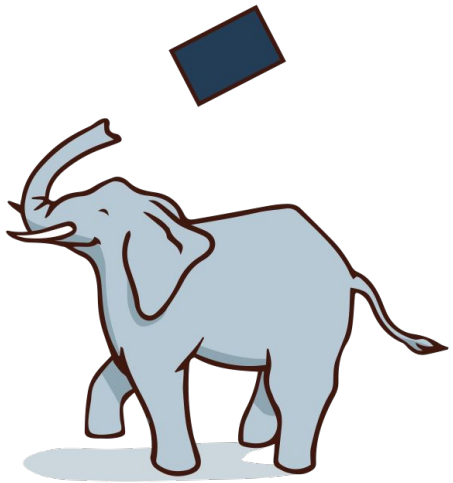



iris.csv
1.3,1.4,...
pytrain.py
model.pkl
model.txt


iris.csv
1.3,1.4,...
pytrain.py
model.pkl
model.txt
pyinfer.py


iris.csv
1.3,1.4,...
pytrain.py
model.pkl
model.txt
1.csv
1.3,1.4,...
pyinfer.py
1
setosa

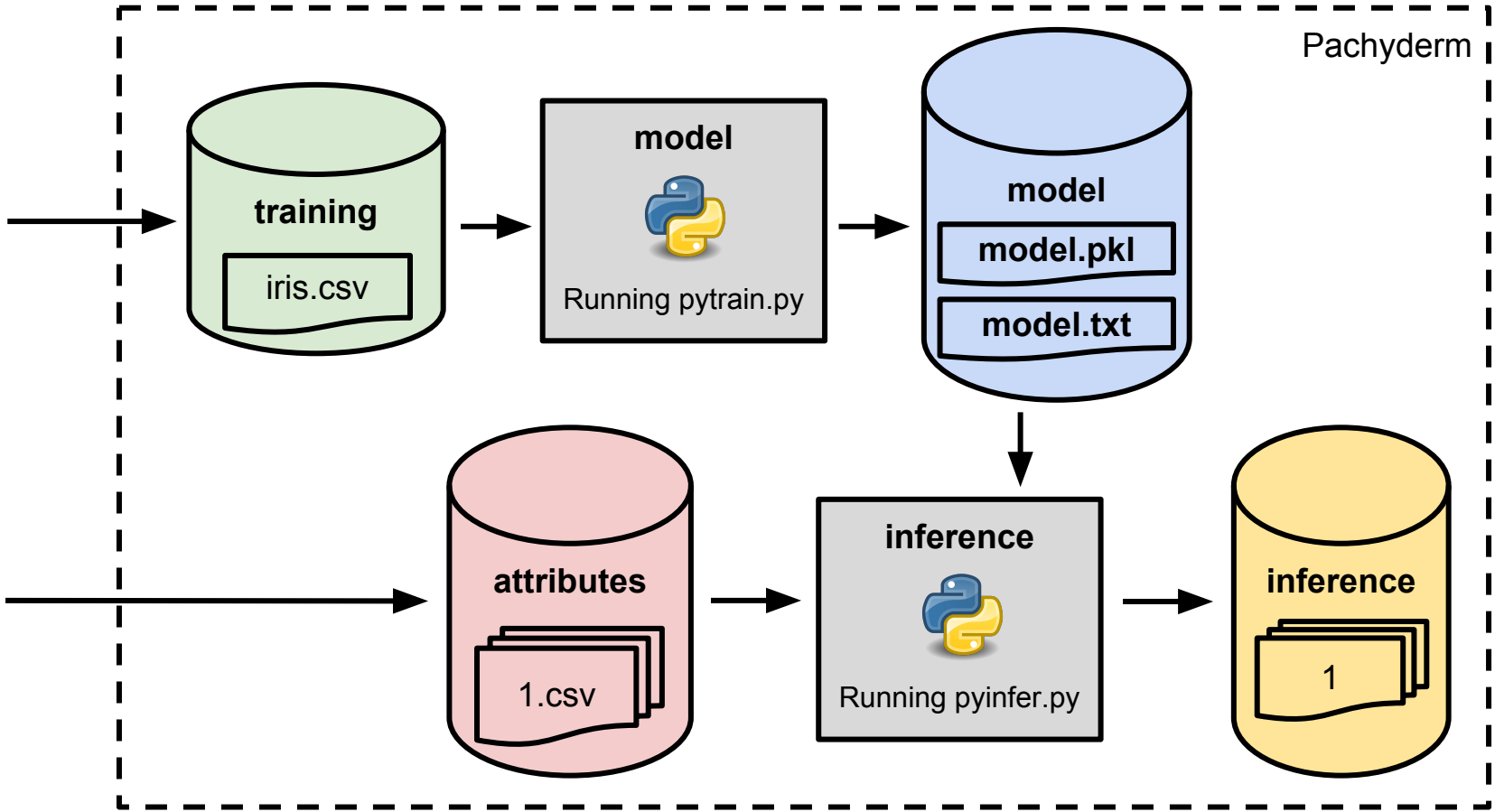

model.pkl
pytrain.py
1
setosa
iris.csv
1.3,1.4,..
1.csv
1.3,1.4,...
model.txt
pyinfer.py

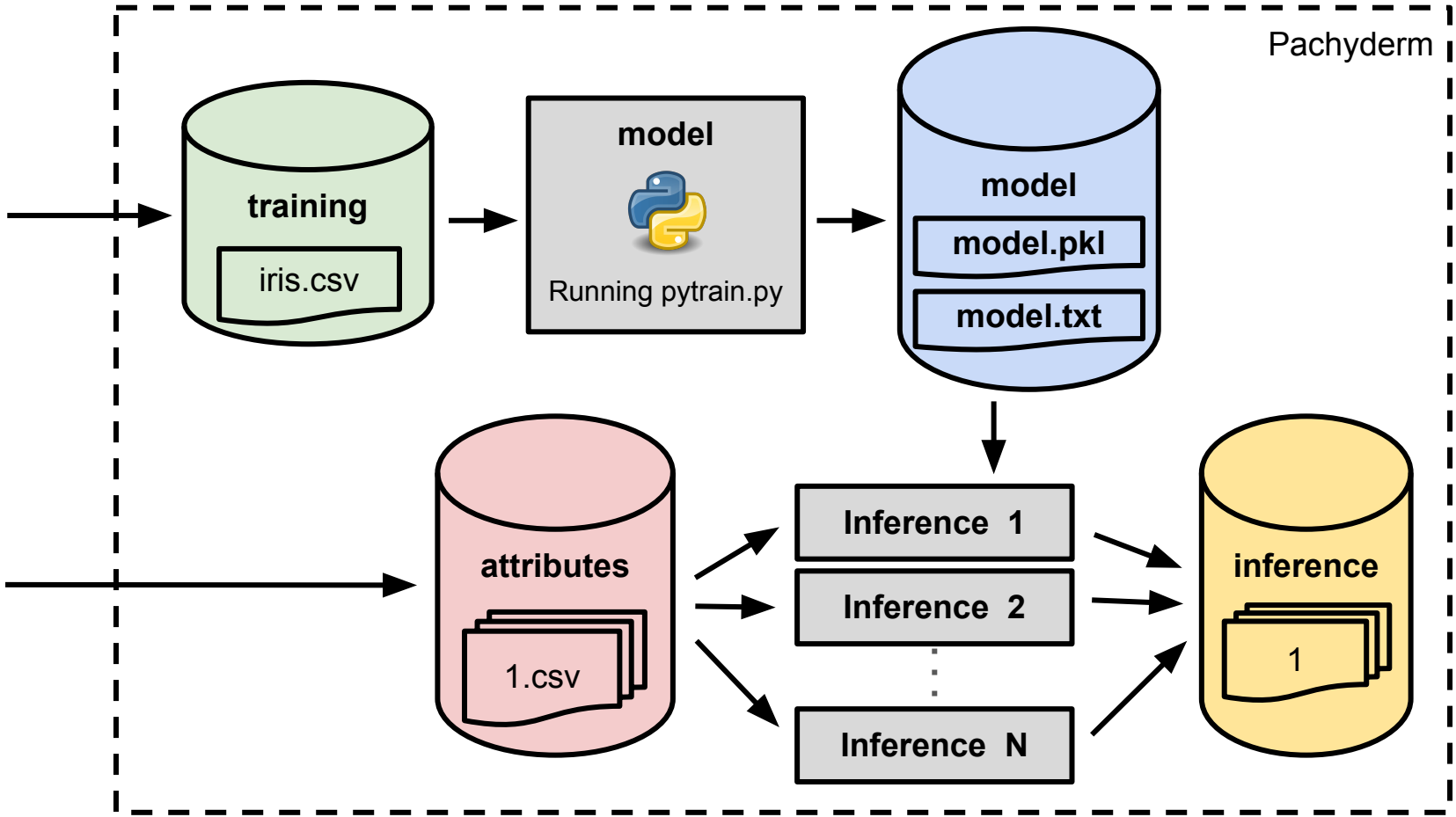


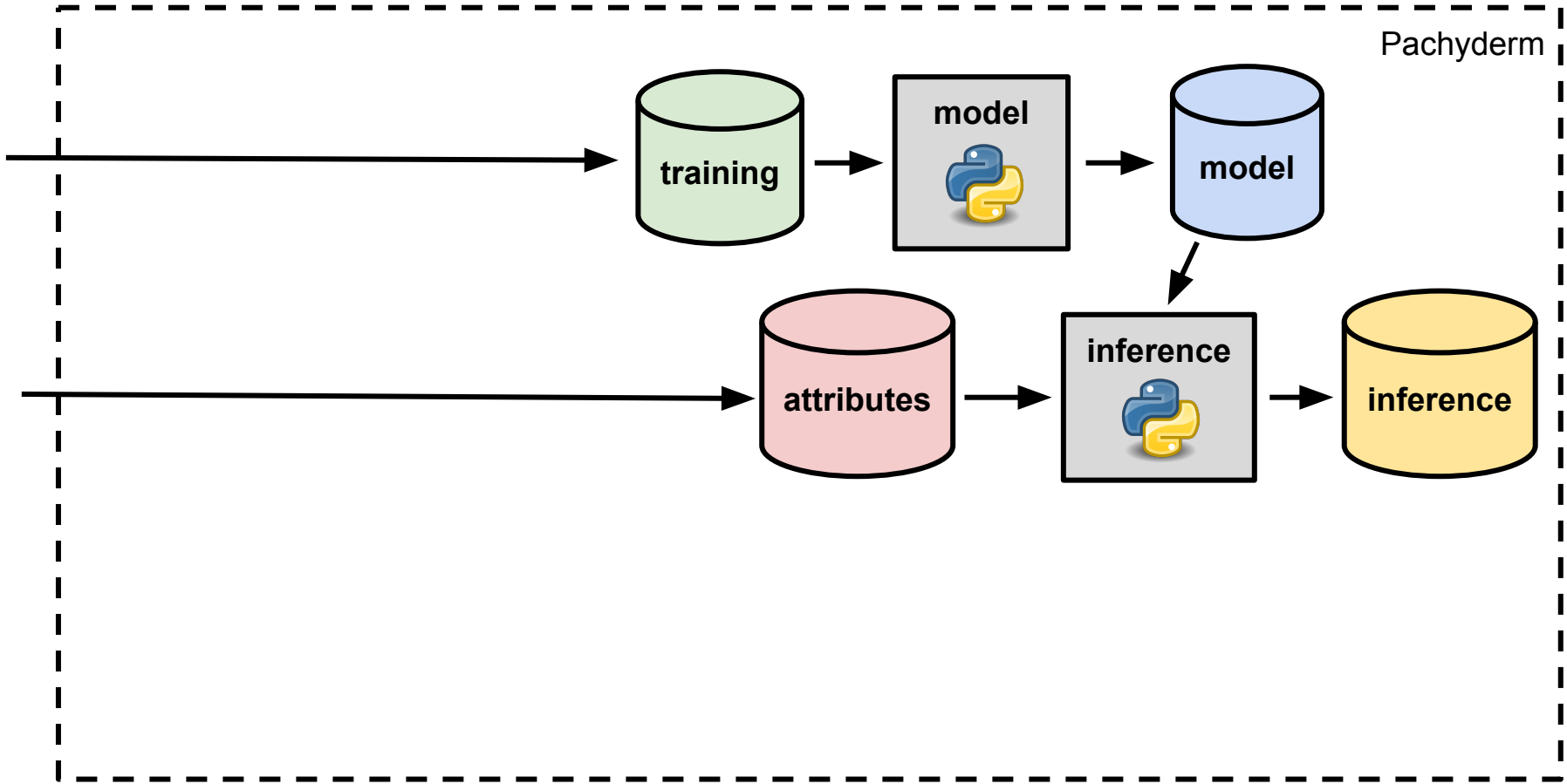
… enter Pachyderm

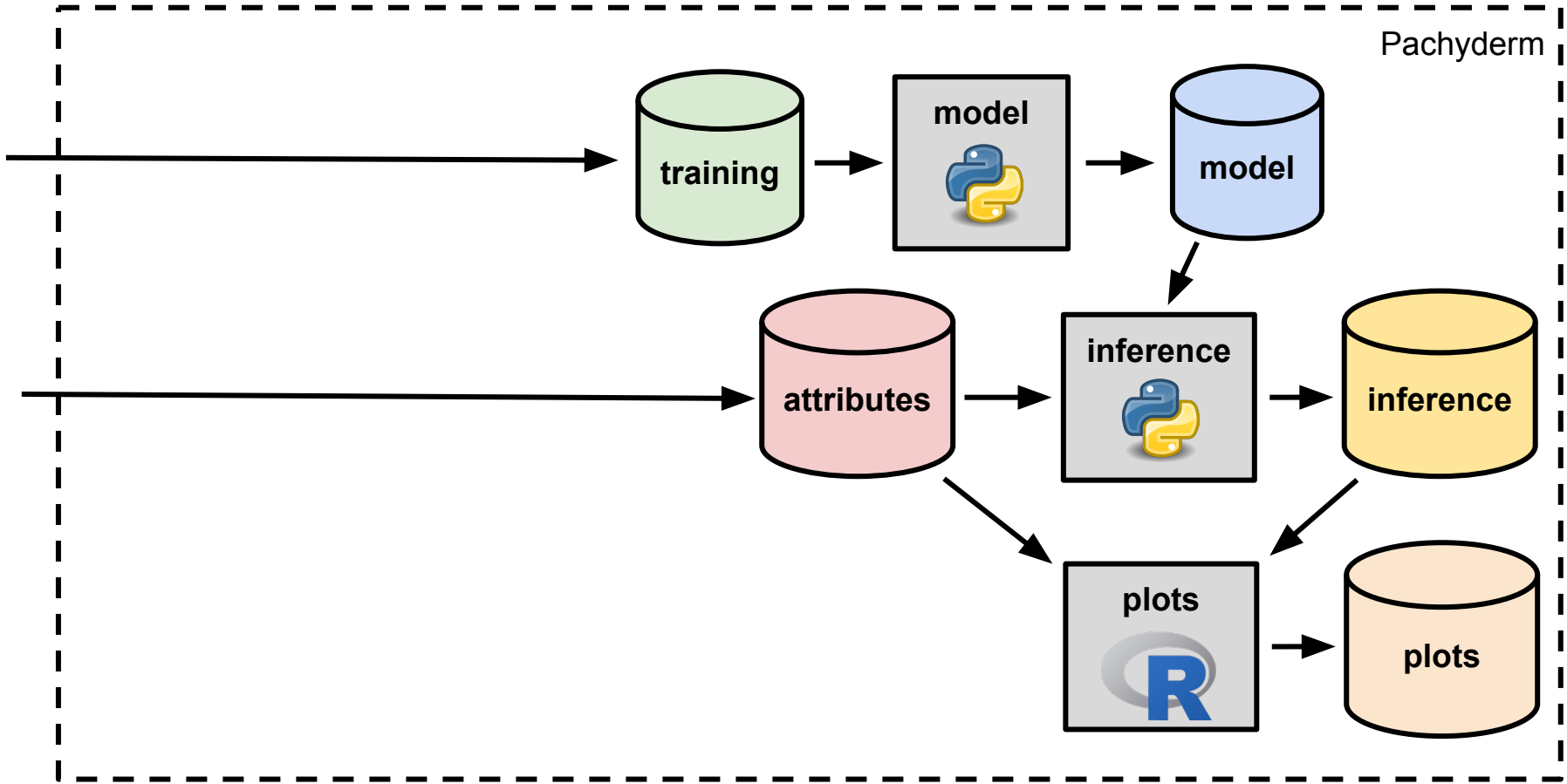
An open source, distributed processing and data versioning framework built on containers.

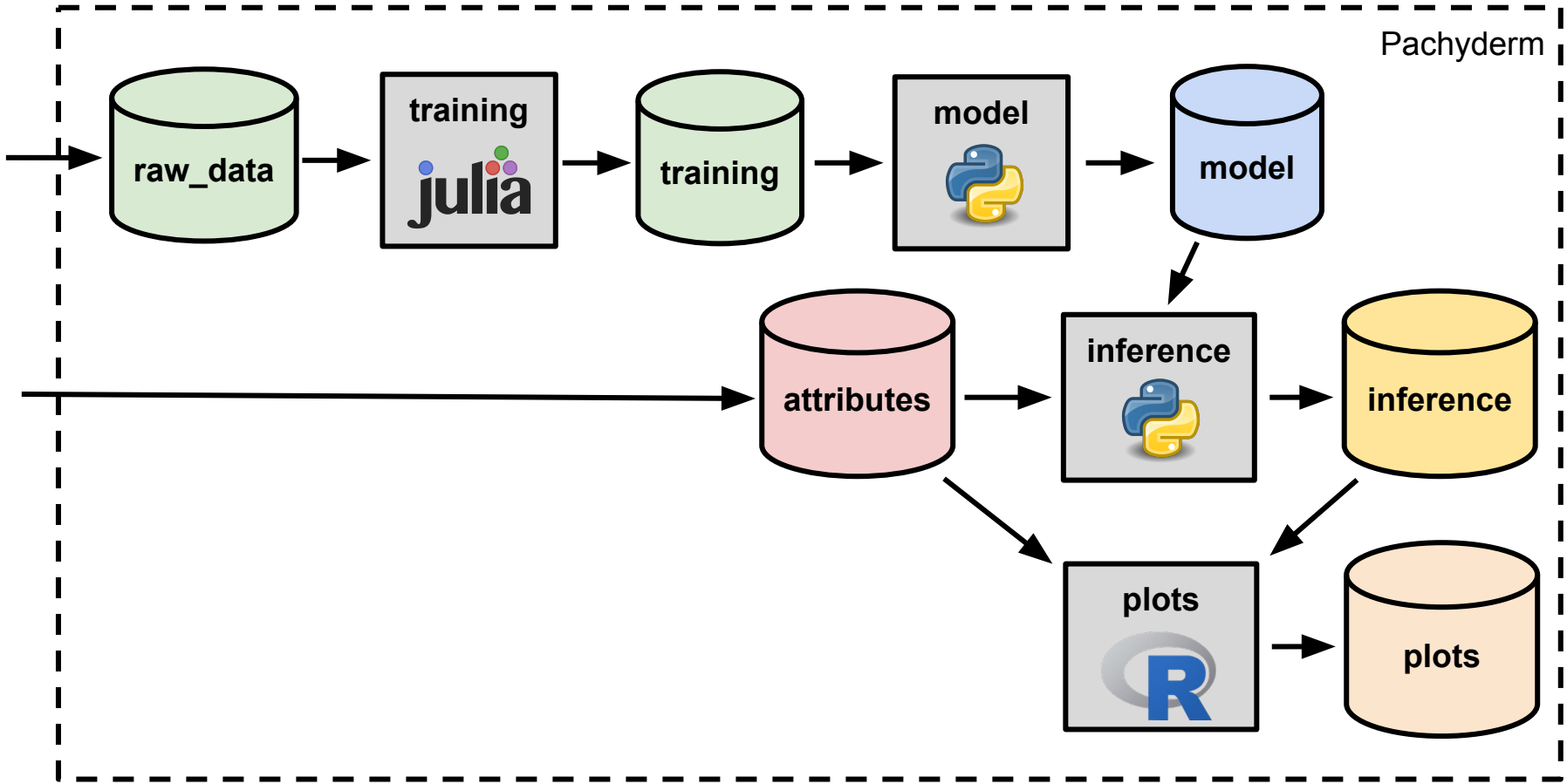
@dwhitena, @PyDataAnnArbor

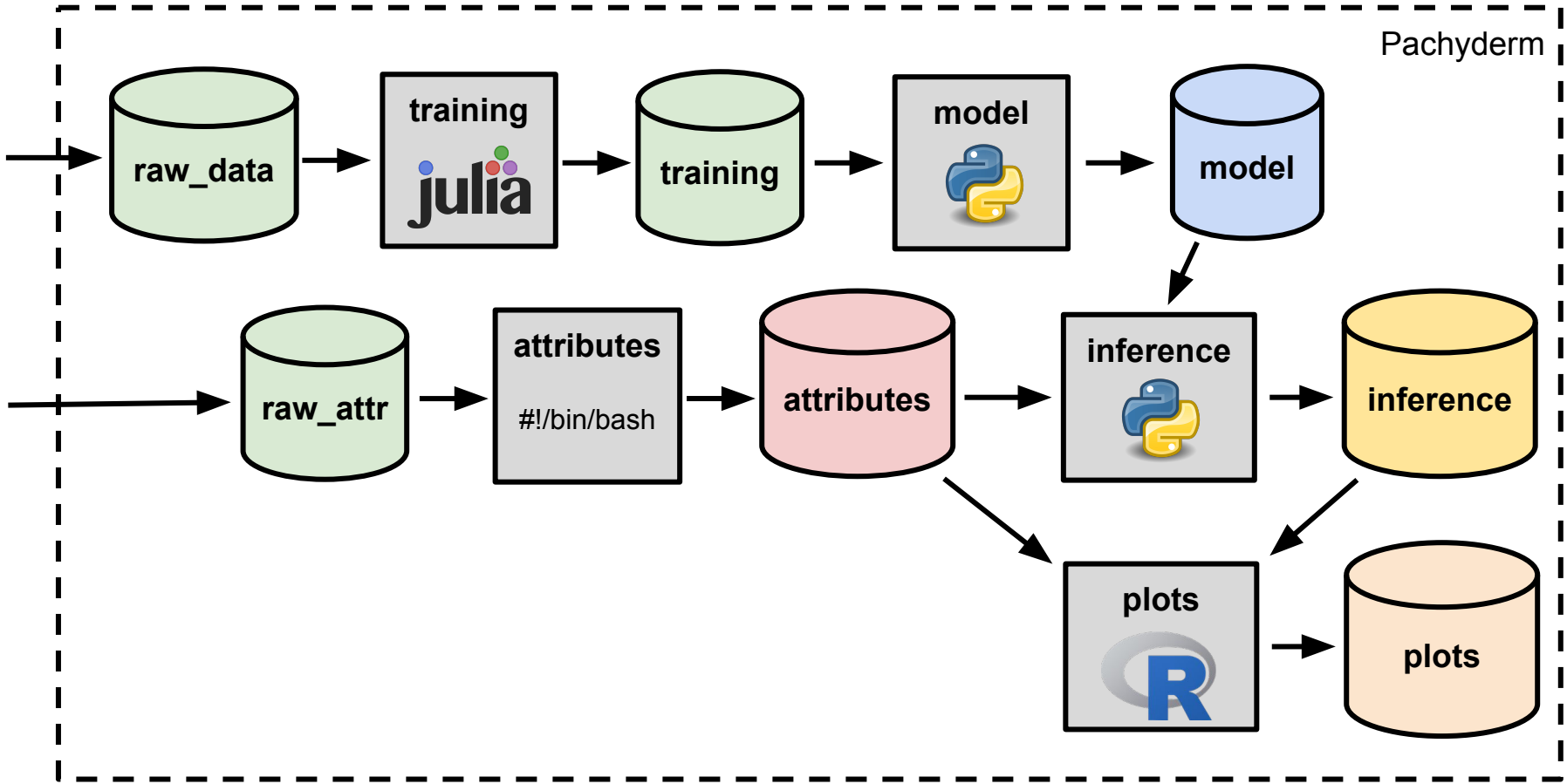

Pachyderm
training
iris.csv
model
Running pytrain.py
model
model.pkl
model.txt
attributes
1.csv
inference
Running pyinfer.py
inference
1

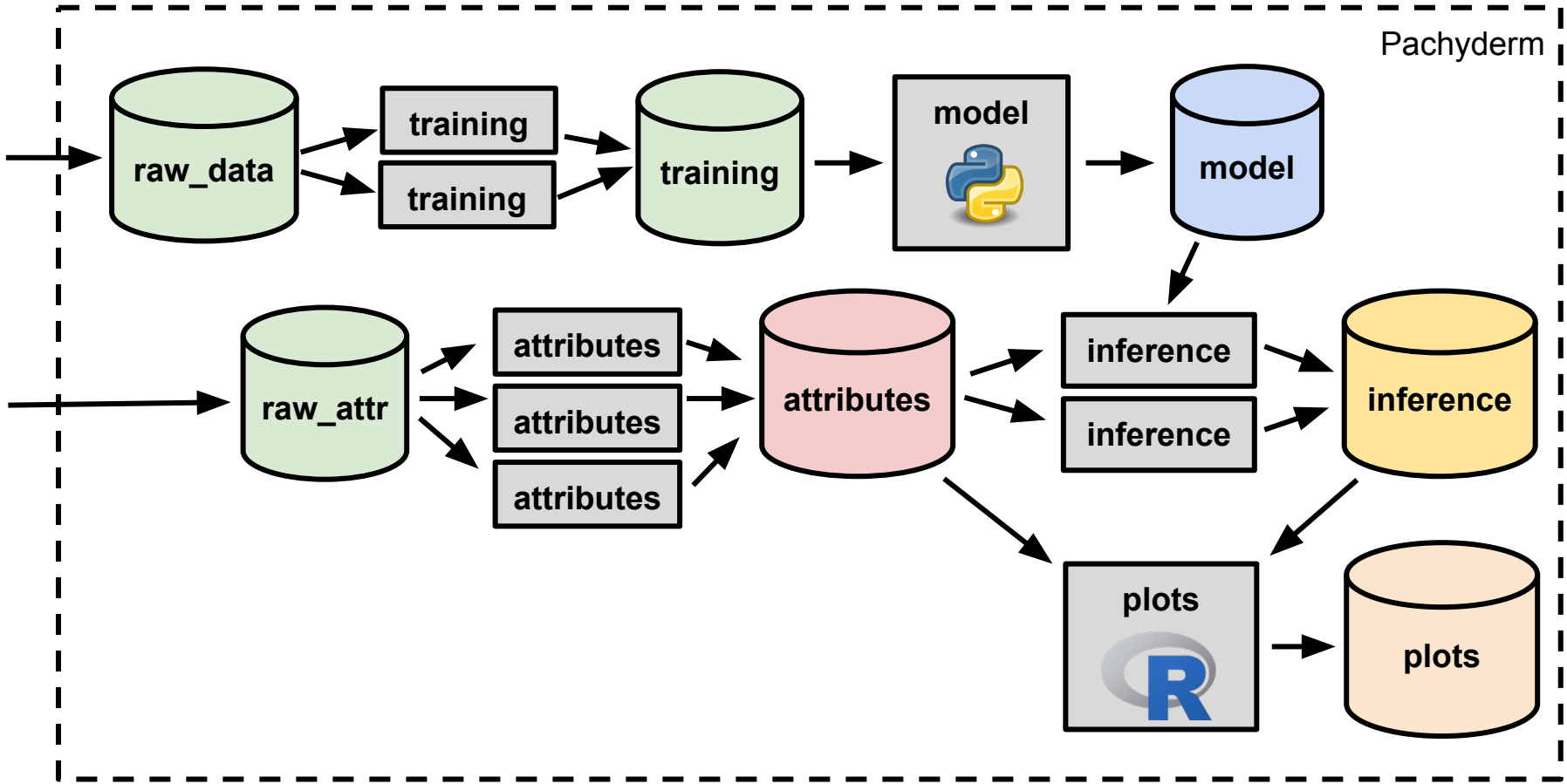

Pachyderm
training
iris.csv
model
Running pytrain.py
model
model.pkl
model.txt
attributes
1.csv
Inference 1
Inference 2
Inference N
inference
1


Pachyderm
training
model
model
attributes
inference
inference


Pachyderm
training
model
model
attributes
inference
inference
plots
plots


Pachyderm
raw_data
training
julia
training
model
model
attributes
inference
inference
plots
R
plots


Pachyderm
raw_data
training
julia
training
model
model
raw_attr
attributes
#!/bin/bash
attributes
inference
inference
plots
R
plots



@dwhitena, @PyDataAnnArbor

Conclusion/Resources

- Run the code/pipeline
- Run other ML examples
- Join the Pachyderm Slack channel
- Check out the Pachyderm docs
- Slack/tweet me @dwhitena
- Read a related article