



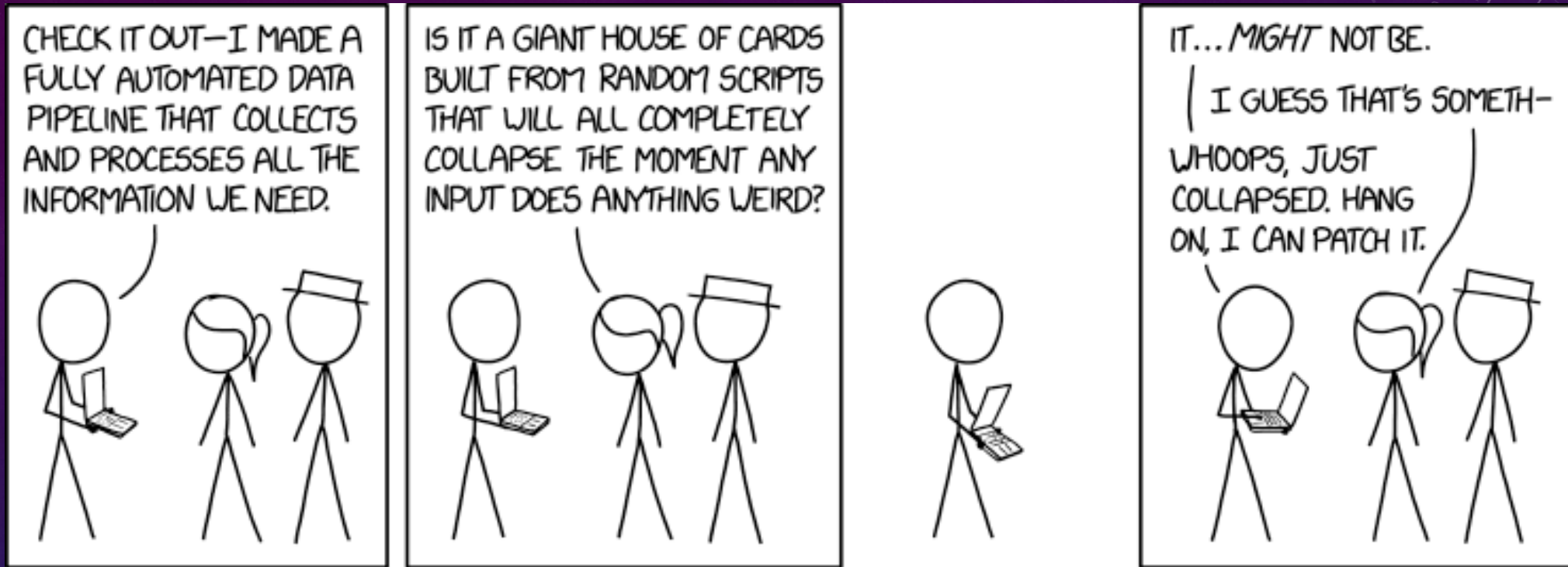
DATA ANALYSIS PIPELINE MANAGEMENT WITH ZIGGY

PETER TENENBAUM

SC 2023, DENVER



xkcd.com. Used under terms of Creative Commons License 2.5.



Science data pipelines need to do a lot more than science:

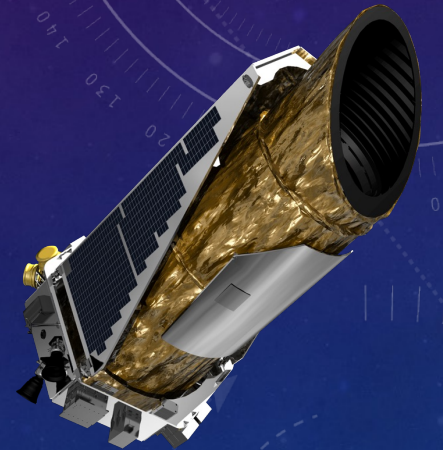
- Logging
- Execution flow
- Execution monitoring
- And much more!
- Data accountability
- Configuration management
- Error handling



ziggy

Ziggy handles all the Not-Science and lets the scientists get on with the Science!

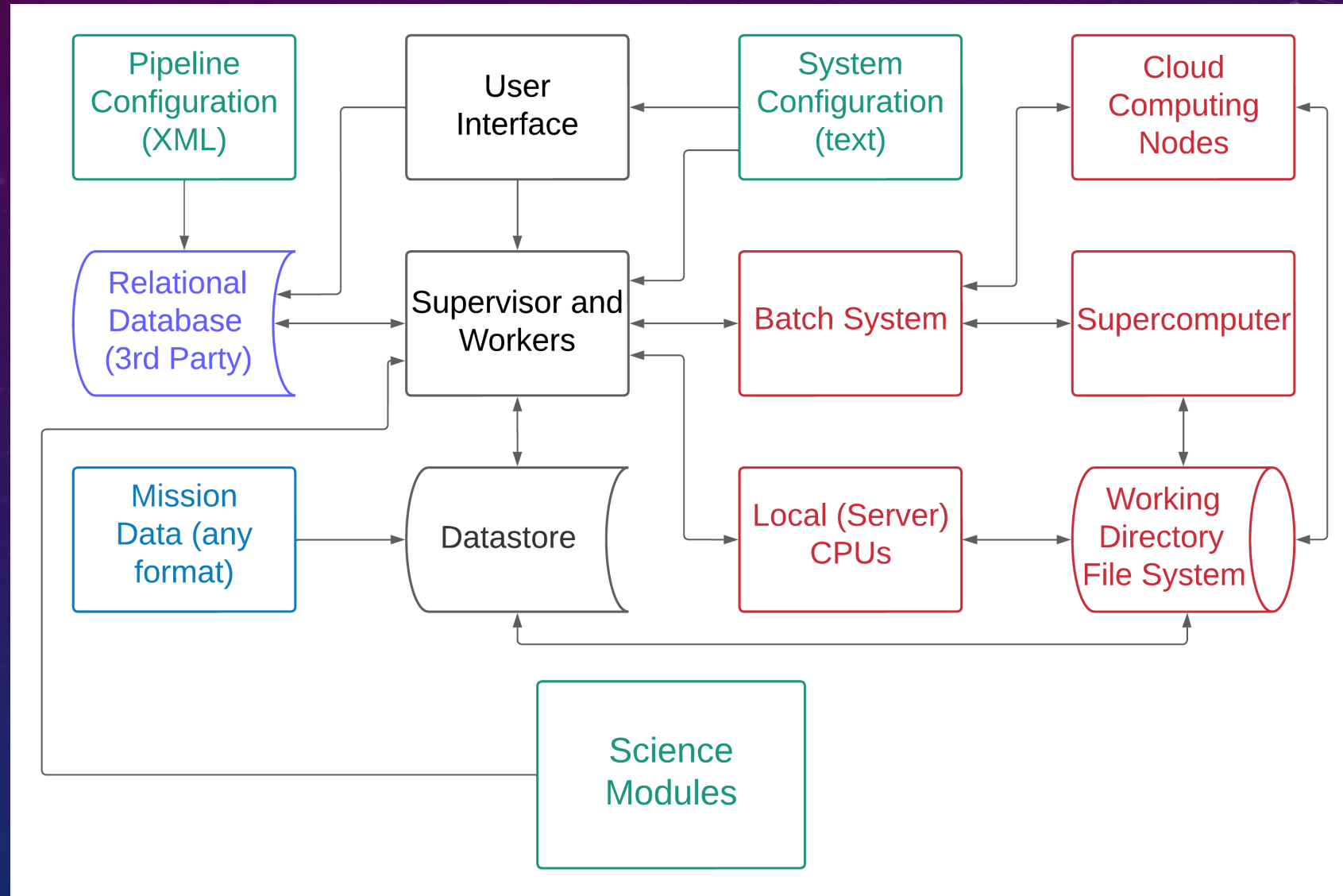
Heritage: Used for science processing of the Kepler mission, where scientists got on with discovering **2879 confirmed exoplanets and 2528 exoplanet candidates!**



Used for science processing of the TESS mission, multi-TB/month data rate and discovery of **172 confirmed exoplanets and 3125 exoplanet candidates!**

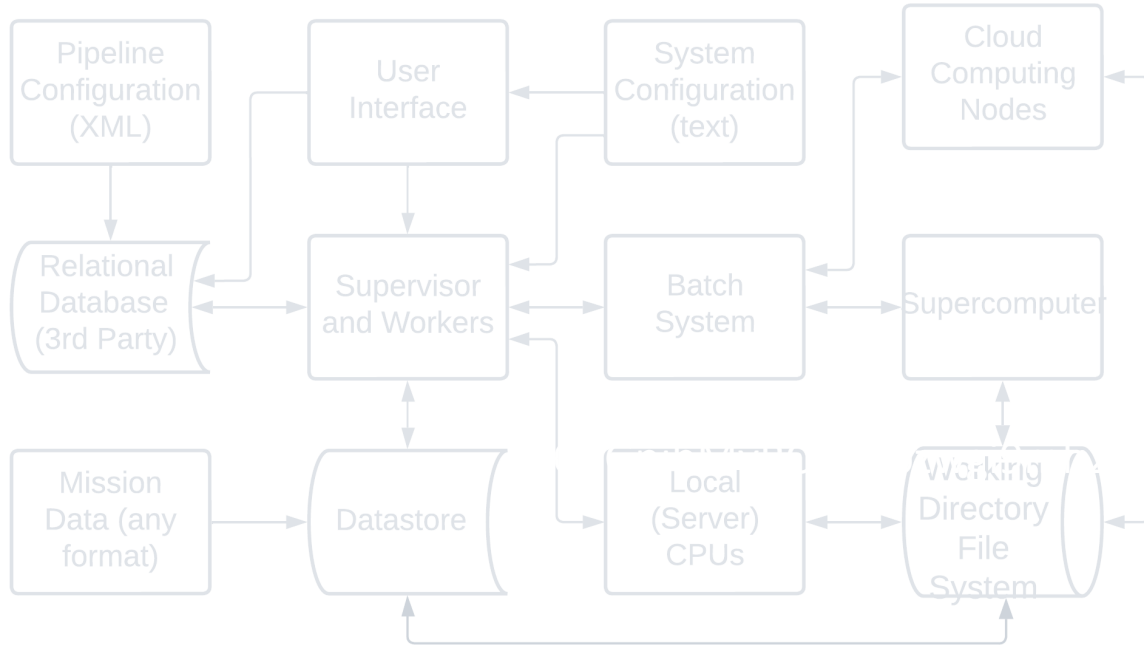


Architecture, or: Under the Hood





Science Modules



Science
Modules

Subject Matter Experts supply the algorithms

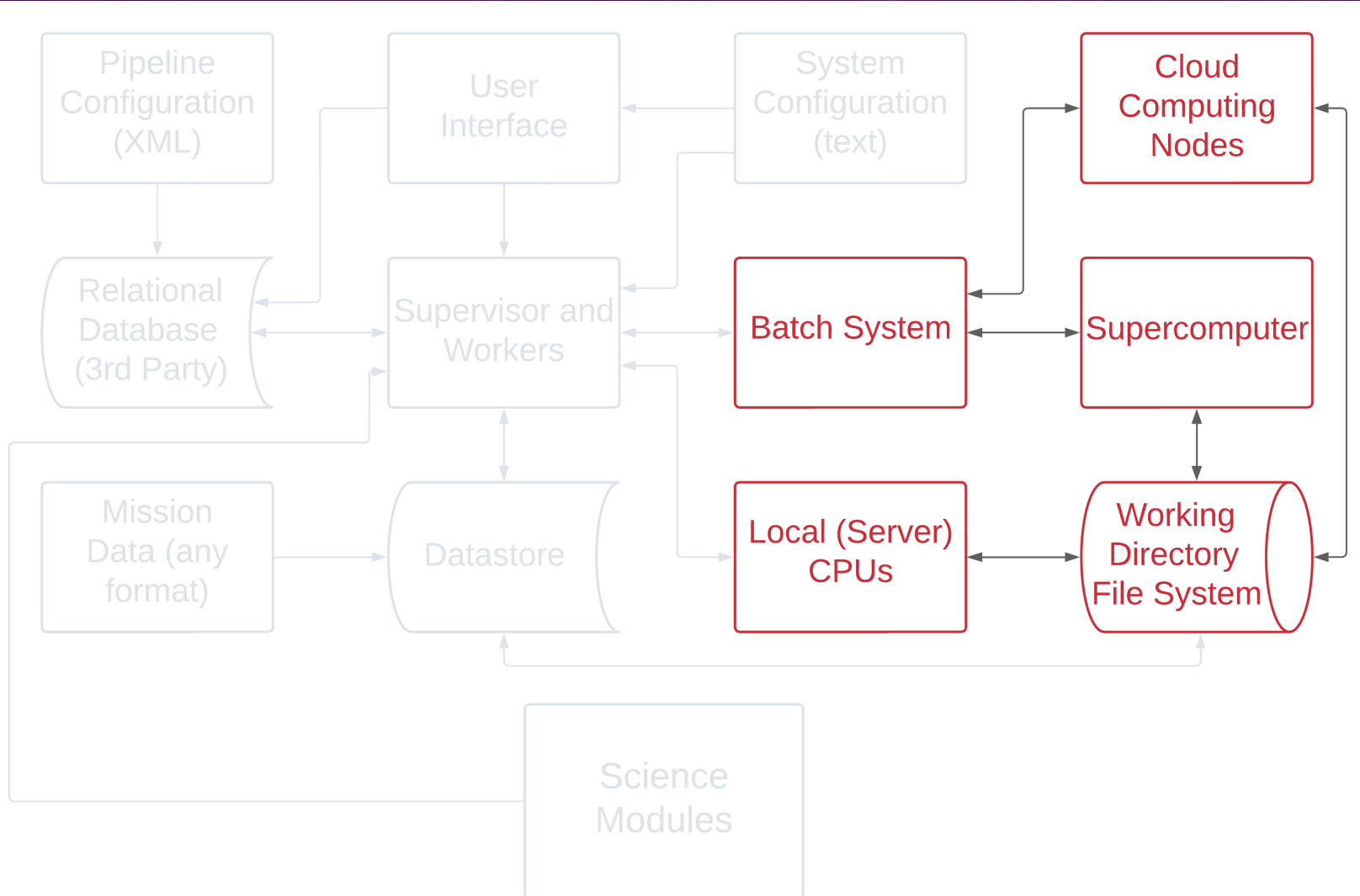
“As-is” – no design rules that must be followed

Polyglot – “mix and match” any languages

So far we’ve used MATLAB, C++, Python, Shell Script – it all just works, no fuss no muss



Execution Options



Select execution in server or in HPC / Cloud on a per-algorithm basis

Current support for NAS (PBS Batch)

Coming soon: support for AWS / Slurm batch



File

Help

NASA

powered by

ziggy

Pi

W

Pr

A

Parameter Library

Pipelines

Instances

Status

Data Receipt

Event Definitions

Module Library

Pipeline instances

Clear filters

Show instances

Started within the last

10

 days

1d

7d

14d

30d

90d

Limit states

☐ Initialized

☐ Errors (running)

☐ Processing

☐ Errors (stalled)

☐ Completed

Instance name

Search

ID	Pipeline name	Date	State	P-time
1	sample: data-recei...	2023-11-06 0...	COMPLETED	00:02:08
2	sample:	2023-11-06 0...	COMPLETED	00:01:57
3	sample: data-recei...	2023-11-06 0...	ERRORS_STALLED	00:20:11

Pipeline tasks

Module	Submitted	Processing	Completed	Failed	Subtasks
data-receipt	0	0	2	0	0 / 0 / 0
permuter	0	0	6	0	24 / 24 / 0
flip	0	0	0	6	8 / 0 / 0
	0	0	8	6	32 / 24 / 0

ID	Module	UOW	State	Worker	P-time	P-state
12	data-receipt	sample-1	COMPLETED	localhost:1	00:00:00	C (0 / 0 / 0)
13	data-receipt	sample-2	COMPLETED	localhost:2	00:00:00	C (0 / 0 / 0)
14	permuter	1	COMPLETED	localhost:2	00:00:35	C (4 / 4 / 0)
15	permuter	2	COMPLETED	localhost:1	00:00:36	C (4 / 4 / 0)
16	permuter	3	COMPLETED	localhost:2	00:00:35	C (4 / 4 / 0)
17	permuter	4	COMPLETED	localhost:1	00:00:35	C (4 / 4 / 0)
18	permuter	5	COMPLETED	localhost:2	00:00:35	C (4 / 4 / 0)
19	permuter	6	COMPLETED	localhost:1	00:00:35	C (4 / 4 / 0)
20	flip	1	ERROR	localhost:1	00:00:03	Ae (4 / 0 / 0)
21	flip	2	ERROR	localhost:2	00:00:03	Ae (4 / 0 / 0)
22	flip	3	ERROR	-	-	I (0 / 0 / 0)
23	flip	4	ERROR	-	-	I (0 / 0 / 0)
24	flip	5	ERROR	-	-	I (0 / 0 / 0)
25	flip	6	ERROR	-	-	I (0 / 0 / 0)



Ziggy is Lightweight

Develop here ...



Dog not included.

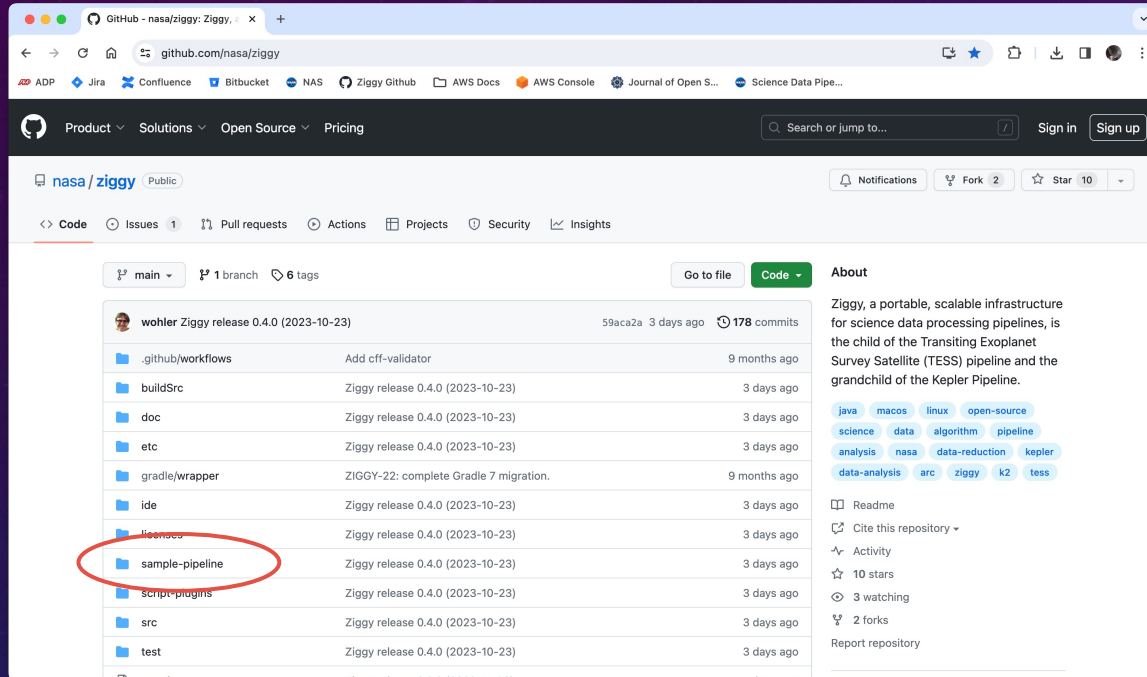
... deploy here!





Ziggy is Open Source!

github.com/nasa/ziggy



Super-basic sample pipeline
that shows you how to do it all

Tons of totally-readable
documentation

