Tibanna - CloudData @ Hyperspeed

read data from HMS or AWS cloud

move data through processing steps
+ monitor / report on status

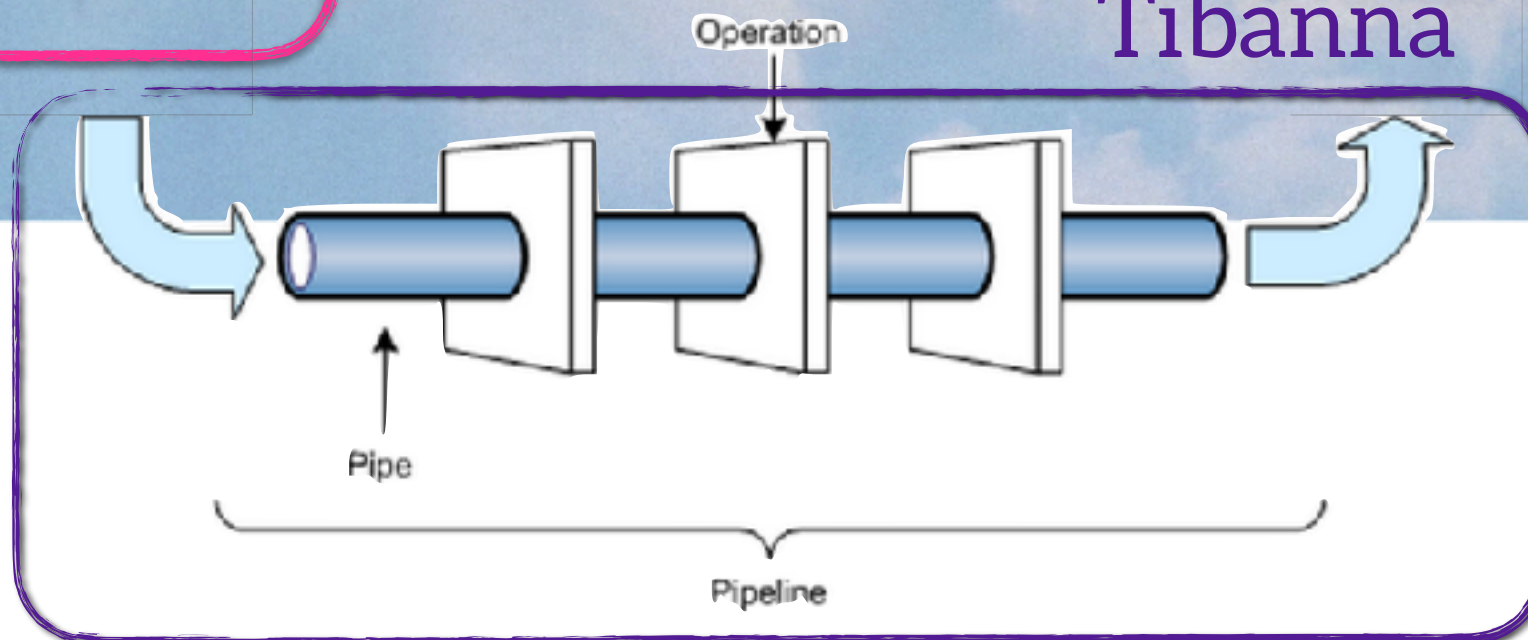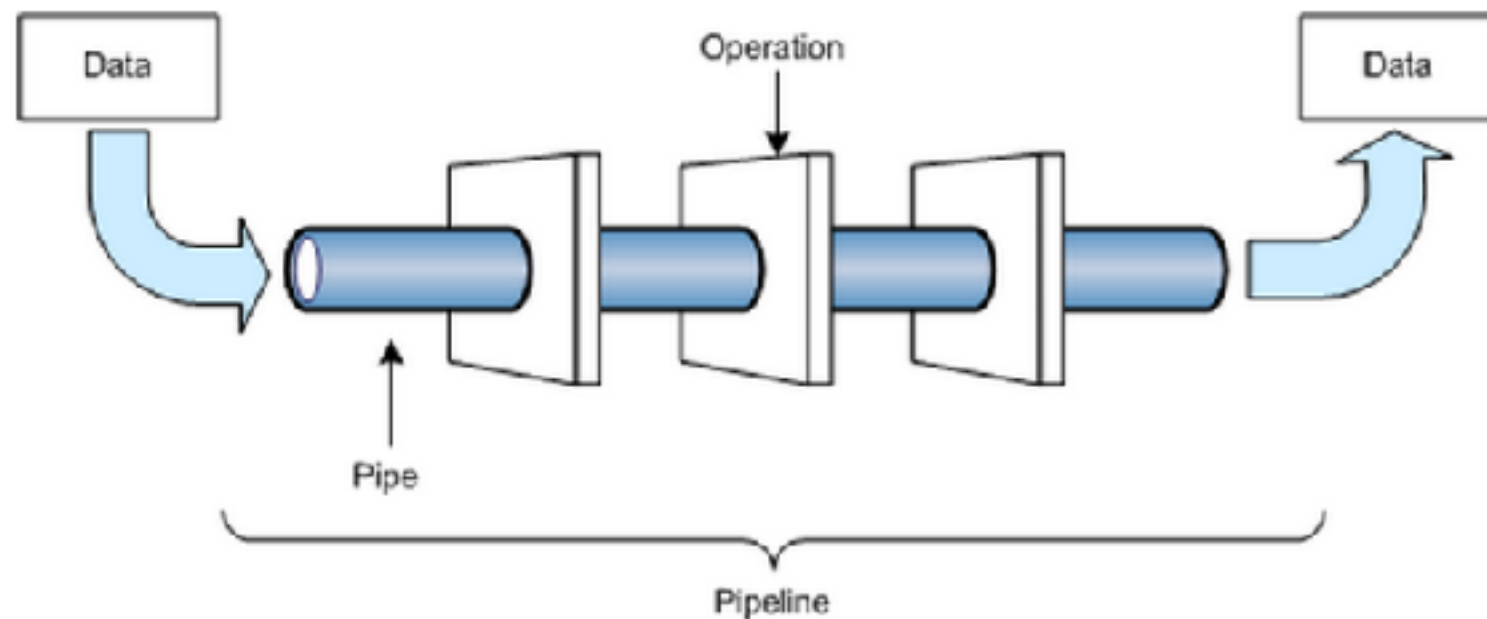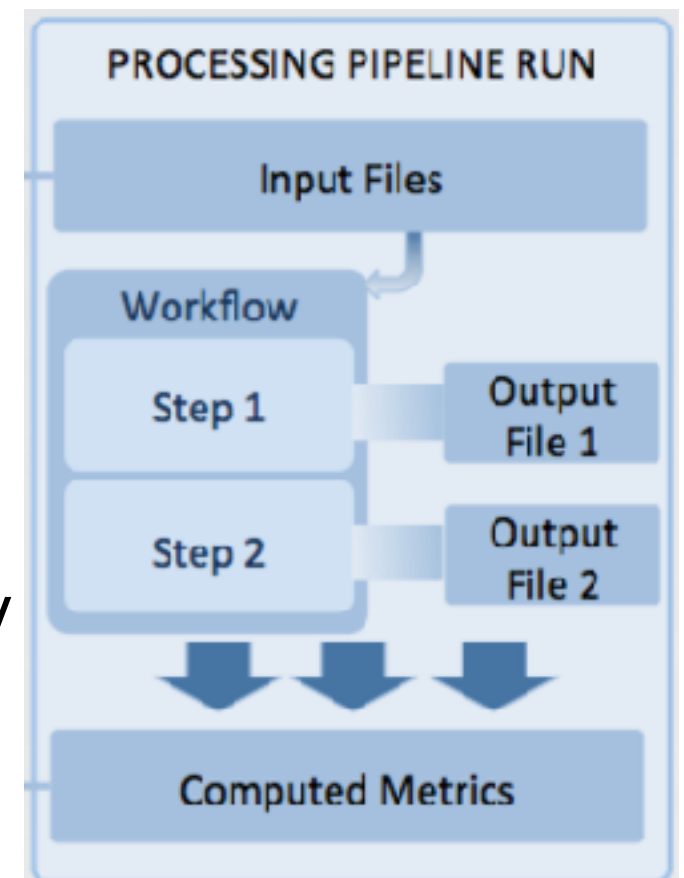- Raw data in from anywhere
- Each Step runs in separate env / container
- Store output files
- Run arbitrary reporting steps
- Each step added to shared reusable library

Reporting / Tracking

Error Reporting

700e60f6-0407-4f6f-322f-4b85522703fa ⓧ

Graph | Code

■ Success ■ Failed ▢ Cancelled ■ In progress

```
          Start
            │
            ▼
       StartRunSbg
            │
            ▼
      CheckImportSbg
            │
            ▼
        RunTaskSbg
            │
            ▼
       CheckTaskSbg
            │
            ▼
       ExportFilesSbg
            │
            ▼
       CheckExportSbg
            │
            ▼
           End
```

▼ Execution Details

Info | Input | Output

error: UnhandledError
cause: {"stackTrace": [["/var/task/service.py", 15,
"handler", "sbg =
utils.create_sbg_workflow(**event.get('workflow'))"]
, ["/var/task/core/utils.py", 113,
"create_sbg_workflow", 'token = get_sbg_keys()"],
["/var/task/core/utils.py", 45, "get_sbg_keys",
"return get_key('sbgkey')"],
["/var/task/core/utils.py", 57, "get_key",
"SSECustomerAlgorithm='AES256')"],
["/var/task/botocore/client.py", 159, "_api_call",
"return self._make_api_call(operation_name,
kwargs)'], ["/var/task/botocore/client.py", 470,
"_make_api_call", "api_params, operation_model,
context=request_context)"],
["/var/task/botocore/client.py", 523,
"_convert_to_request_dict", 'api_params,
operation_model)"],
["/var/task/botocore/validate.py", 270,
"serialize_to_request", "raise
ParamValidationError(report=report.generate_repor
t())"]], "errorType": "ParamValidationError",
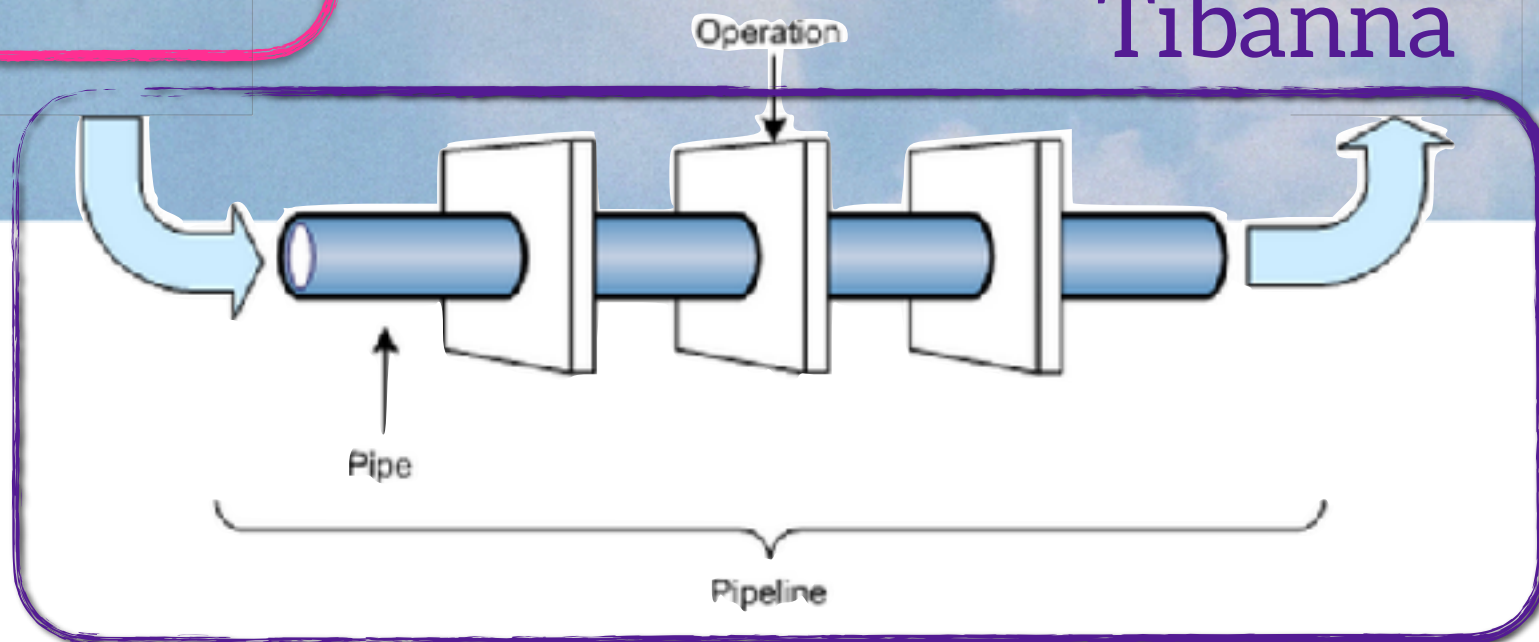
▲ Step Details

Current

S3 Output

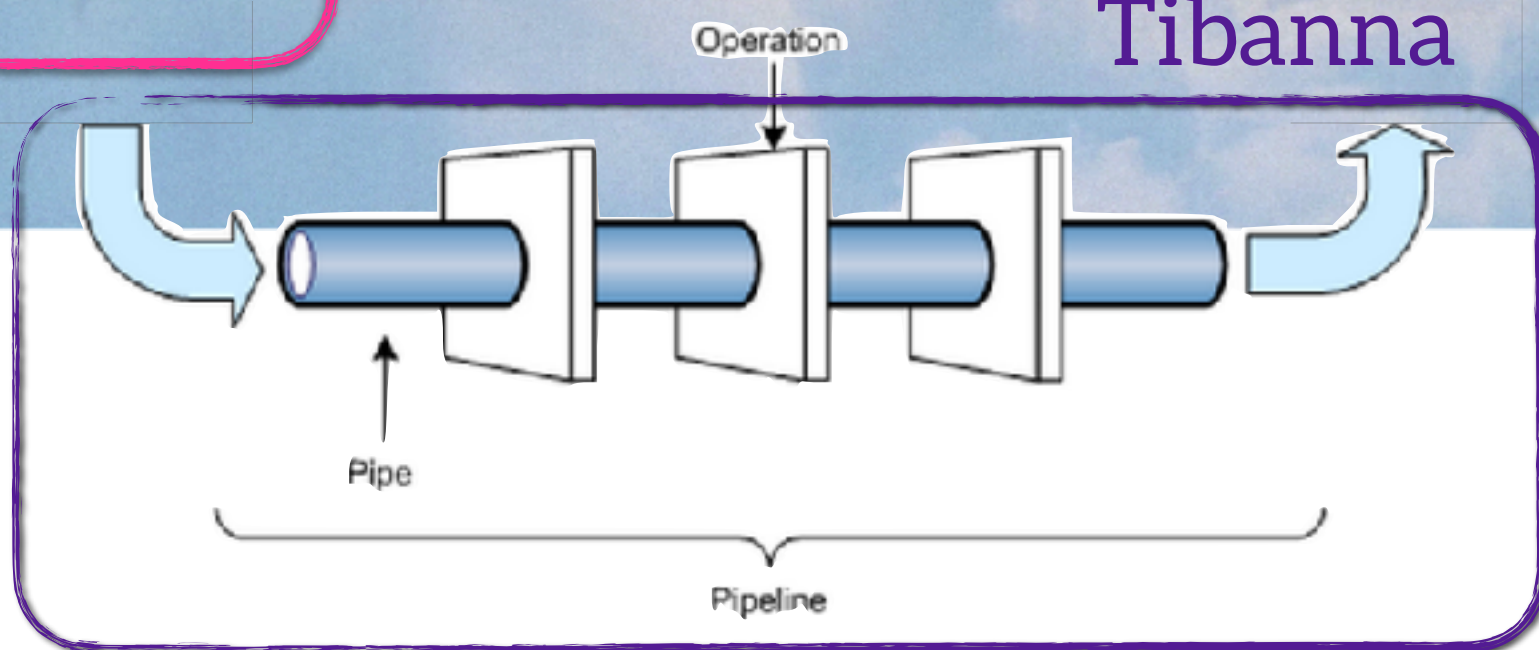S3 Input only

runs on proprietary backed call SBG

Tibanna

Operation

Pipe

Pipeline

Desired

Output Anywhere

Input from Anywhere

Run on EC2 + run on orchestra

Tibanna

Operation

Pipe

Pipeline

https://github.com/hms-dbmi/tibanna

We need you

- Convert bash scripts to python running on lambda
- Get Data plus pipelines to run
- Convert pipelines / software installs to docker / CWL
- Test running stuff on pipelines
- Reporting Integrations? (Slack / phone notifications /etc..)
- Anything else to make it more awesome