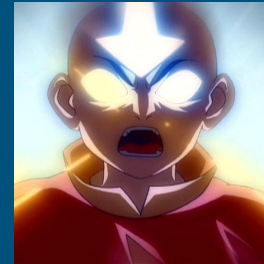


Weekly Deliverables: 4 December 2017



AVATR



AVATR Sprint 2 Progress

0% Complete

>0%, <100% Complete

100% Complete

- Last Week:
 - Finalize Sprint 2 Deliverables
 - Work on database schema
- This week:
 - Tech evaluations on pipeline logging
 - Finalized database schema

Sprint 2	12/20	<ul style="list-style-type: none">● Annotation Pipeline<ul style="list-style-type: none">■ DoD: Demo with ≤ 2 terminal commands● Develop MVP for LIMS with the following features:<ul style="list-style-type: none">○ Dataset registration○ Dataset queries○ Documentation<ul style="list-style-type: none">■ DoD: Demo demonstrating features● Tech Evaluations on Pipeline Logging to see feasibility<ul style="list-style-type: none">■ DoD: Two Tech Evaluations on Logging Tools
----------	-------	---

Our finalized database schema for m2g.io data is organized by subject and data modality.

- Thanks to discussions with Eric
- Choose by subject and modality
 - Optional session filter
- We build off BIDS but aren't constrained
 - Will have assortment of ingest scripts

A	B	C	D	E	F	G	H	I	J	K
SUBID	SESSION	AGE_AT_SCAN	SEX	DSM_IV_TR	FIQ	VIQ	PIQ	BMI	RESTING_STATE	VISUAL_STIM
25864	Baseline	22	#	2	#	#	#	#	Lie Still and F	4
25864	Retest_1	#	#	#	#	#	#	#	#	#
25865	Baseline	24	#	2	#	#	#	#	Lie Still and F	4
25865	Retest_1	#	#	#	#	#	#	#	#	#
25866	Baseline	27	#	2	#	#	#	#	Lie Still and F	4
25866	Retest_1	#	#	#	#	#	#	#	#	#
25867	Baseline	22	#	2	#	#	#	#	Lie Still and F	4
25867	Retest_1	#	#	#	#	#	#	#	#	#
25868	Baseline	27	#	2	#	#	#	#	Lie Still and F	4
25868	Retest_1	#	#	#	#	#	#	#	#	#
25869	Baseline	22	#	1	#	#	#	#	Lie Still and F	4

```
1 pipeline/
2   subject-level/
3     scan-level/
4       anat/
5         preproc/
6           # preprocessing derivatives
7         registered/
8           # registration derivatives
9       qa/
10        # anatomical qa figures, for each of the above
11      preproc/
12        # preprocessing qa figures
13      registered/
14        # registration qa figures
15    func/
16      preproc/
17        # fmri preprocessing derivatives
18      registered/
19        # fmri registration derivatives
20      cleaned/
21        # fmri nuisance-corrected derivatives
22    timeseries/
23      parcellation/
24        # fmri timeseries
25    connectomes/
26      parcellation/
27        # fmri connectomes
28    qa/
29      # fmri qa, for each from above
30    ...
31  dwi/
32    preproc/
33      # dwi preprocessing derivatives
34    registered/
35      # dwi registration derivatives
36    tensor/
37      # dwi tensor model
38    fibers/
39    connectomes/
40      parcellation/
41        # dwi connectomes
42    qa/
43      # dwi qa, for each from above
44    ...
```

We need a better way to parse m2g.io derivative links into our LIMS

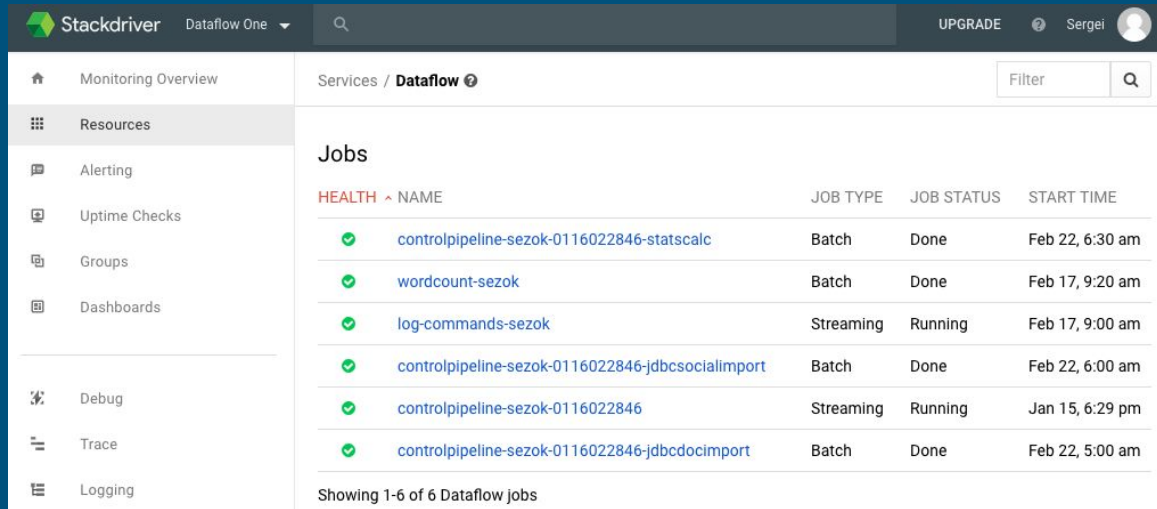
- Acquire links to s3 buckets
 - But they load via JS
- Can do dynamic js scraping (PyQt)
 - Is there another way?

```
BNU1
http://fcon_1000.projects.nitrc.org/indi/CoRR/html/bnu_1.html
[@csv]
http://mrneurodata.s3-website-us-east-1.amazonaws.com/data/BNU1/BNU1.csv
Aligned Images
http://mrneurodata.s3-website-us-east-1.amazonaws.com/BNU1/ndmg_0-0-48/reg_dti/
Tensors
http://mrneurodata.s3-website-us-east-1.amazonaws.com/BNU1/ndmg_0-0-48/tensors/
Fibers
http://mrneurodata.s3-website-us-east-1.amazonaws.com/BNU1/ndmg_0-0-48/fibers/
Graphs
http://mrneurodata.s3-website-us-east-1.amazonaws.com/BNU1/ndmg_0-0-48/graphs/
QA
http://mrneurodata.s3-website-us-east-1.amazonaws.com/BNU1/ndmg_0-0-48/qa/
v0.0.48
https://github.com/neurodata/ndmg/tree/v0.0.48
```

Last Modified	Size	Key
2017-03-07T21:58:23.000Z	163.1 MB	../sub-0025864_ses-1_dwi_aligned.nii.gz
2017-03-07T21:34:18.000Z	164.0 MB	sub-0025864_ses-2_dwi_aligned.nii.gz
2017-03-07T21:46:05.000Z	161.1 MB	sub-0025865_ses-1_dwi_aligned.nii.gz
2017-03-07T22:09:07.000Z	161.2 MB	sub-0025865_ses-2_dwi_aligned.nii.gz
2017-03-07T21:17:25.000Z	164.2 MB	sub-0025866_ses-1_dwi_aligned.nii.gz
2017-03-07T22:11:03.000Z	164.2 MB	sub-0025866_ses-2_dwi_aligned.nii.gz
2017-03-07T21:59:35.000Z	163.4 MB	sub-0025867_ses-1_dwi_aligned.nii.gz
2017-03-07T21:37:41.000Z	160.6 MB	sub-0025867_ses-2_dwi_aligned.nii.gz
2017-03-07T22:05:50.000Z	164.1 MB	sub-0025868_ses-1_dwi_aligned.nii.gz
2017-03-07T21:57:00.000Z	159.4 MB	sub-0025868_ses-2_dwi_aligned.nii.gz
2017-03-07T22:02:57.000Z	159.7 MB	sub-0025869_ses-1_dwi_aligned.nii.gz

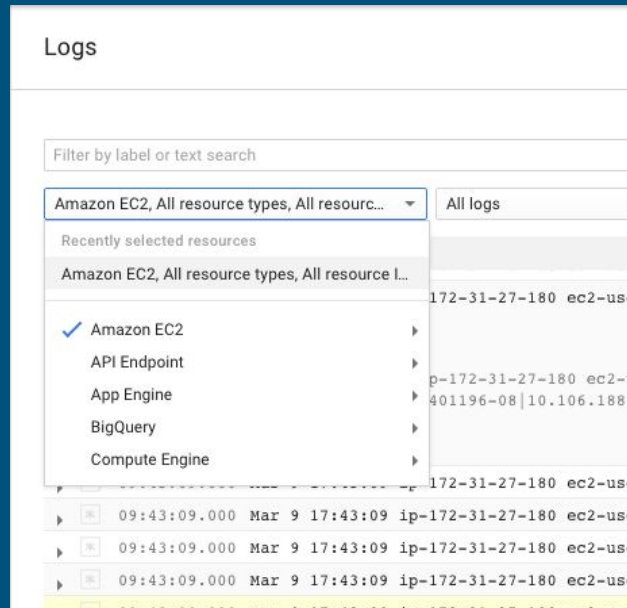
```
<!DOCTYPE html>
<html>
<head>
<title>MR-NeuroData</title>
</head>
<body>
<!-- <div id="navigation"></div> -->
<div id="listing"></div>
<script src="https://ajax.googleapis.com/ajax/libs/jquery/3.1.1/jquery.min.js" type="text/javascript"></script>
<script type="text/javascript">
  var S3BL_IGNORE_PATH = false;
  var BUCKET_URL = 'https://mrneurodata.s3.amazonaws.com';
  var S3B_ROOT_DIR = 'data/';
  var S3B_SORT = 'Z2A';
</script>
<script src="https://cdn.rawgit.com/gkjar/a0e79ba5362528739dad5d92bde86d9/raw/289427e3613048a418b5ac57276e060ff3a02c7e/list.js" type="text/javascript"></script>
</body>
</html>
```

Tech Eval: StackDriver provides custom metrics you can log for your cloud applications.



The screenshot shows the Stackdriver Dataflow console. The left sidebar contains navigation links: Monitoring Overview, Resources, Alerting, Uptime Checks, Groups, Dashboards, Debug, Trace, and Logging. The main content area is titled 'Services / Dataflow' and displays a table of jobs. The table has columns for HEALTH, NAME, JOB TYPE, JOB STATUS, and START TIME. There are six jobs listed, all with a green checkmark in the HEALTH column. The jobs are: controlpipeline-sezok-0116022846-statscalc (Batch, Done, Feb 22, 6:30 am), wordcount-sezok (Batch, Done, Feb 17, 9:20 am), log-commands-sezok (Streaming, Running, Feb 17, 9:00 am), controlpipeline-sezok-0116022846-jdbcsocialimport (Batch, Done, Feb 22, 6:00 am), controlpipeline-sezok-0116022846 (Streaming, Running, Jan 15, 6:29 pm), and controlpipeline-sezok-0116022846-jdbcdocimport (Batch, Done, Feb 22, 5:00 am). Below the table, it says 'Showing 1-6 of 6 Dataflow jobs'.

HEALTH	NAME	JOB TYPE	JOB STATUS	START TIME
✓	controlpipeline-sezok-0116022846-statscalc	Batch	Done	Feb 22, 6:30 am
✓	wordcount-sezok	Batch	Done	Feb 17, 9:20 am
✓	log-commands-sezok	Streaming	Running	Feb 17, 9:00 am
✓	controlpipeline-sezok-0116022846-jdbcsocialimport	Batch	Done	Feb 22, 6:00 am
✓	controlpipeline-sezok-0116022846	Streaming	Running	Jan 15, 6:29 pm
✓	controlpipeline-sezok-0116022846-jdbcdocimport	Batch	Done	Feb 22, 5:00 am



The screenshot shows the Stackdriver Logs console. The top bar has a search filter and a user profile. The main content area is titled 'Logs' and has a search bar. Below the search bar, there is a dropdown menu for resource selection. The dropdown menu is open, showing a list of resources. The selected resource is 'Amazon EC2, All resource types, All resource I...'. Other resources listed include 'API Endpoint', 'App Engine', 'BigQuery', and 'Compute Engine'. The right side of the console shows a list of log entries with timestamps and IP addresses.

Logs

Filter by label or text search

Amazon EC2, All resource types, All resource I... All logs

Recently selected resources

Amazon EC2, All resource types, All resource I...

✓ Amazon EC2

API Endpoint

App Engine

BigQuery

Compute Engine

09:43:09.000 Mar 9 17:43:09 ip-172-31-27-180 ec2-use

09:43:09.000 Mar 9 17:43:09 ip-172-31-27-180 ec2-use

09:43:09.000 Mar 9 17:43:09 ip-172-31-27-180 ec2-use

StackDriver isn't very LIMS friendly or useful for data pipelines.

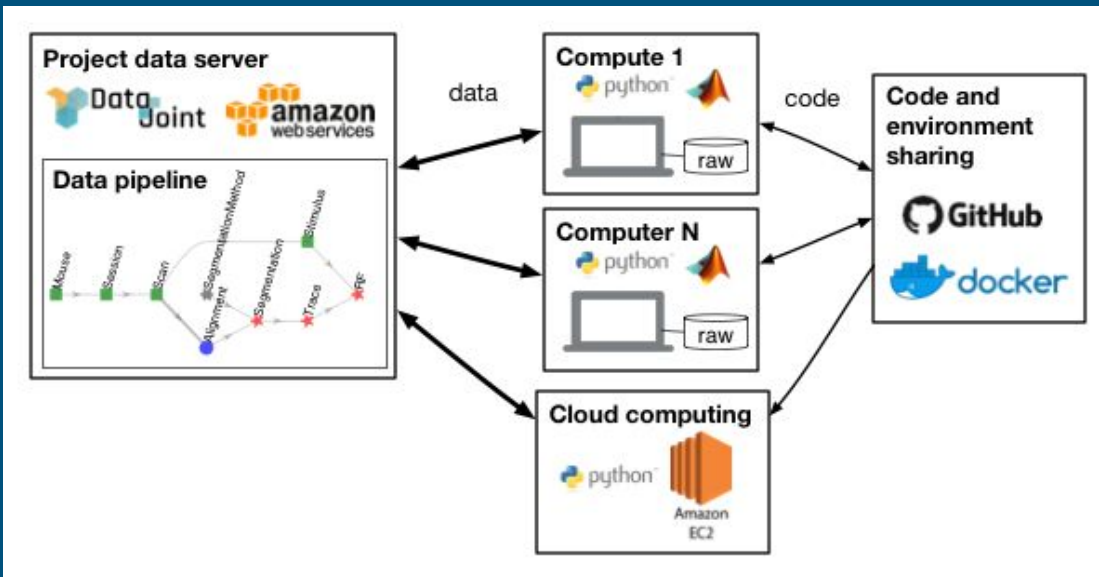
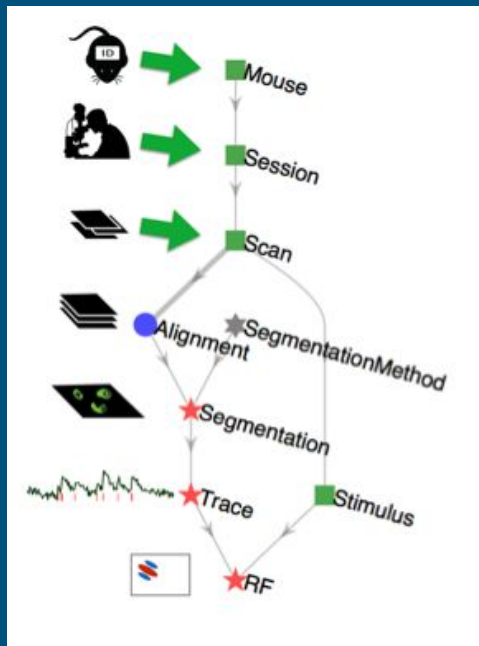
Pros:

- Works well on EC2 Instances (Batch)
- Can specify which metrics of your application you want to track

Cons:

- Need to learn and pay for Google Cloud Platform
- A lot of the functionality is meant for Google Cloud tools
- Will need to write an additional parser to scrape Google Logs to enter into LIMS
 - Might as well use CloudWatch at this point instead

Tech Eval: DataJoint provides a framework to automatically add newly computed data into a database.



For next week...

- Finish scraping
- Finish data registration
 - Using our current schema, organizing data in a mongoDB
- Talk with annotators about revising pipeline
 - Locally hosted, click buttons