splunk> .conf2017

Splunking the 2016 Presidential Election

Corey Marshall | Splunk4Good Director

Satoshi Kawasaki | Splunk4Good Ninja

September 27th, 2017 | Washington, DC

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Bio: Corey Marshall

Splunk4Good Director

Splunk4Good

BA in Political Science from Lewis & Clark College

Master's in Public Policy from the University of Chicago

- Advising government and non-profits on open data for more than 15 years, including working with
  - City and County of San Francisco
  - Accenture
  - Office of Chicago Mayor Richard M. Daley
- Joined Splunk in 2013
- Lead company’s efforts in
  - employee service and engagement
  - community giving
  - social impact initiatives

Bio: Satoshi Kawasaki

Splunk4Good Ninja

BS in Aerospace Engineering from Georgia Tech

hobbes3

- Also joined Splunk in 2013
  - 3 years of Professional Services (PS)
  - 1+ year of Splunk4Good
- Unofficially became a dashboard/visualization specialist in PS
  - .conf 2014: *I Want that Cool Viz in Splunk!*
  - .conf 2015: *Enhancing Dashboards with JavaScript!*
- Doing 3 talks this year
  - .conf 2017: *Speed up your searches!*
  - .conf 2017: *Splunking to fight human trafficking*
  - .conf 2017: *Splunking the 2016 presidential election*

You are here.

About Splunk4Good

Big data can make a big difference

WOUNDED WARRIOR PROJECT®

NETHOPE

TEAM RUBICON

npower

- $100 million Splunk Pledge has issued licenses and training worth over $6 million
- Provide workforce training to veterans and opportunity youth to train the workforce of tomorrow
- Engaging our partners in initiatives to promote STEM and develop shared solutions for humanitarian response and human trafficking
- Supporting life-changing research at top universities
- More than 70,000 hours of paid volunteer time

2016 Presidential Election

Spoiler Alert!

elections.splunk4good.com

splunk> .conf2017

Our goals and requirements

Goals

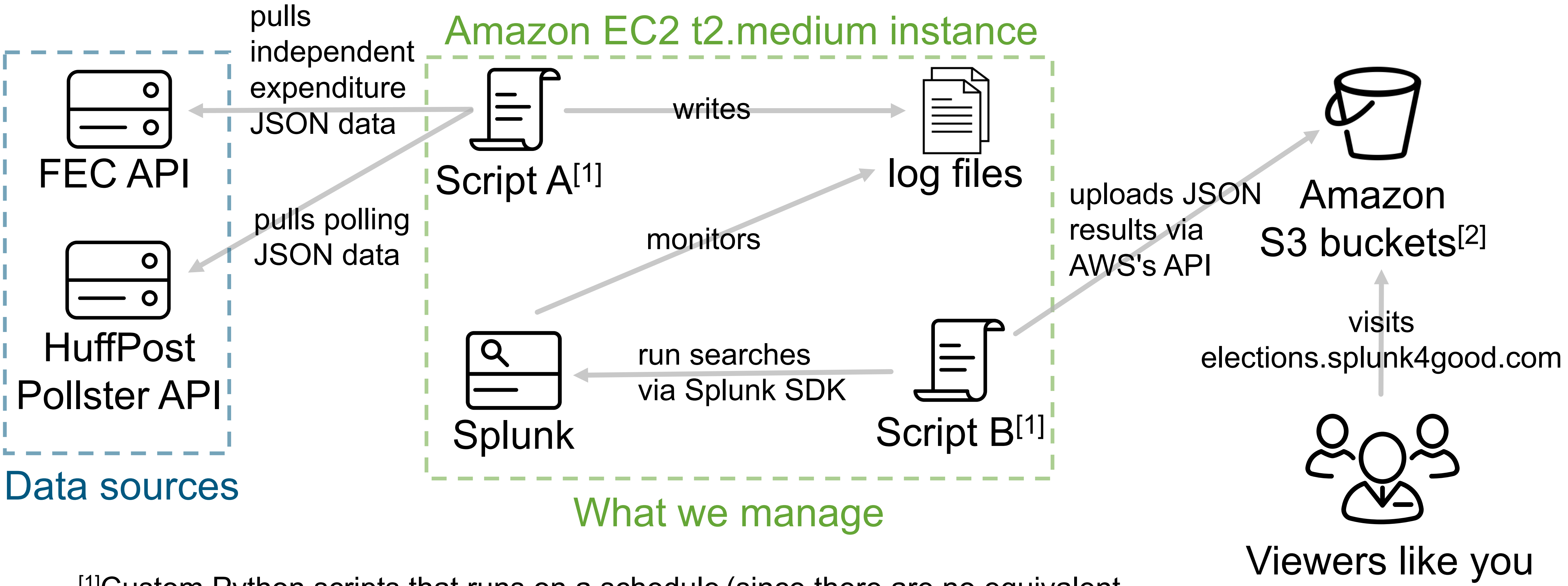
- Publically showcase Splunk's ability to ingest and analyze non-traditional[1] and open data
- Show how Splunk can correlate data from different sources
- Provide a meaningful story or discovery

Requirements

- Create a *public-facing* interface or website
- Scale to handle public traffic
- Try to be unbiased and neutral
- Show off some custom, kick-ass visualizations

[1]Not security or IT data

The "easy" architecture



Amazon EC2 t2.medium instance

pulls independent expenditure JSON data

FEC API

pulls polling JSON data

HuffPost Pollster API

Data sources

Script A[1]

writes

log files

monitors

Splunk

run searches via Splunk SDK

Script B[1]

What we manage

uploads JSON results via AWS's API

Amazon S3 buckets[2]

visits elections.splunk4good.com

Viewers like you

[1]Custom Python scripts that runs on a schedule (since there are no equivalent functionality from Splunkbase apps)

[2]Hosting html, css, and javascript as a static website (Amazon managed service)

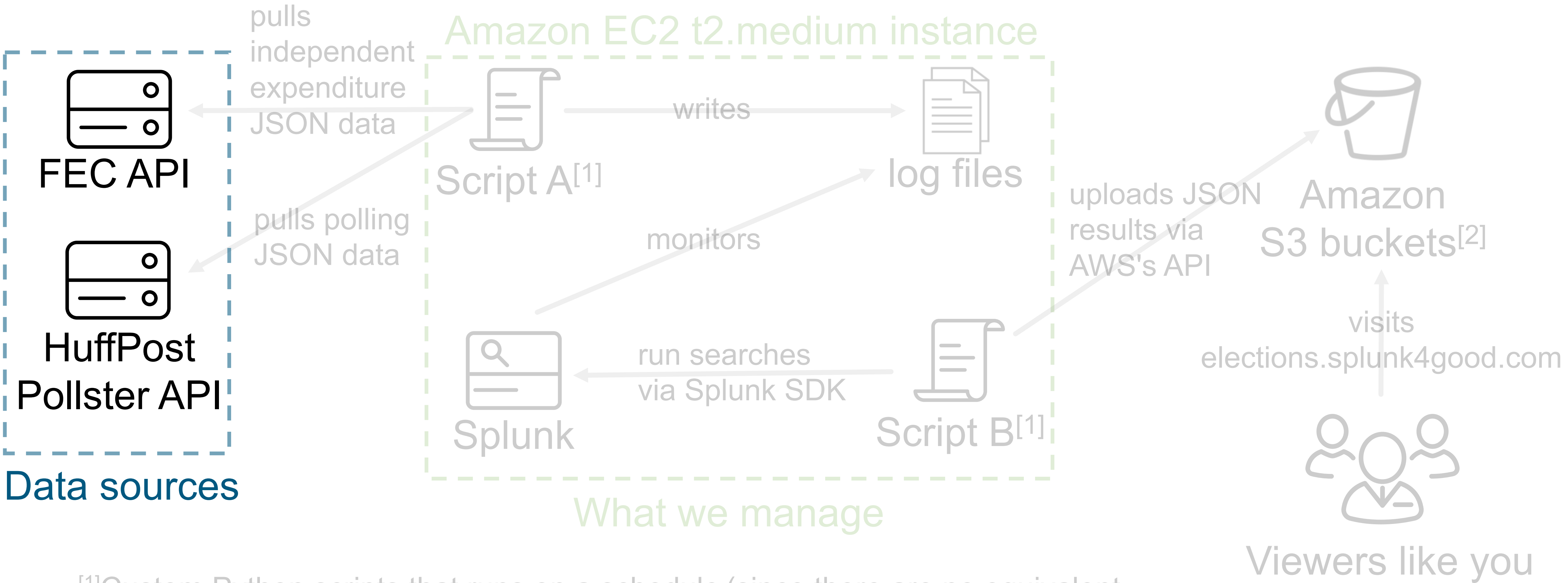
The easy[1] steps

How to go from a private Splunk instance to a public website

1. Preview the data
2. Record the data
3. Index the data
4. Upload the data
5. Serve the data

[1] It's actually not that easy

Step 1: Preview the data



Data sources

FEC API

HuffPost Pollster API

pulls independent expenditure JSON data

pulls polling JSON data

Amazon EC2 t2.medium instance

Script A[1]

writes

log files

monitors

Splunk

run searches via Splunk SDK

Script B[1]

What we manage

uploads JSON results via AWS's API

Amazon S3 buckets[2]

visits elections.splunk4good.com

Viewers like you

[1]Custom Python scripts that runs on a schedule (since there are no equivalent functionality from Splunkbase apps)

[2]Hosting html, css, and javascript as a static website (Amazon managed service)

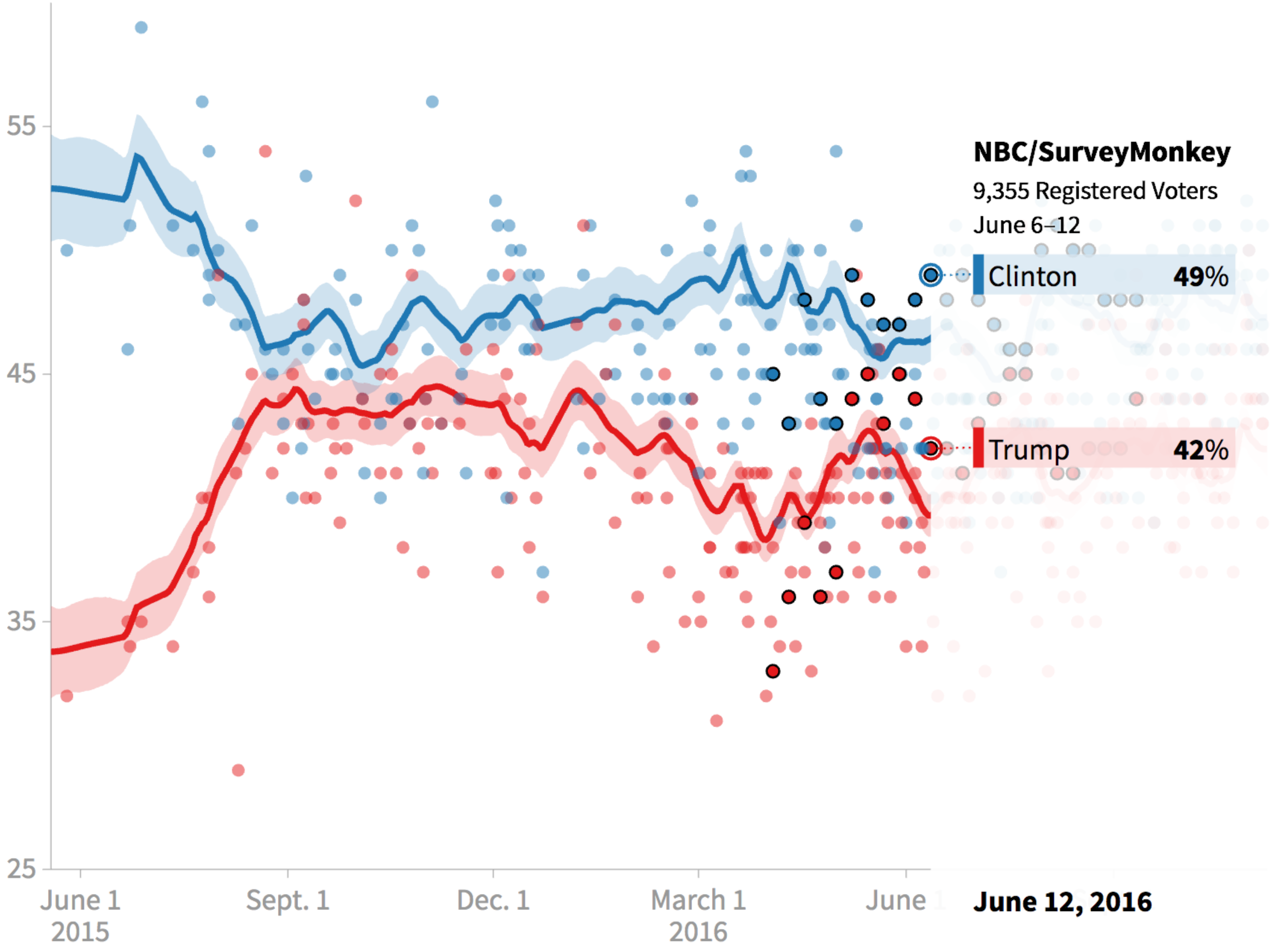
Data source #1: Federal Election Commission (FEC)

- FEC is an independent regulatory agency whose purpose is to enforce campaign finance law in federal elections
- We decided to mostly focus on independent expenditures (aka schedule e) of the "Super PACs"[1]
- Provides campaign finance data via *https://www.fec.gov/data/*
- Also provides a documented REST API on the same dataset: *https://api.open.fec.gov/developers/*

[1]The creation of the Super PACs came from the landmark ruling of *Citizens United v. FEC (2010)*

Data source #2: HuffPost Pollster

- HuffPost is a politically liberal American news and opinion website and blog
- HuffPost Pollster tracks and aggregates thousands of public polls and provides a documented REST API on those dataset: *https://app.swaggerhub.com/apis/huffpostdata/pollster-api/2.0.0*



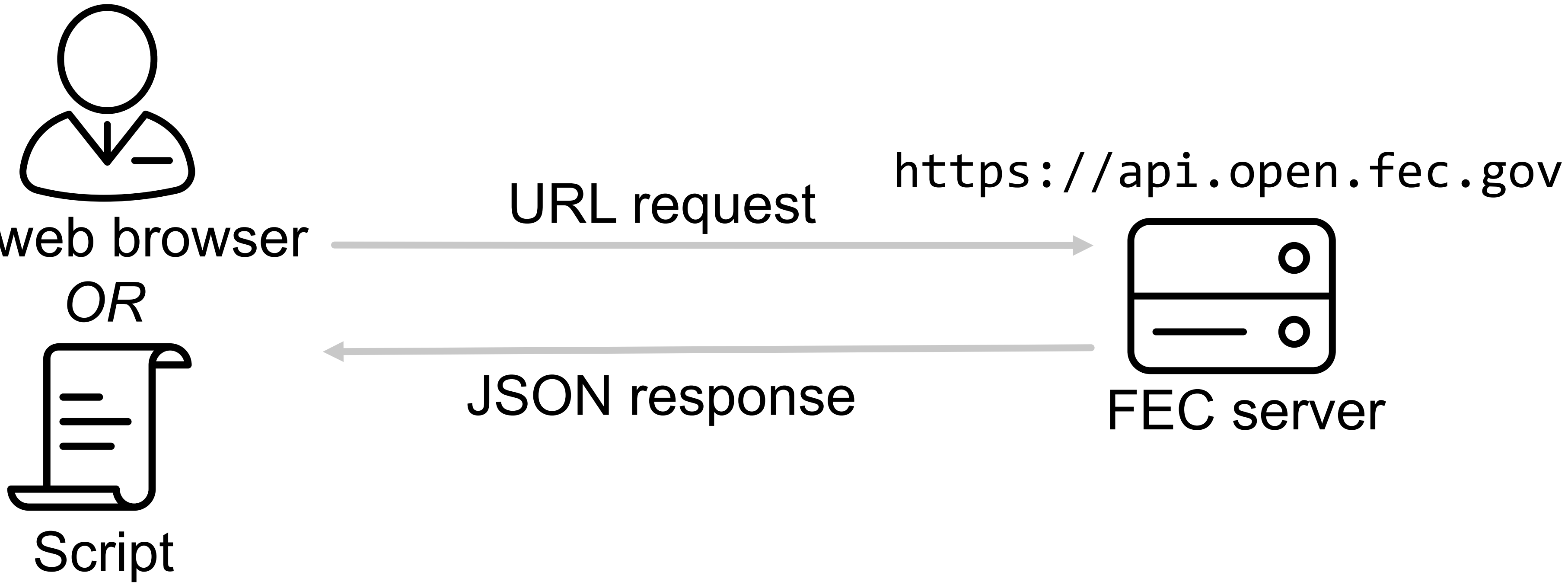
HUFFPOST

splunk> .conf2017

REST API

What is REST API?

A REST API defines a set of functions which developers can perform requests and receive responses via HTTP protocol such as GET and POST.

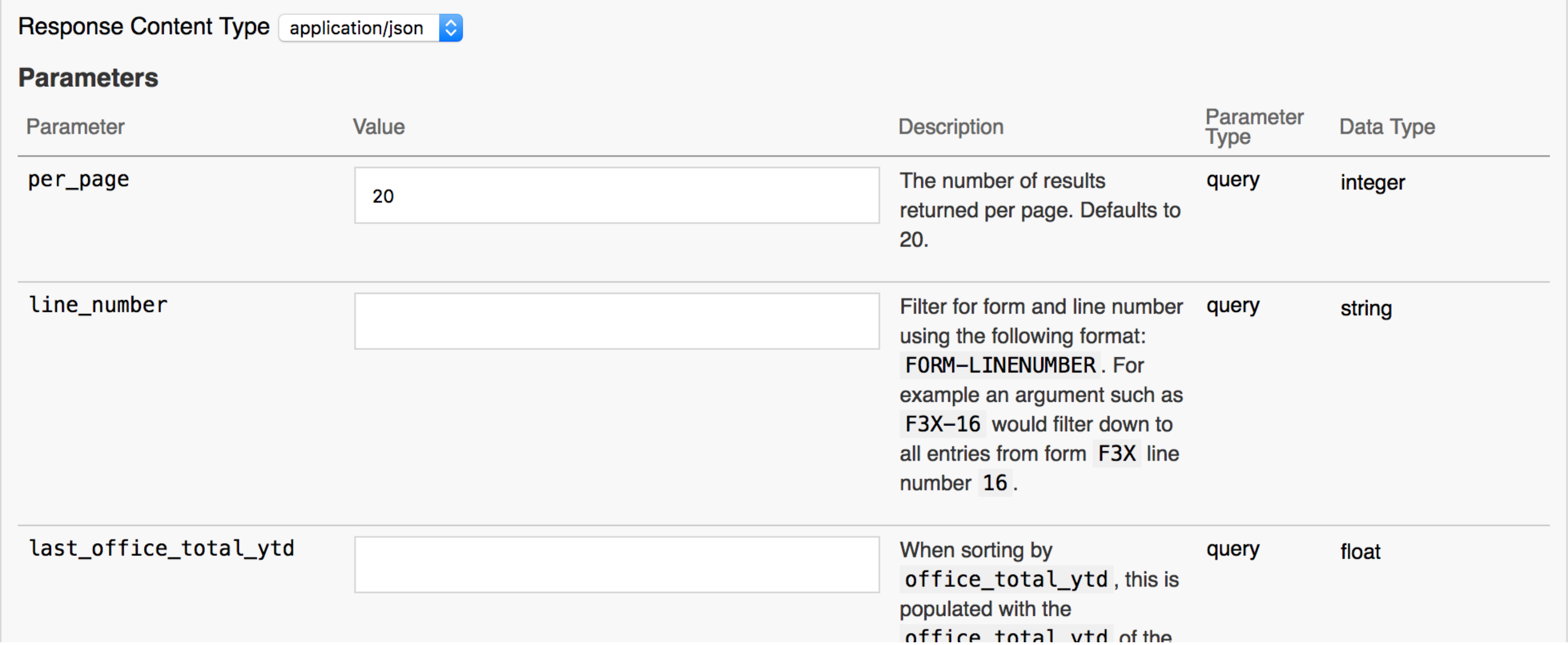


Example URL: *https://api.open.fec.gov/v1/candidate/P80001571/?api_key=DEMO_KEY*

Example: FEC REST API URL

Find the correct URL from the API documentation

```
https://api.open.fec.gov/v1/schedules/schedule_e/?candidate_id=P80001571&
per_page=100&is_notice=false&cycle=2016&api_key=DEMO_KEY
```



Response Content Type: application/json

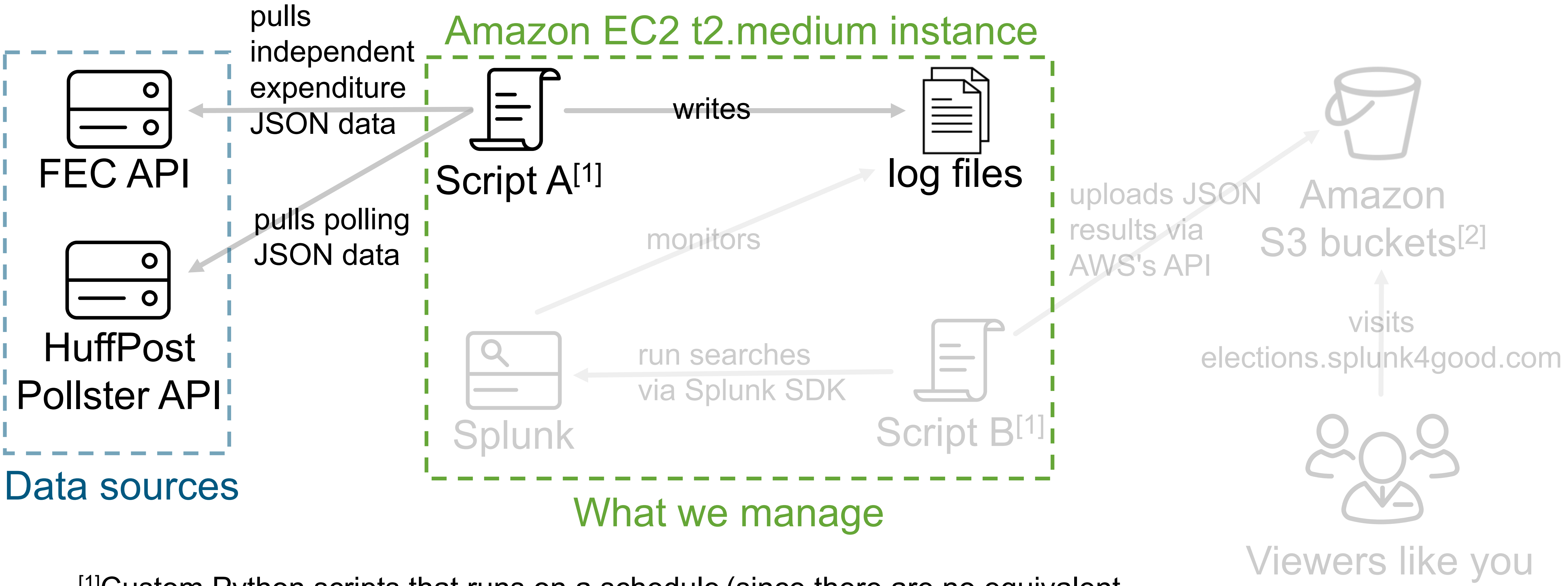
Parameters

Parameter	Value	Description	Parameter Type	Data Type
`per_page`	20	The number of results returned per page. Defaults to 20.	query	integer
`line_number`		Filter for form and line number using the following format: `FORM-LINENUMBER`. For example an argument such as `F3X-16` would filter down to all entries from form `F3X` line number `16`.	query	string
`last_office_total_ytd`		When sorting by `office_total_ytd`, this is populated with the `office_total_ytd` of the	query	float

We use `is_notice=false` to exclude 24- and 48-hour reports, ie we want the completed reports.

splunk> .conf2017

Step 2: Record the data



Data sources

FEC API

HuffPost Pollster API

pulls independent expenditure JSON data

pulls polling JSON data

Amazon EC2 t2.medium instance

Script A[1]

writes

log files

monitors

Splunk

run searches via Splunk SDK

Script B[1]

What we manage

uploads JSON results via AWS's API

Amazon S3 buckets[2]

visits elections.splunk4good.com

Viewers like you

[1]Custom Python scripts that runs on a schedule (since there are no equivalent functionality from Splunkbase apps)

[2]Hosting html, css, and javascript as a static website (Amazon managed service)

FEC JSON response

FEC API is limited up to 100 results per response

{"api_version":"1.0","pagination":{"count":18207,"pages":183,"last_indexes":{"last_index":"4010420171358323494","last_expenditure_date":"2016-11-28T00:00:00"},"per_page":100},"results":[{"payee_name":"ACTBLUE TECHNICAL SERVICES","office_total_ytd":603.07,"conduit_committee_id":"C00626234","payee_street_1":"366 SUMMER STREET","report_type":"YE","expenditure_description":"CREDIT CARD PROCESSING FEES","filer_suffix":null,"original_sub_id":null,"conduit_committee_street1":null,"conduit_committee_name":null,"image_number":"201701319042196565","payee_suffix":null,"conduit_committee_city":null,"conduit_committee_zip":null,"payee_prefix":null,"independent_sign_name":"RANDOLPH, SUSANNAH","expenditure_amount":18.74,"back_reference_transaction_id":null,"file_number":1144979,"payee_middle_name":null,"cand_office_state":null,"expenditure_date":"2016-12-31T00:00:00","memo_code_full":null,"cand_office_district":null,"report_year":2016,"candidate_id":"P80001571","candidate_prefix":null,"notary_sign_name":null,"filer_first_name":"SUSANNAH","filing_form":"F3X","action_code_full":"ADD","category_code":"001","candidate_first_name":"DONALD","filer_last_name":"RANDOLPH","committee_id":"C00626234","candidate_suffix":null,"memoed_subtotal":false,"payee_city":"SOMERVILLE","election_type":"G2020","filer_prefix":null,"candidate_last_name":"TRUMP","payee_zip":"021443132","schedule_type":"SE","conduit_committee_state":null,"payee_state":"MA","conduit_committee_street2":null,"filer_middle_name":null,"candidate":{"two_year_period":2016.0,"idx":88448,"candidate_id":"P80001571"},"payee_first_name":null,"schedule_type_full":"ITEMIZED INDEPENDENT EXPENDITURES","dissemination_date":"2016-12-21T00:00:00","notary_commission_expiration_date":null,"link_id":4013120171369074356,"candidate_middle_name":"J","election_type_full":null,"action_code":"A","is_notice":false,"payee_last_name":null,"support_oppose_indicator":"S","memo_code":null,"pdf_url":"http:\/\/docquery.fec.gov\/cgi-bin\/fecimg\/?201701319042196565","payee_street_2":null,"line_number":"24","committee":{"city":"ORLANDO","party_full":null,"street_1":"701 DELANEY PARK DRIVE","cycles":[2018,2016],"party":null,"candidate_ids":[],"committee_type_full":"Super PAC (Independent Expenditure-Only)","street_2":null,"organization_type":null,"zip":"32806","designation":"U","cycle":2016,"treasurer_name":"SUSANNAH RANDOLPH","designation_full":"Unauthorized","state":"FL","organization_type_full":null,"committee_id":"C00626234","state_full":"Florida","committee_type":"O","name":"HELPING ELECT REFORMERS"},"sub_id":"4021020171370394552","independent_sign_date":"2017-01-31T00:00:00","memo_text":null,"notary_sign_date":null,"back_reference_schedule_name":null,"candidate_office":"P","category_code_full":"Administrative\/Salary\/Overhead Expenses ","candidate_name":"TRUMP, DONALD J"},{"payee_name":"WESTERN TRAILS GUN AND KNIFE SHOWS","office_total_ytd":9315895.8800000008,"conduit_committee_id":"C00580100","payee_street_1":"ATTN: KARL LANGE","report_type":"YE","expenditure_description":"VOID - BOOTH RENTAL - EVENT CANCELLED","filer_suffix":null,"original_sub_id":null,"conduit_committee_street1":null,"conduit_committee_name":null,"image_number":"201705049053505223","payee_suffix":null,"conduit_committee_city":null,"conduit_committee_zip":null,"cand_office_state":null,"independent_sign_name":"ADKINS, MARY ROSE","expenditure_amount":-9.17,"back_reference_transaction_id":null,"file_number":1161245,"payee_middle_name":null,

....

FEC API calls

Script A paginates to get the full results

```
https://api.open.fec.gov/v1/schedules/schedule_e/?candidate_id=P80001571&
per_page=100&is_notice=false&cycle=2016&api_key=DEMO_KEY
```

↓ to fetch the next set of results

```
https://api.open.fec.gov/v1/schedules/schedule_e/?candidate_id=P80001571&
per_page=100&is_notice=false&cycle=2016&api_key=DEMO_KEY&last_index=40104
20171358323494&last_expenditure_date=2016-11-28T00:00:00
```

↓ to fetch the next set of results

```
https://api.open.fec.gov/v1/schedules/schedule_e/?candidate_id=P80001571&
per_page=100&is_notice=false&cycle=2016&api_key=DEMO_KEY&last_index=40210
20171370392792&last_expenditure_date=2016-11-08T00:00:00
```

↓ to fetch the next set of results

Script A repeats until finished (takes about 200 times)

HuffPost Pollster JSON response

No need to paginate here

```
{"id":624,"title":"2016 General Election: Trump vs. Clinton","slug":"2016-general-election-trump-vs-clinton","topic":"2016-president",
"state":"US","short_title":"2016 President: Trump vs. Clinton","election_date":"2016-11-08","poll_count":377,"last_updated":"2016-11-0
8T17:20:03.000Z","url":"http://elections.huffingtonpost.com/pollster/2016-general-election-trump-vs-clinton","estimates":[{"choice":"C
linton","value":47.3,"lead_confidence":100.0,"first_name":"Hillary","last_name":"Clinton","party":"Dem","incumbent":false},{"choice":"
Trump","value":"42.0","lead_confidence":0.0,"first_name":"Donald","last_name":"Trump","party":"Rep","incumbent":false},{"choice":"Othe
r","value":5.2,"lead_confidence":null,"first_name":"","last_name":"Other","party":null,"incumbent":false}],"estimates_by_date":[{"date
":"2016-11-08","estimates":[{"choice":"Trump","value":41.98},{"choice":"Clinton","value":47.29},{"choice":"Other","value":5.17},{"choi
ce":"Undecided","value":5.57}]},{"date":"2016-11-07","estimates":[{"choice":"Trump","value":41.97},{"choice":"Clinton","value":47.29},
{"choice":"Other","value":5.17},{"choice":"Undecided","value":5.57}]},{"date":"2016-11-06","estimates":[{"choice":"Trump","value":41.9
8},{"choice":"Clinton","value":47.29},{"choice":"Other","value":5.17},{"choice":"Undecided","value":5.56}]},{"date":"2016-11-05","esti
mates":[{"choice":"Trump","value":42.02},{"choice":"Clinton","value":47.29},{"choice":"Other","value":5.1},{"choice":"Undecided","valu
e":5.59}]},{"date":"2016-11-04","estimates":[{"choice":"Trump","value":42.08},{"choice":"Clinton","value":47.32},{"choice":"Other","va
lue":5.01},{"choice":"Undecided","value":5.59}]},{"date":"2016-11-03","estimates":[{"choice":"Trump","value":42.19},{"choice":"Clinton
","value":47.42},{"choice":"Other","value":4.85},{"choice":"Undecided","value":5.54}]},{"date":"2016-11-02","estimates":[{"choice":"Tr
ump","value":42.28},{"choice":"Clinton","value":47.53},{"choice":"Other","value":4.72},{"choice":"Undecided","value":5.47}]},{"date":"
2016-11-01","estimates":[{"choice":"Trump","value":42.37},{"choice":"Clinton","value":47.64},{"choice":"Other","value":4.66},{"choice"
:"Undecided","value":5.33}]},{"date":"2016-10-31","estimates":[{"choice":"Trump","value":42.52},{"choice":"Clinton","value":47.88},{"c
hoice":"Other","value":4.63},{"choice":"Undecided","value":4.97}]},{"date":"2016-10-30","estimates":[{"choice":"Trump","value":42.76},
{"choice":"Clinton","value":48.27},{"choice":"Other","value":4.59},{"choice":"Undecided","value":4.38}]},{"date":"2016-10-29","estimat
es":[{"choice":"Trump","value":42.84},{"choice":"Clinton","value":48.49},{"choice":"Other","value":4.56},{"choice":"Undecided","value"
:4.12}]},{"date":"2016-10-28","estimates":[{"choice":"Trump","value":42.87},{"choice":"Clinton","value":48.69},{"choice":"Other","valu
e":4.55},{"choice":"Undecided","value":3.89}]},{"date":"2016-10-27","estimates":[{"choice":"Trump","value":42.68},{"choice":"Clinton",
"value":48.67},{"choice":"Other","value":4.55},{"choice":"Undecided","value":4.1}]},{"date":"2016-10-26","estimates":[{"choice":"Trump
","value":42.15},{"choice":"Clinton","value":48.3},{"choice":"Other","value":4.56},{"choice":"Undecided","value":4.99}]},{"date":"2016
-10-25","estimates":[{"choice":"Trump","value":41.66},{"choice":"Clinton","value":48.0},{"choice":"Other","value":4.67},{"choice":"Und
ecided","value":5.67}]},{"date":"2016-10-24","estimates":[{"choice":"Trump","value":41.25},
                                                        .........
```

Write the JSON to log files

It's best practice to write logs to files first

Script A

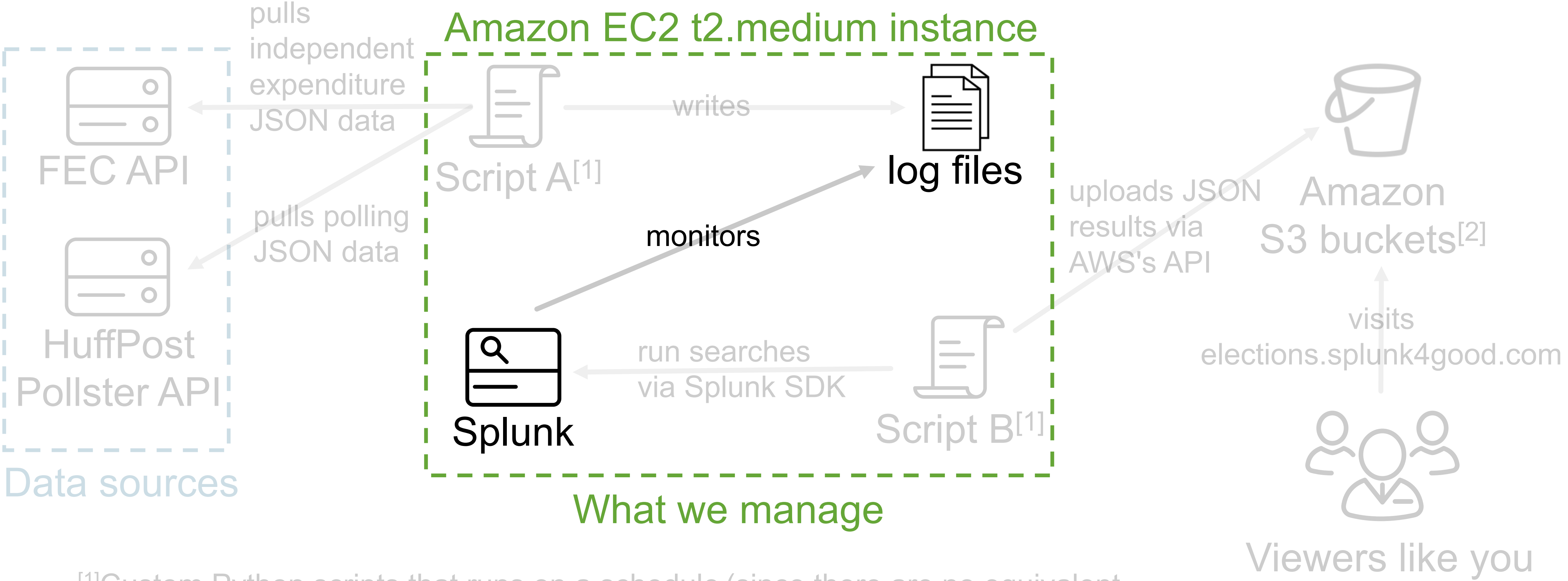
Script A runs daily, pulls from both data sources, and writes 3 files:

- `clinton_schedule_e_<DATE>.json`: the completely paginated JSON results for Clinton from FEC
- `trump_schedule_e_<DATE>.json`: and for Trump
- `polls_<DATE>.json`: the HuffPost polling chart JSON

where `<DATE>` is the date the script ran.

If you worry about using too much disk then you can set a cron job to look for files older than X days and delete it via the `find` command.

Step 3: Index the data



[1]Custom Python scripts that runs on a schedule (since there are no equivalent functionality from Splunkbase apps)

[2]Hosting html, css, and javascript as a static website (Amazon managed service)

Monitor the JSON log files

inputs.conf

```
[monitor:///home/splunk/data/*_schedule_e_*.json]
index = fec
sourcetype = fec_schedule_e
crcSalt = <SOURCE>

[monitor:///home/splunk/data/polls_*.json]
index = huffpost
sourcetype = huffpost_poll
crcSalt = <SOURCE>
```

* accommodates for different dates and `crcSalt` is set to make sure every filename gets indexed.
But before monitoring, we must set the proper props.conf and transforms.conf for both sourcetypes (continued)...

FEC JSON response

Breaking up the individual expenditures

"header"

```
{"api_version":"1.0","pagination":{"count":18207,"pages":183,"last_indexes":{"last_index":"4010420171358323494","last_expenditure_date":"2016-11-28T00:00:00"},"per_page":100},"results":[{"payee_name":"ACTBLUE TECHNICAL SERVICES","office_total_ytd":603.07,"conduit_committee_id":"C00626234","payee_street_1":"366 SUMMER STREET","report_type":"YE","expenditure_description":"CREDIT CARD PROCESSING FEES","filer_suffix":null,"original_sub_id":null,"conduit_committee_street1":null,"conduit_committee_name":null,"image_number":"201701319042196565","payee_suffix":null,"conduit_committee_city":null,"conduit_committee_zip":null,"payee_prefix":null,"independent_sign_name":"RANDOLPH, SUSANNAH","expenditure_amount":18.74,"back_reference_transaction_id":null,"file_number":1144979,"payee_middle_name":null,"cand_office_state":null,"expenditure_date":"2016-12-31T00:00:00","memo_code_full":null,"cand_office_district":null,"report_year":2016,"candidate_id":"P80001571","candidate_prefix":null,"notary_sign_name":null,"filer_first_name":"SUSANNAH","filing_form":"F3X","action_code_full":"ADD","category_code":"001","candidate_first_name":"DONALD","filer_last_name":"RANDOLPH","committee_id":"C00626234","candidate_suffix":null,"memoed_subtotal":false,"payee_city":"SOMERVILLE","election_type":"G2020","filer_prefix":null,"candidate_last_name":"TRUMP","payee_zip":"021443132","schedule_type":"SE","conduit_committee_state":null,"payee_state":"MA","conduit_committee_street2":null,"filer_middle_name":null,"candidate":{"two_year_period":2016.0,"idx":88448,"candidate_id":"P80001571"},"payee_first_name":null,"schedule_type_full":"ITEMIZED INDEPENDENT EXPENDITURES","dissemination_date":"2016-12-21T00:00:00","notary_commission_expiration_date":null,"link_id":4013120171369074356,"candidate_middle_name":"J","election_type_full":null,"action_code":"A","is_notice":false,"payee_last_name":null,"support_oppose_indicator":"S","memo_code":null,"pdf_url":"http:\/\/docquery.fec.gov\/cgi-bin\/fecimg\/?201701319042196565","payee_street_2":null,"line_number":"24","committee":{"city":"ORLANDO","party_full":null,"street_1":"701 DELANEY PARK DRIVE","cycles":[2018,2016],"party":null,"candidate_ids":[],"committee_type_full":"Super PAC (Independent Expenditure-Only)","street_2":null,"organization_type":null,"zip":"32806","designation":"U","cycle":2016,"treasurer_name":"SUSANNAH RANDOLPH","designation_full":"Unauthorized","state":"FL","organization_type_full":null,"committee_id":"C00626234","state_full":"Florida","committee_type":"O","name":"HELPING ELECT REFORMERS"},"sub_id":"4021020171370394552","independent_sign_date":"2017-01-31T00:00:00","memo_text":null,"notary_sign_date":null,"back_reference_schedule_name":null,"candidate_office":"P","category_code_full":"Administrative\/Salary\/Overhead Expenses ","candidate_name":"TRUMP, DONALD J"},{"payee_name":"WESTERN TRAILS GUN AND KNIFE SHOWS","office_total_ytd":9315895.8800000008,"conduit_committee_id":"C00580100","payee_street_1":"ATTN: KARL LANGE","report_type":"YE",

.........

"memo_text":null,"notary_sign_date":null,"back_reference_schedule_name":null,"candidate_office":"P","category_code_full":"Solicitation and Fundraising Expenses ","candidate_name":"TRUMP, DONALD J"}]}
```

extra closing brackets

FEC JSON response

Identify the time of each event

```
{"api_version":"1.0","pagination":{"count":18207,"pages":183,"last_indexes":{"last_index":"4010420171358323494","last_expenditure_date
":"2016-11-28T00:00:00"},"per_page":100},"results":[{"payee_name":"ACTBLUE TECHNICAL SERVICES","office_total_ytd":603.07,"conduit_comm
ittee_id":"C00626234","payee_street_1":"366 SUMMER STREET","report_type":"YE","expenditure_description":"CREDIT CARD PROCESSING FEES",
"filer_suffix":null,"original_sub_id":null,"conduit_committee_street1":null,"conduit_committee_name":null,"image_number":"201701319042
196565","payee_suffix":null,"conduit_committee_city":null,"conduit_committee_zip":null,"payee_prefix":null,"independent_sign_name":"RA
NDOLPH, SUSANNAH","expenditure_amount":18.74,"back_reference_transaction_id":null,"file_number":1144979,"payee_middle_name":null,"cand
_office_state":null,"expenditure_date":"2016-12-31T00:00:00","memo_code_full":null,"cand_office_district":null,"report_year":2016,"can
didate_id":"P80001571","candidate_prefix":null,"notary_sign_name":null,"filer_first_name":"SUSANNAH","filing_form":"F3X","action_code_
full":"ADD","category_code":"001","candidate_first_name":"DONALD","filer_last_name":"RANDOLPH","committee_id":"C00626234","candidate_s
uffix":null,"memoed_subtotal":false,"payee_city":"SOMERVILLE","election_type":"G2020","filer_prefix":null,"candidate_last_name":"TRUMP
","payee_zip":"021443132","schedule_type":"SE","conduit_committee_state":null,"payee_state":"MA","conduit_committee_street2":null,"fil
er_middle_name":null,"candidate":{"two_year_period":2016.0,"idx":88448,"candidate_id":"P80001571"},"payee_first_name":null,"schedule_t
ype_full":"ITEMIZED INDEPENDENT EXPENDITURES","dissemination_date":"2016-12-21T00:00:00","notary_commission_expiration_date":null,"lin
k_id":4013120171369074356,"candidate_middle_name":"J","election_type_full":null,"action_code":"A","is_notice":false,"payee_last_name":
null,"support_oppose_indicator":"S","memo_code":null,"pdf_url":"http:\/\/docquery.fec.gov\/cgi-bin\/fecimg\/?201701319042196565","paye
e_street_2":null,"line_number":"24","committee":{"city":"ORLANDO","party_full":null,"street_1":"701 DELANEY PARK DRIVE","cycles":[2018
,2016],"party":null,"candidate_ids":[],"committee_type_full":"Super PAC (Independent Expenditure-Only)","street_2":null,"organization_
type":null,"zip":"32806","designation":"U","cycle":2016,"treasurer_name":"SUSANNAH RANDOLPH","designation_full":"Unauthorized","state"
:"FL","organization_type_full":null,"committee_id":"C00626234","state_full":"Florida","committee_type":"O","name":"HELPING ELECT REFOR
MERS"},"sub_id":"4021020171370394552","independent_sign_date":"2017-01-31T00:00:00","memo_text":null,"notary_sign_date":null,"back_ref
erence_schedule_name":null,"candidate_office":"P","category_code_full":"Administrative\/Salary\/Overhead Expenses ","candidate_name":"
TRUMP, DONALD J"},{"payee_name":"WESTERN TRAILS GUN AND KNIFE SHOWS","office_total_ytd":9315895.8800000008,"conduit_committee_id":"C00
580100","payee_street_1":"ATTN: KARL LANGE","report_type":"YE","expenditure_description":"VOID - BOOTH RENTAL - EVENT CANCELLED","file
r_suffix":null,"original_sub_id":null,"conduit_committee_street1":null,"conduit_committee_name":null,"image_number":"20170504905350522
3","payee_suffix":null,"conduit_committee_city":null,"conduit_committee_zip":null,"cand_office_state":null,"independent_sign_name":"AD
KINS, MARY ROSE","expenditure_amount":-9.17,"back_reference_transaction_id":null,"file_number":1161245,"payee_middle_name":null,"payee
_prefix":null,"expenditure_date":"2016-12-30T00:00:00","memo_code_full":null,"cand_office_district":null,"report_year":2016,
                              .........
```

FEC Splunk settings

For proper line breaks, timestamps, and field extractions

props.conf

```
[fec_schedule_e]
LINE_BREAKER = (,){"payee_name"
TRUNCATE = 7000
SHOULD_LINEMERGE = false
TIME_PREFIX = expenditure_date":"
TIME_FORMAT = %F
MAX_TIMESTAMP_LOOKAHEAD = 10
MAX_DAYS_AGO = 10951
SEDCMD-0 = s/^{.+?"results":\[//
SEDCMD-1 = s/]}$//

KV_MODE = json
```

Remove the "header" from the first event

Remove the extra closing brackets from the last event

HuffPost Pollster JSON response

Also in similar format

"header"

```
{"id":624,"title":"2016 General Election: Trump vs. Clinton","slug":"2016-general-election-trump-vs-clinton","topic":"2016-president",
"state":"US","short_title":"2016 President: Trump vs. Clinton","election_date":"2016-11-08","poll_count":377,"last_updated":"2016-11-0
8T17:20:03.000Z","url":"http://elections.huffingtonpost.com/pollster/2016-general-election-trump-vs-clinton","estimates":[{"choice":"C
linton","value":47.3,"lead_confidence":100.0,"first_name":"Hillary","last_name":"Clinton","party":"Dem","incumbent":false},{"choice":"
Trump","value":"42.0","lead_confidence":0.0,"first_name":"Donald","last_name":"Trump","party":"Rep","incumbent":false},{"choice":"Othe
r","value":5.2,"lead_confidence":null,"first_name":"","last_name":"Other","party":null,"incumbent":false}],"estimates_by_date":[{"date
":"2016-11-08","estimates":[{"choice":"Trump","value":41.98},{"choice":"Clinton","value":47.29},{"choice":"Other","value":5.17},{"choi
ce":"Undecided","value":5.57}]},{"date":"2016-11-07","estimates":[{"choice":"Trump","value":41.97},{"choice":"Clinton","value":47.29},
{"choice":"Other","value":5.17},{"choice":"Undecided","value":5.57}]},{"date":"2016-11-06","estimates":[{"choice":"Trump","value":41.9
8},{"choice":"Clinton","value":47.29},{"choice":"Other","value":5.17},{"choice":"Undecided","value":5.56}]},{"date":"2016-11-05","esti
mates":[{"choice":"Trump","value":42.02},{"choice":"Clinton","value":47.29},{"choice":"Other","value":5.1},{"choice":"Undecided","valu
e":5.59}]},{"date":"2016-11-04","estimates":[{"choice":"Trump","value":42.08},{"choice":"Clinton","value":47.32},{"choice":"Other","va
lue":5.01},{"choice":"Undecided","value":5.59}]},{"date":"2016-11-03","estimates":[{"choice":"Trump","value":42.19},{"choice":"Clinton
","value":47.42},{"choice":"Other","value":4.85},{"choice":"Undecided","value":5.54}]},{"date":"2016-11-02","estimates":[{"choice":"Tr
ump","value":42.28},{"choice":"Clinton","value":47.53},{"choice":"Other","value":4.72},{"choice":"Undecided","value":5.47}]},{"date":"
2016-11-01","estimates":[{"choice":"Trump","value":42.37},{"choice":"Clinton","value":47.64},{"choice":"Other","value":4.66},{"choice"
:"Undecided","value":5.33}]},{"date":"2016-10-31","estimates":[{"choice":"Trump","value":42.52},{"choice":"Clinton","value":47.88},{"c
hoice":"Other","value":4.63},{"choice":"Undecided","value":4.97}]},{"date":"2016-10-30","estimates":[{"choice":"Trump","value":42.76},
{"choice":"Clinton","value":48.27},{"choice":"Other","value":4.59},{"choice":"Undecided","value":4.38}]},{"date":"2016-10-29","estimat
es":[{"choice":"Trump","value":42.84},{"choice":"Clinton","value":48.49},{"choice":"Other","value":4.56},{"choice":"Undecided","value"
:4.12}]},{"date":"2016-10-28","estimates":[{"choice":"Trump","value":42.87},{"choice":"Clinton","value":48.69},{"choice":"Other","valu
e":4.55},{"choice":"Undecided","value":3.89}]},{"date":"2016-10-27","estimates":[{"choice":"Trump","value":42.68},{"choice":"Clinton",
"value":48.67},{"choice":"Other","value":4.55},{"choice":"Undecided","value":4.1}]},
                                    .........
{"date":"2015-05-19","estimates":[{"choice":"Trump","value":33.79},{"choice":"Clinton","value":52.5},{"choice":"Other","value":3.94},{
"choice":"Undecided","value":9.78}]}]}
```

extra closing brackets

splunk> .conf2017

HuffPost Pollster Splunk settings

Similar format means similar settings

props.conf

```
[huffpost_poll]
LINE_BREAKER = (,){"date"
TRUNCATE = 2000
SHOULD_LINEMERGE = false
TIME_PREFIX = date":"
TIME_FORMAT = %F
MAX_TIMESTAMP_LOOKAHEAD = 10
MAX_DAYS_AGO = 10951
SEDCMD-0 = s/^{.+?,"estimates_by_date":\[//
SEDCMD-1 = s/]}]}$/]}/

REPORT-0 = huffpost_poll_kv
KV_MODE = json
```

Remove the "header" from the first event

Remove the extra closing brackets from the last event

Continued in transforms.conf (continued)...

splunk> .conf2017

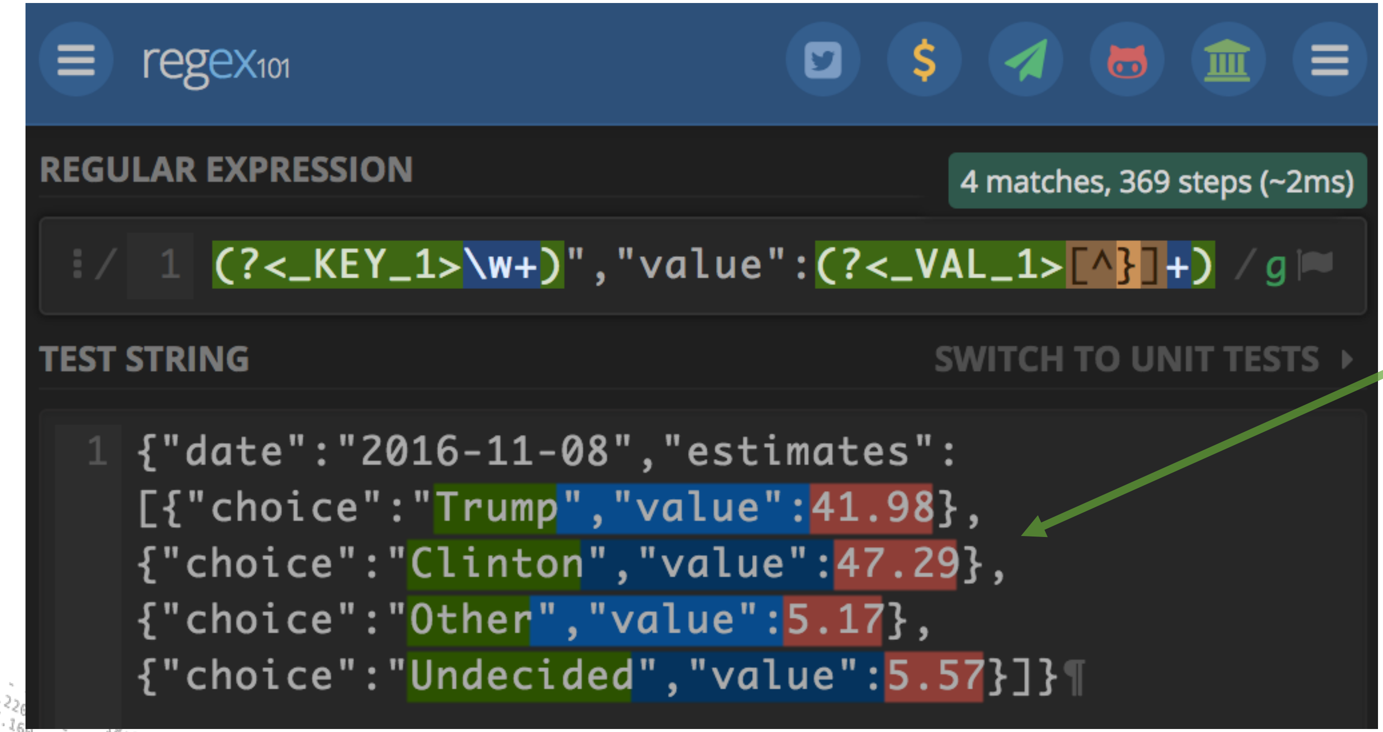
HuffPost Pollster Splunk settings

Dynamic field name extractions

transforms.conf

```
[huffpost_poll_kv]
REGEX = (?<_KEY_1>\w+)","value":(?<_VAL_1>[^}]+)
```

Referenced by props.conf



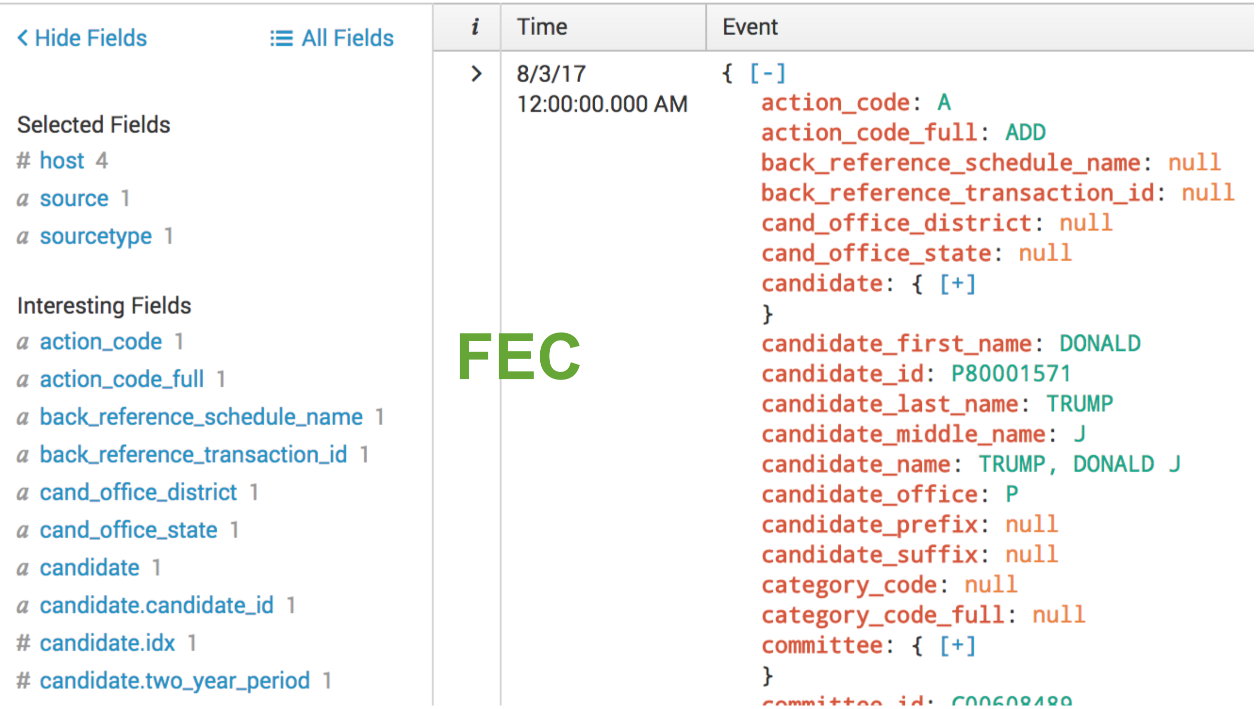
```
/ (?<_KEY_1>\w+)","value":(?<_VAL_1>[^}]+) /g
```

```
{"date":"2016-11-08","estimates":
[{"choice":"Trump","value":41.98},
{"choice":"Clinton","value":47.29},
{"choice":"Other","value":5.17},
{"choice":"Undecided","value":5.57}]}¶
```

The **green** capture is the field name (_KEY_1)
The **red** capture is the value of the field (_VAL_1)

splunk> .conf2017

Clean data in Splunk!



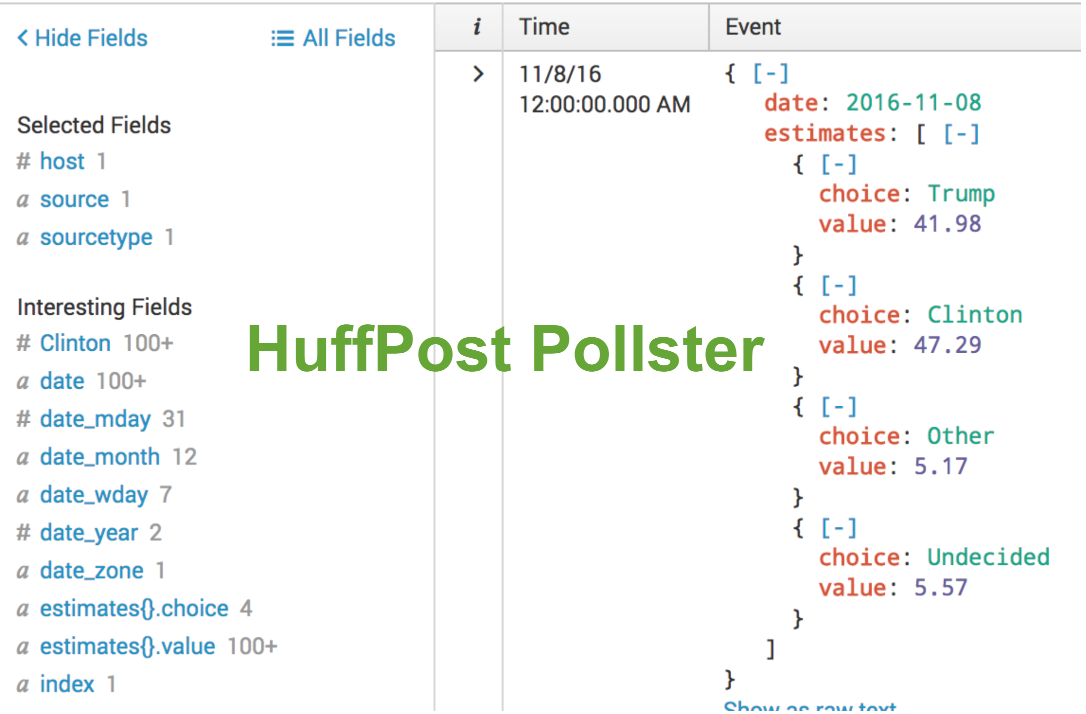
Selected Fields
- # host 4
- a source 1
- a sourcetype 1

Interesting Fields
- a action_code 1
- a action_code_full 1
- a back_reference_schedule_name 1
- a back_reference_transaction_id 1
- a cand_office_district 1
- a cand_office_state 1
- a candidate 1
- a candidate.candidate_id 1
- # candidate.idx 1
- # candidate.two_year_period 1

> 8/3/17 12:00:00.000 AM

```
{ [-]
   action_code: A
   action_code_full: ADD
   back_reference_schedule_name: null
   back_reference_transaction_id: null
   cand_office_district: null
   cand_office_state: null
   candidate: { [+]
   }
   candidate_first_name: DONALD
   candidate_id: P80001571
   candidate_last_name: TRUMP
   candidate_middle_name: J
   candidate_name: TRUMP, DONALD J
   candidate_office: P
   candidate_prefix: null
   candidate_suffix: null
   category_code: null
   category_code_full: null
   committee: { [+]
   }
   committee_id: C00608489
```

FEC



Selected Fields
- # host 1
- a source 1
- a sourcetype 1

Interesting Fields
- # Clinton 100+
- a date 100+
- # date_mday 31
- a date_month 12
- a date_wday 7
- # date_year 2
- a date_zone 1
- a estimates{}.choice 4
- a estimates{}.value 100+
- a index 1

> 11/8/16 12:00:00.000 AM

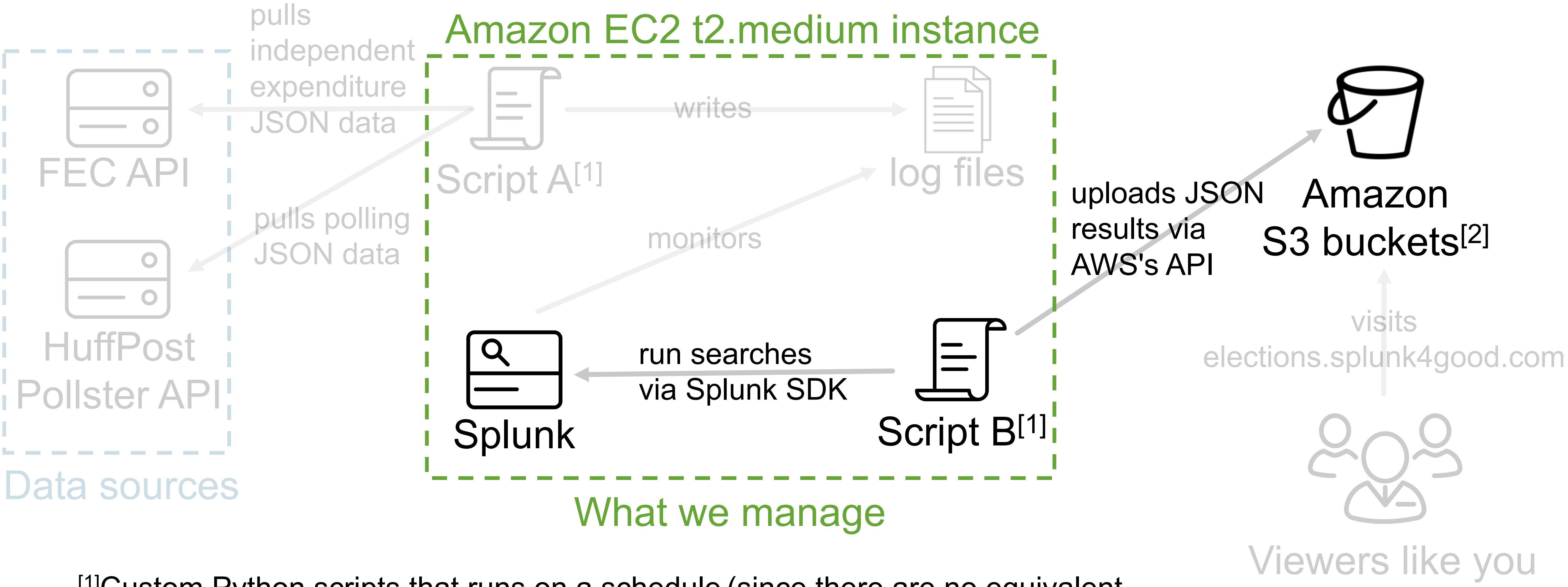
```
{ [-]
   date: 2016-11-08
   estimates: [ [-]
     { [-]
       choice: Trump
       value: 41.98
     }
     { [-]
       choice: Clinton
       value: 47.29
     }
     { [-]
       choice: Other
       value: 5.17
     }
     { [-]
       choice: Undecided
       value: 5.57
     }
   ]
}
Show as raw text
```

HuffPost Pollster

- Timestamps are properly set
- Each event is a valid JSON thanks to the LINE_BREAKER and SEDCMD regexes (malformed JSON won't have color highlighting)
- JSON key values are automatically extracted
- The dynamic field extraction from transforms.conf creates the "Trump", "Clinton", "Other", and "Undecided" fields for Pollster events

splunk> .conf2017

Step 4: Upload the data



Amazon EC2 t2.medium instance

pulls independent expenditure JSON data

FEC API

pulls polling JSON data

HuffPost Pollster API

Data sources

Script A[1]

writes

log files

monitors

Splunk

run searches via Splunk SDK

Script B[1]

What we manage

uploads JSON results via AWS's API

Amazon S3 buckets[2]

visits elections.splunk4good.com

Viewers like you

[1]Custom Python scripts that runs on a schedule (since there are no equivalent functionality from Splunkbase apps)

[2]Hosting html, css, and javascript as a static website (Amazon managed service)

Running searches and uploading to S3

Script B uses the Splunk SDK to authenticate and run 3 searches. Splunk returns the search results in JSON:

- `stats.json` (groups the expenditures by candidate, committees, and supporting/opposing)
- `timechart.json` (correlates the expenditures with polls)
- `latest.json` (simply gets the current time and last expenditure date[1])

The script knows to search only the latest dataset by using the correct date for `source`.

Then it uses AWS API to authenticate and upload these files to the S3 bucket.

[1]The expenditures can be delayed by about a month since the committees have filing deadlines, ie they only need to file their completed reports every month or so. Remember we excluded 24- and 48- reports via `is_notice=false` for the REST API.

The Splunk searches

Even the searches ain't easy

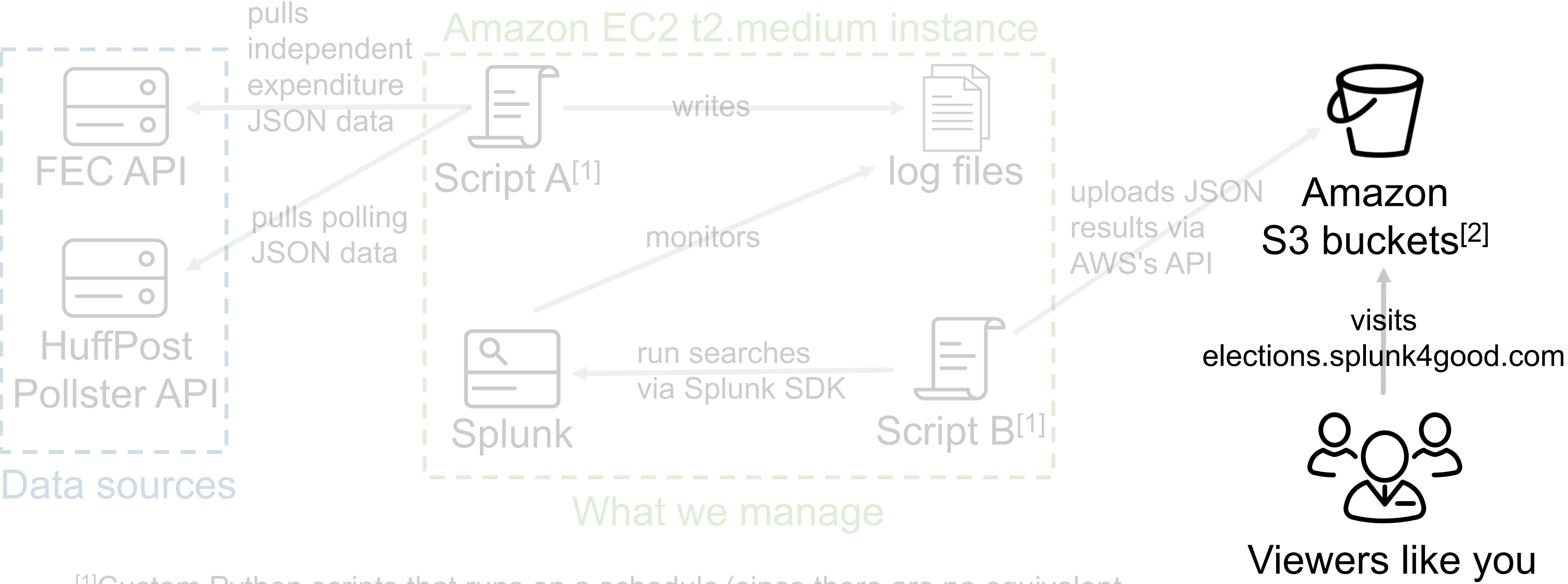
stats.json

```
index=fec sourcetype=fec_schedule_e
| stats sum(expenditure_amount) as spent by committee_id committee.committee_type_full
committee.name toward candidate candidate_id
| sort 0 -spent
| streamstats count as rank by toward candidate
| eval committee_id=if(rank<=5, committee_id, "none")
| eval committee.name=if(rank<=5, 'committee.name', "others ".toward." ".candidate)
| eval committee.committee_type_full=if(rank<=5, 'committee.committee_type_full', "none")
| stats sum(spent) as spent by committee_id committee.name committee.committee_type_full toward
candidate candidate_id
```

timechart.json

```
(index=fec sourcetype=fec_schedule_e) OR (index=huffpost sourcetype=huffpost_poll)
| rename Trump as poll_trump Clinton as poll_clinton
| eval id="fec"."_".candidate."_".toward
| timechart span=1w sum(expenditure_amount) avg(poll_trump) avg(poll_clinton) by id
| rename "avg(*): NULL" as * "sum(expenditure_amount): *" as *
| fillnull
```

Step 5: Serve the data



Data sources

FEC API

HuffPost Pollster API

pulls independent expenditure JSON data

pulls polling JSON data

Amazon EC2 t2.medium instance

Script A[1]

writes

log files

monitors

Splunk

run searches via Splunk SDK

Script B[1]

What we manage

uploads JSON results via AWS's API

Amazon S3 buckets[2]

visits elections.splunk4good.com

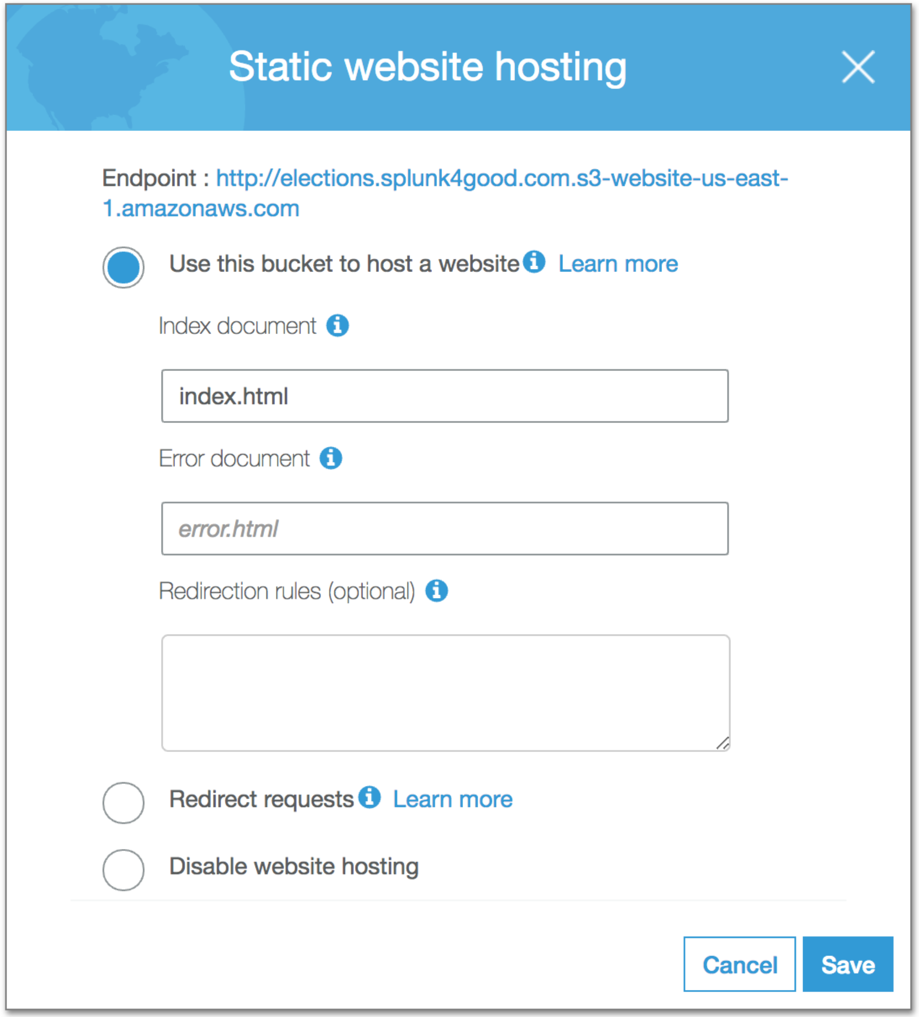
Viewers like you

[1] Custom Python scripts that runs on a schedule (since there are no equivalent functionality from Splunkbase apps)

[2] Hosting html, css, and javascript as a static website (Amazon managed service)

Let Amazon handle the "web server"

Pay as you go



Static website hosting

Endpoint : http://elections.splunk4good.com.s3-website-us-east-1.amazonaws.com

● Use this bucket to host a website Learn more

Index document: index.html

Error document: error.html

Redirection rules (optional)

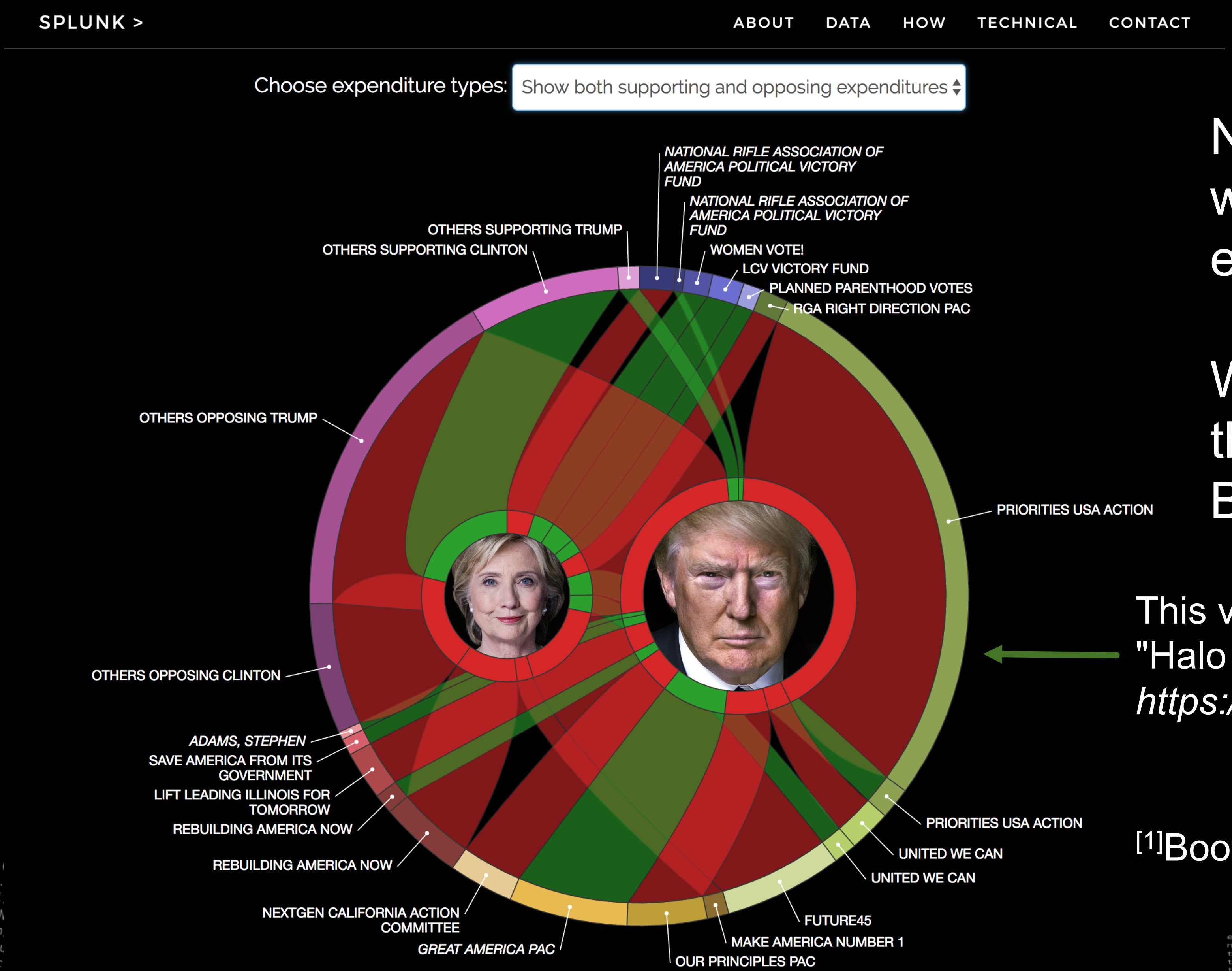
○ Redirect requests Learn more

○ Disable website hosting

S3 is a managed service, which means we don't need to administer or scale our own web servers.

If we need even more performance, then we can use Amazon CloudFront (CDN) for multiple regional caching.

The website



Choose expenditure types: Show both supporting and opposing expenditures

No need to reinvent the wheel when we can just search for existing free themes and styles!

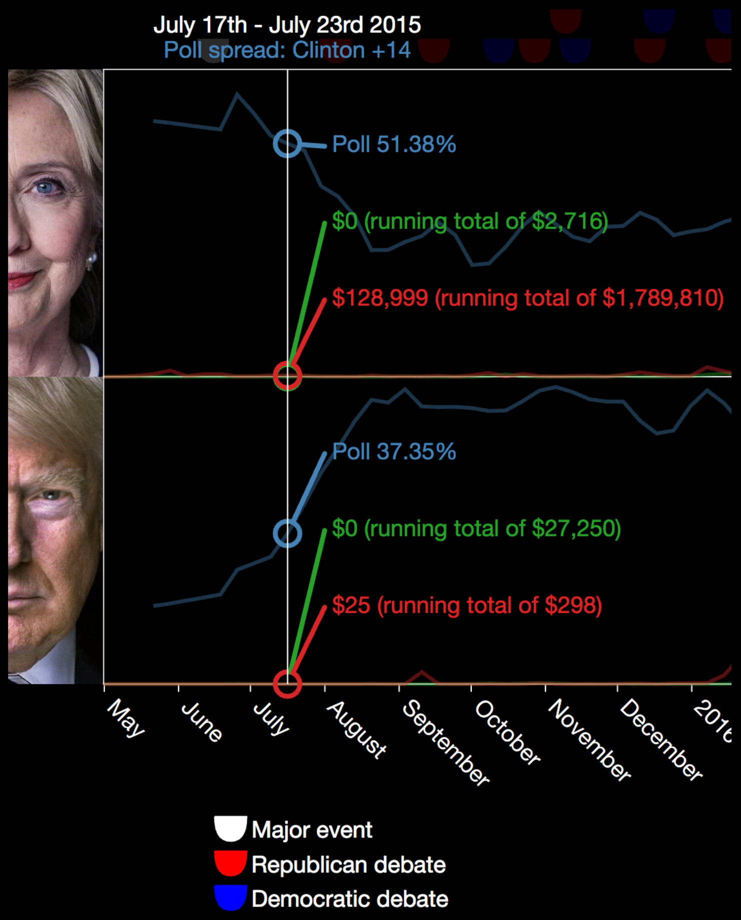
We modified a Bootstrap[1] theme called "Grayscale" by Blackrock Digital for the site.

This visualization is available as an app called "Halo – Custom Visualization" on SplunkBase: *https://splunkbase.splunk.com/app/3514/*

[1]Bootstrap is a front-end framework by Twitter

The JavaScript (JS) magic

Parsing and displaying the JSON data



- RequireJS loads all the necessary JS libraries
- D3.js asynchronously loads the 3 JSON files and "loops" through the JSON to draw the visualizations using <SVG> elements
- D3.js also uses Underscore.js to heavily manipulate and format the JSON for easier parsing

D3.js is *not* easy... you must draw almost every line and shape from scratch. Your math- and coordinate-fu must be strong.

The data challenges

It gets even harder...

Confession: Every regex for FEC in this presentation is in "easy mode".

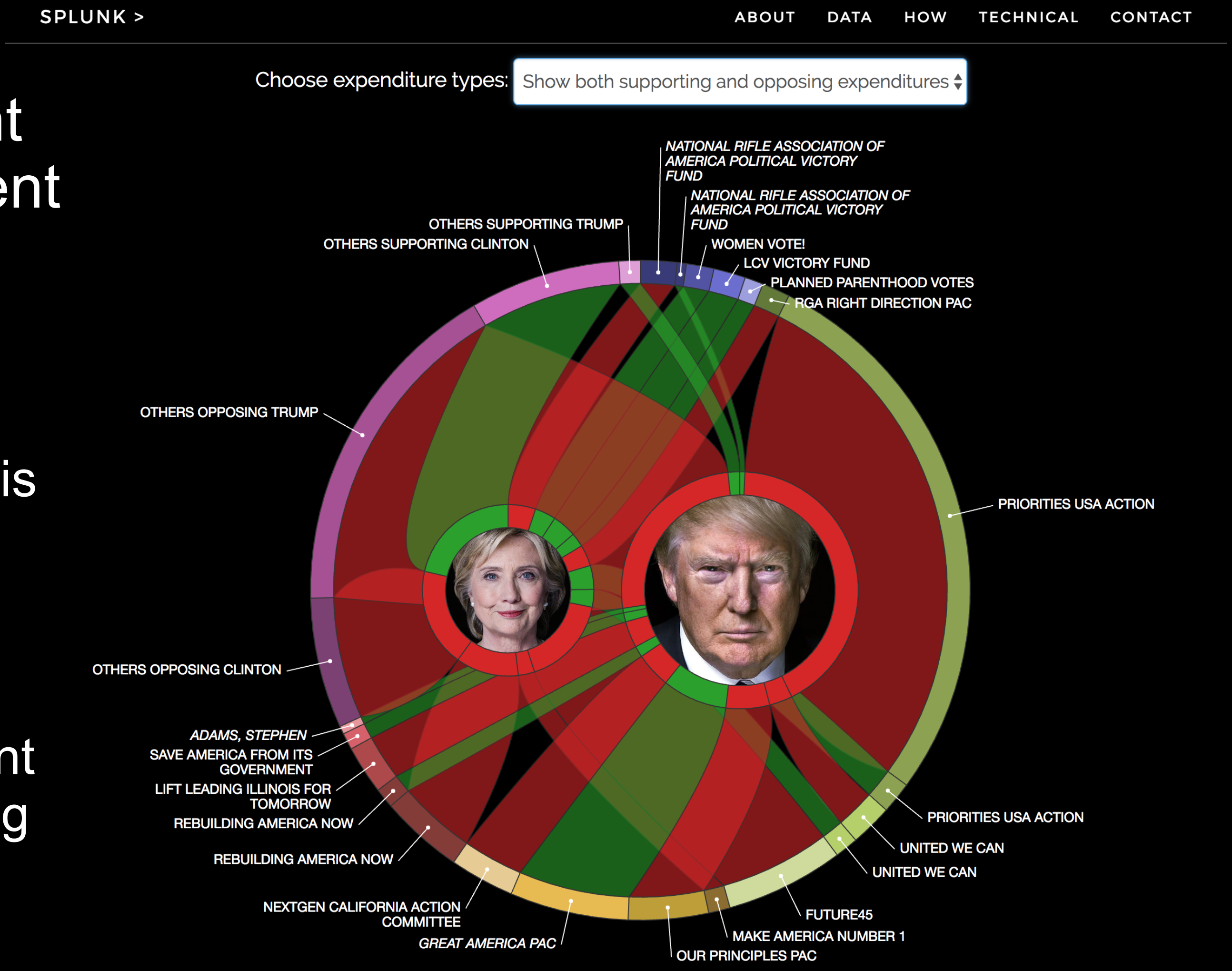
- The FEC API was in early beta during the election (with incorrect values).
- The FEC JSON key order is inconsistent, which is valid for JSON, but this makes the regex much more complicated.
  - For example, we fall back on `dissemination_date` if `expenditure_date` is undefined. But since the key order is inconsistent, the regex becomes very complicated:
  - `TIME_PREFIX = expenditure_date":"|dissemination_date":"(?=.+?expenditure_date":null)|expenditure_date":null.+?dissemination_date":"`
- We have to pull the complete FEC results every time (due to a new pagination's `last_index`). Indexing historical data repeatedly creates "bucket spread" and can slow down searches (but I know what I'm doing).

Findings

Not what you would have expected

Conclusions

- There was a lot of soft money spent in the 2016 election, but wasn't spent in the ways that you might expect
  - $417,457,906 spent just on Clinton and Trump – this is only soft money
  - 56% of money all soft money spent in this race went to defeat Trump ($234M)
  - 86% of money spent on Trump was opposing him
  - Clinton was no angel: 60% of funds spent on Clinton ($85.5M) were spent opposing her



SPLUNK > ABOUT DATA HOW TECHNICAL CONTACT

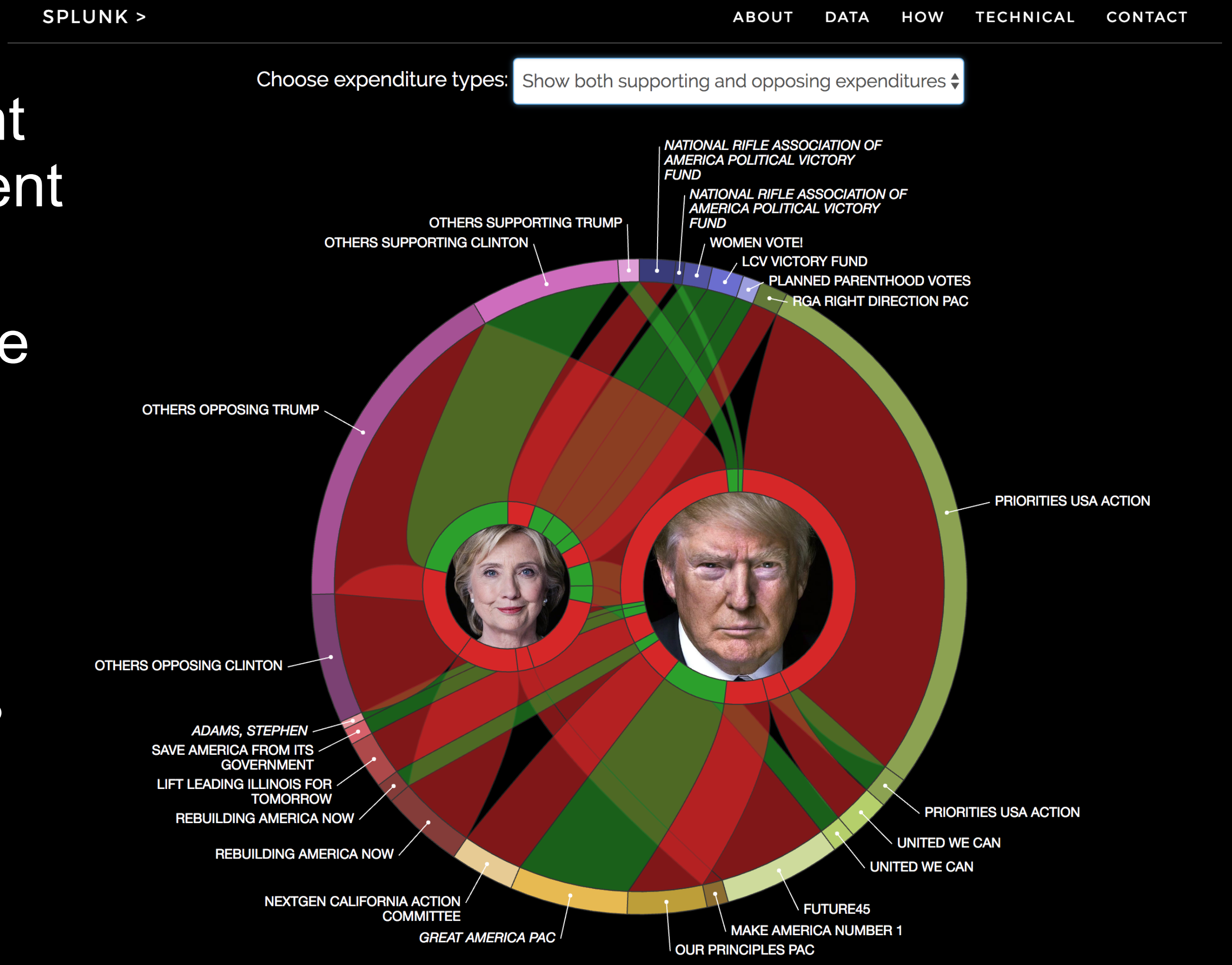
Choose expenditure types: Show both supporting and opposing expenditures

Findings

Not what you would have expected

Conclusions

- There was a lot of soft money spent in the 2016 election, but wasn’t spent in the ways that you might expect
- So much good information available
- There are some weaknesses in election reporting
  - Where (and for what) are funds actually expended?
  - From whom do funds actually originate?
  - Clearly there are new challenges with tracking of foreign spending in online advertising spend



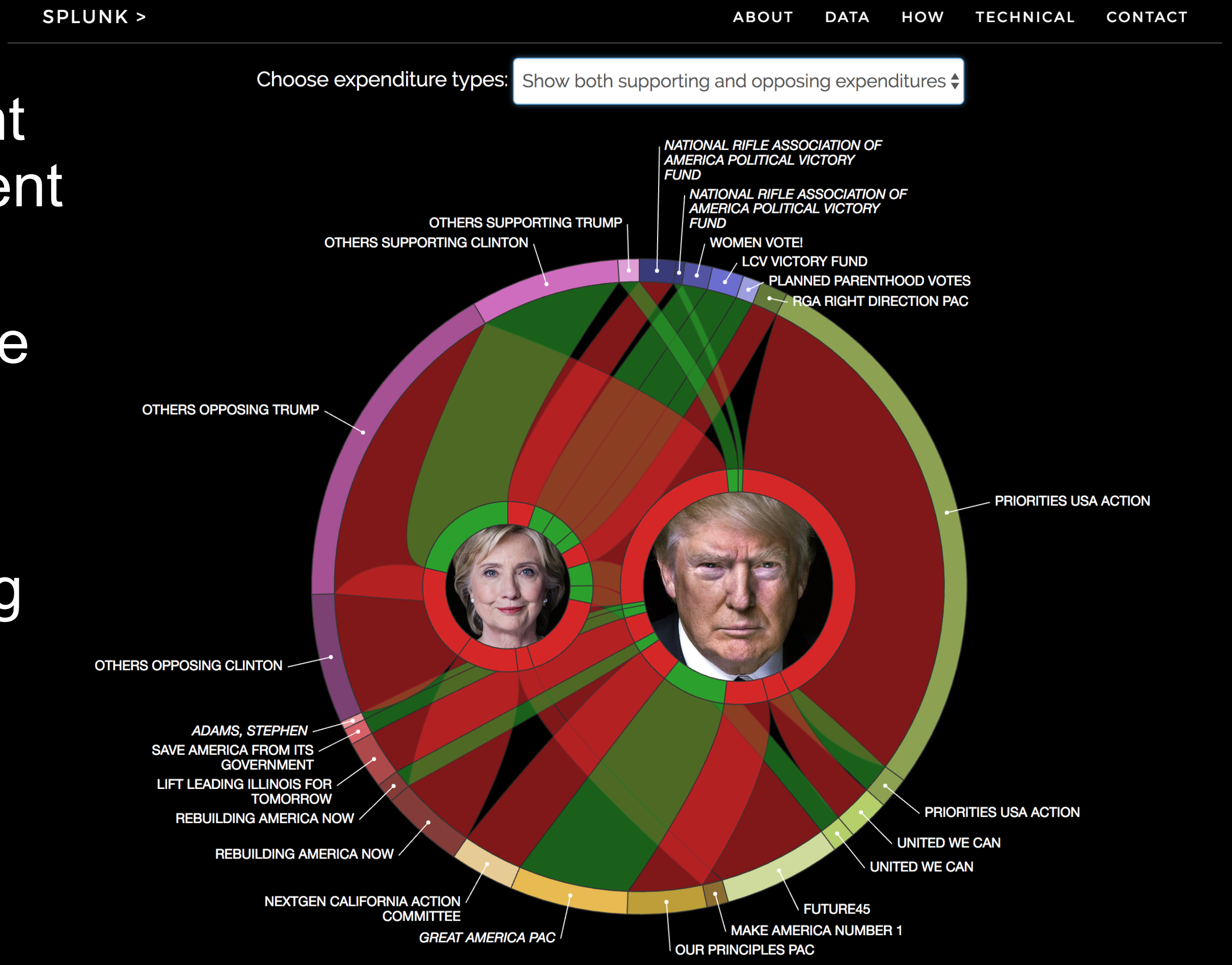
Choose expenditure types: Show both supporting and opposing expenditures

Findings

Not what you would have expected

Conclusions

- There was a lot of soft money spent in the 2016 election, but wasn't spent in the ways that you might expect
- So much good information available
- There are some weaknesses in election reporting
- Spending has not stopped following the campaign



SPLUNK >

ABOUT DATA HOW TECHNICAL CONTACT

Choose expenditure types: Show both supporting and opposing expenditures

elections.splunk4good.com

Closing remarks

Corey Marshall | Splunk4Good Director

As flexible as you think Splunk is…

Big data can make a big difference

Lots of opportunities to make an impact with data and Splunk

- Fascinating way to explore the impacts of money on our electoral system
- Lots of data available right under our noses, but very few are aware of it

Splunk is a powerful tool to explore interesting and impactful new use cases

- Great way to experiment with Splunk outside of traditional IT
- Find ways to leverage open and public data sources to enrich your work
- Showcase Splunk to an entirely new audience through compelling visualizations

There's always more we can do

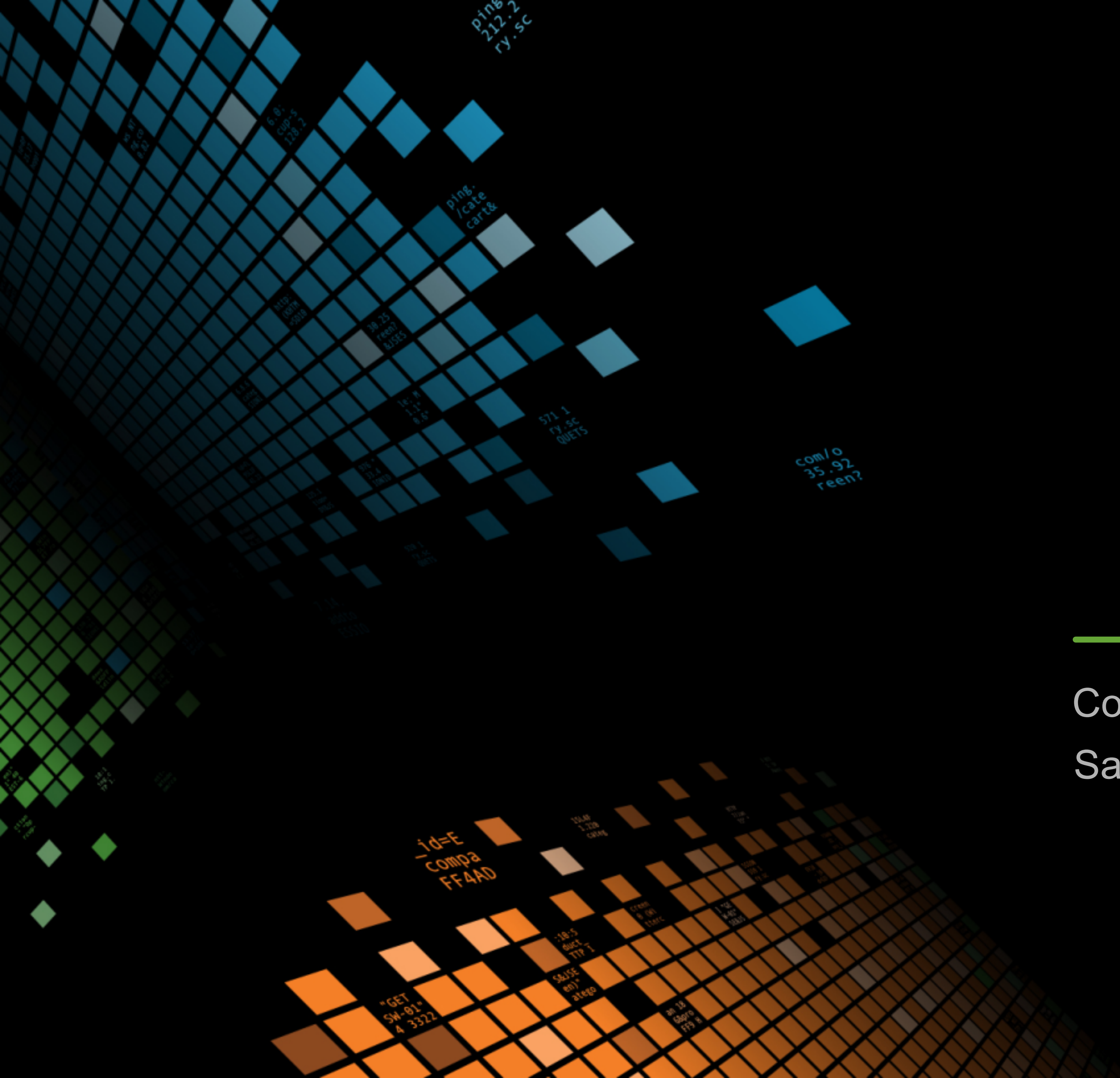
- Interesting use case that improves visibility and transparency
- What other causes could benefit from Splunk expertise?

Thank You!

Shout-out to the **18F group** on continual feedbacks during the development of the FEC API
Shout-out to **Eric Grant** as a our content delivery manager

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017



Q&A

Corey Marshall | Splunk4Good Director

Satoshi Kawasaki | Splunk4Good Ninja