SEARCH.GOV PaaS ARCHITECTURE

General Services Administration

Dawn McCleskey — Program Manager, Search.gov

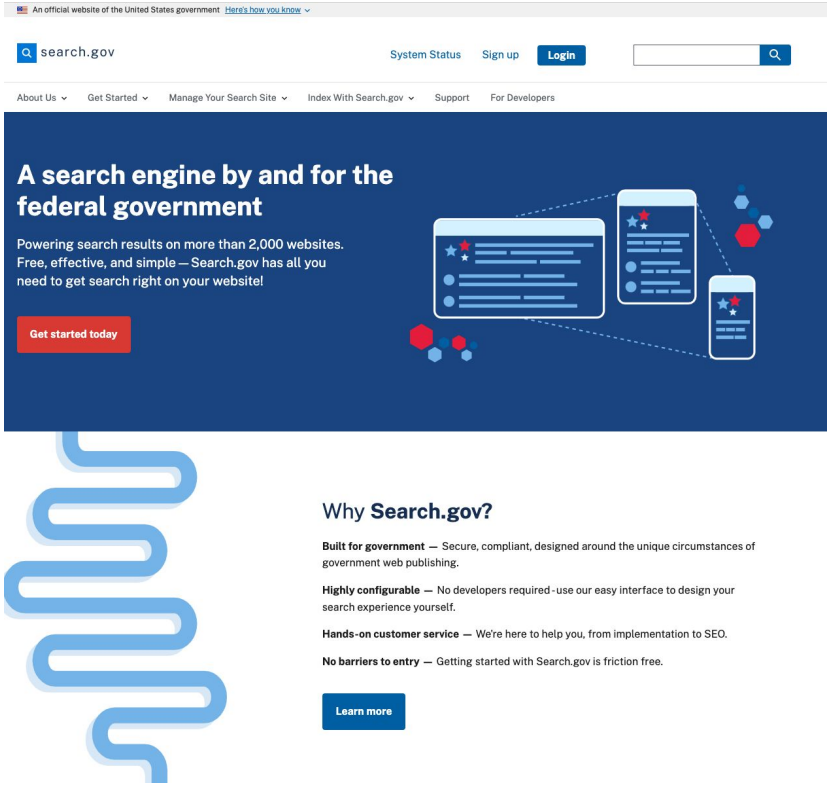
CIVIC DIGITAL FELLOWSHIP

GSA

BENJAMIN DAHROOGE
University of Rhode Island
Computer Science & Cybersecurity

WHAT IS SEARCH.GOV?

- GSA/TTS developed **SaaS** search engine for federal government websites
- Powers the intra-site search boxes on ~33% of federal .gov websites
- Enables searching all federal and state government websites on usa.gov
- Integrations built for government (e.g., USAjobs API, Federal Register)
- Free for federal agencies to use

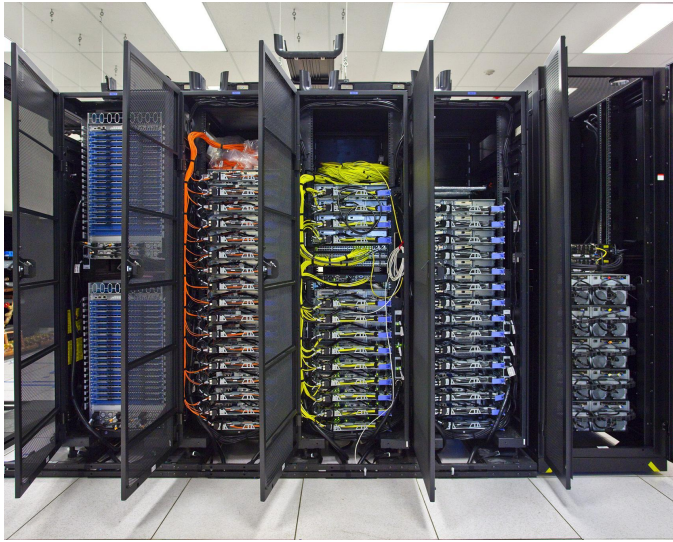


Search.gov Homepage/Documentation, August 2022

CIVIC DIGITAL FELLOWSHIP

PROJECT BACKGROUND

- Exploring the feasibility and procedures to run on Cloud.gov, a Cloud Foundry based **PaaS**
    - Deploy code, minimal configuration
- Existing Search.gov infrastructure exists primarily in **IaaS** (AWS EC2 virtual machines)
    - Provision & manage virtual infrastructure (e.g., disks, CPU, memory, network/IPs, load balancing, database)
    - Deploy and update code programatically



CIVIC DIGITAL FELLOWSHIP

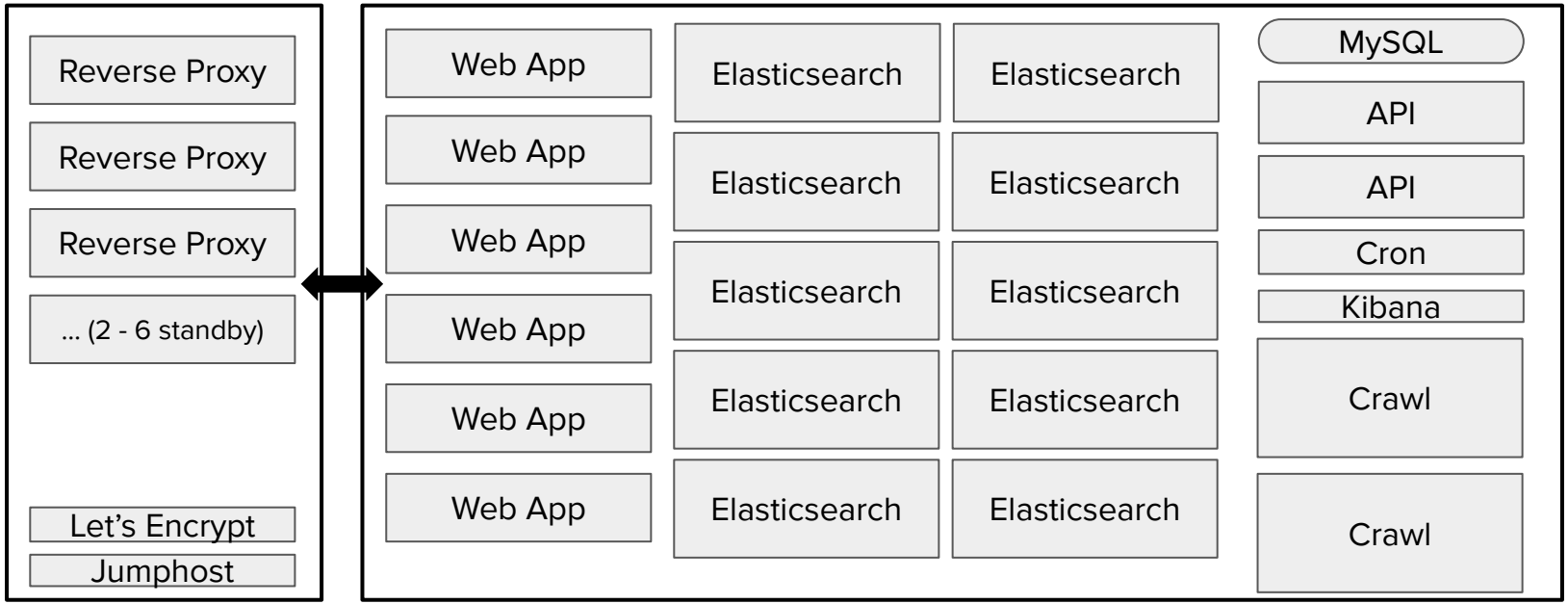
PLATFORM-AS-A-SERVICE

- **Software-as-a-Service**
  (e.g., Google Suite, Search.gov)
  Someone else manages the software and hardware, you are the end user
- **Platform-as-a-Service**
  (e.g., Cloud.gov, AWS Beanstalk, Heroku)
  Bring your code, minor configurations
- **Infrastructure-as-a-Service**
  (e.g., AWS EC2, GCP Compute Engine)
  Bring OS, software, configuration, firewall, handle security patches manually

Currently, Search.gov runs directly on AWS IaaS, making the search.gov team responsible for platform-level tasks that could be centralized.

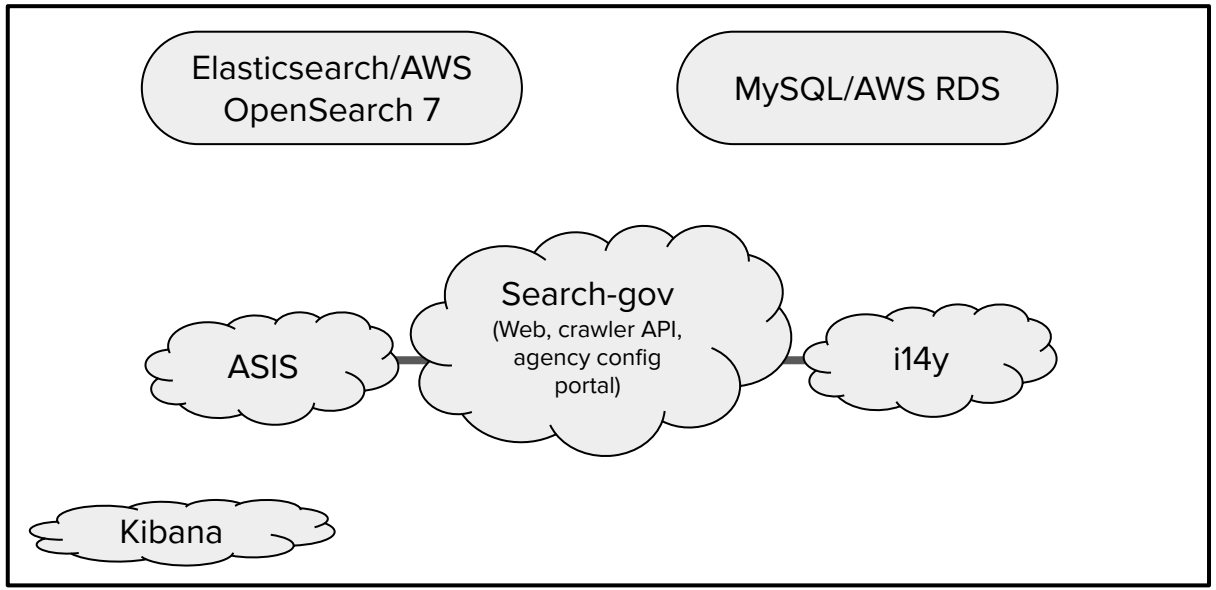
CIVIC DIGITAL FELLOWSHIP

CURRENT INFRASTRUCTURE - AWS



Each box represents a VM instance, with larger boxes denoting more computing resources (e.g., vCPU, memory)

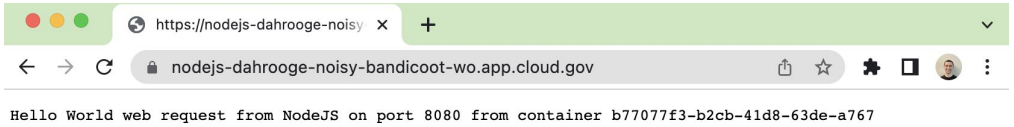
TARGET STATE IN CLOUD.GOV



Methodology changes from provisioning servers, to thinking about individual apps;
Cloud denotes a scalable containerized app, with 1 or more instance(s)

TESTING PROCESS

- Experiment with sandbox (Hello World examples) ✅
- Determine necessary capabilities (custom domains, jumphost/SSH, service brokers, backend processes, container-to-container networking) ✅
- Test various deployment methods (Supported Buildpack, Custom Buildpack, Docker image) ✅
- Test service brokers & container-to-container networking
- Full test; *aka "Monster Test Plan"* - verifies all functionality



Screenshot of Hello World app running in Cloud.gov sandbox

CIVIC DIGITAL FELLOWSHIP

SUPPORTED BUILDPACK

- Cloud Foundry buildpacks contain instructions, dependencies and configuration to run an app written in a particular language/framework
    - Docker base images before they were cool 😎
- Search-gov (github.com/GSA/search-gov) is a Ruby on Rails application, that powers most of search.gov's core functionality (e.g., web requests, API requests, crawling)

```
applications:
- name: ruby-searchgov
 random-route: false
 buildpacks:
   - ruby_buildpack
 memory: 512MB
```

manifest.yml, minimum configuration needed for Cloud.gov

```
cf push
```

```
Using manifest file ...

Creating app ruby-searchgov in org / space
OK
requested state:    started
**routes:           ruby-searchgov.app.cloud.gov**
last uploaded:      Tue 06 June 13:40:39 PST 2022
stack:              cflinuxfs3
buildpacks:         ruby_buildpack

state     since #0    running    2022-06-06 02:53:29 PM
```

Example command output when running `cf push`, adapted from Cloud.gov documentation

CIVIC DIGITAL FELLOWSHIP

SUPPORTED BUILDPACK

- To minimize footprint & security exposure, supported buildpacks are barebones for each supported language
- More security capabilities and support from Cloud.gov team
- Result: Unable to install OS-level packages needed by search-gov

```
Fetching active_scaffold 3.6.10
Installing active_scaffold 3.6.10
Fetching active_scaffold_export 3.4.0
Installing active_scaffold_export 3.4.0
Gem::Ext::BuildError: ERROR: Failed to build gem native extension.

current directory:
/tmp/contents4194593497/deps/0/vendor_bundle/ruby/2.7.0/gems/cld3-3.4.4/ext/cld3
/tmp/contents4194593497/deps/0/ruby/bin/ruby -I
/tmp/contents4194593497/deps/0/ruby/lib/ruby/site_ruby/2.7.0 -r
./siteconf20220628-90-uoiyz3.rb extconf.rb
Failed to locate protobuf
```

Error message about missing native dependencies when running `cf push`

Packages

Use the package manager of your choice to install the following packages:

- C++ compiler - required by the cld3 gem, which we use for language detection
- Google's protocol buffers - also required by the cld3 gem
- Java Runtime Environment
- ImageMagick - required by the Paperclip gem, used for image attachments
- MySQL client - required by the mysql2 gem
- V8 - required by the libv8 gem

Screenshot of the GSA/search-gov Github Readme containing a list of native dependencies

CIVIC DIGITAL FELLOWSHIP

DOCKER IMAGE

- Cloud.gov/Cloud Foundry also supports deploying Docker images, a flexible widely adopted method for deploying applications
- Based on **containerd** - Isolates processes while ensuring access to specific system libraries
- Docker can be run on almost any machine, therefore allows for local testing; portable to another PaaS if needed
- Container image gets uploaded a to container registry

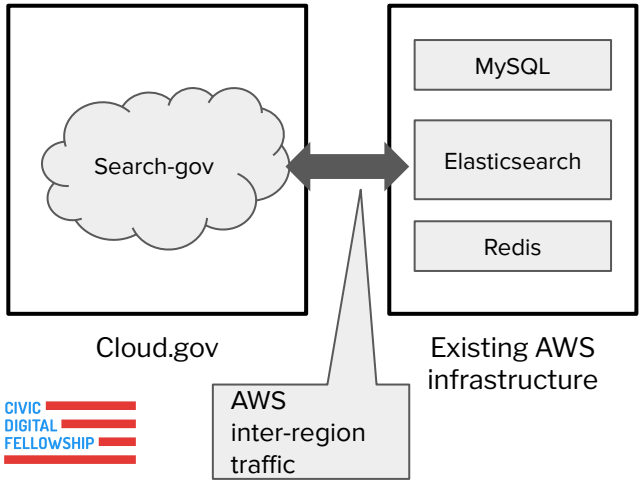
```
FROM ruby:2.7.5

# Install OS package dependencies needed by gems
RUN apt-get update
RUN apt-get install protobuf-compiler libprotobuf-dev imagemagick
default-jre default-mysql-client -y

WORKDIR /app

# Copy Gemfile over to working directory
# This allows cached gem installation to be cached
# if there are no changes to the gem files
COPY Gemfile /app
COPY Gemfile.lock /app

RUN gem install bundler -v 2.3.8
RUN bundle install

# Copy the application to working directory
COPY . /app

# Copy the development secrets
RUN cp config/secrets.yml.dev config/secrets.yml

EXPOSE 3000

# Run the rails app, listening on all interfaces
CMD ["rails", "server", "-b", "0.0.0.0"]
```

Search-gov example Dockerfile configuration

CIVIC DIGITAL FELLOWSHIP

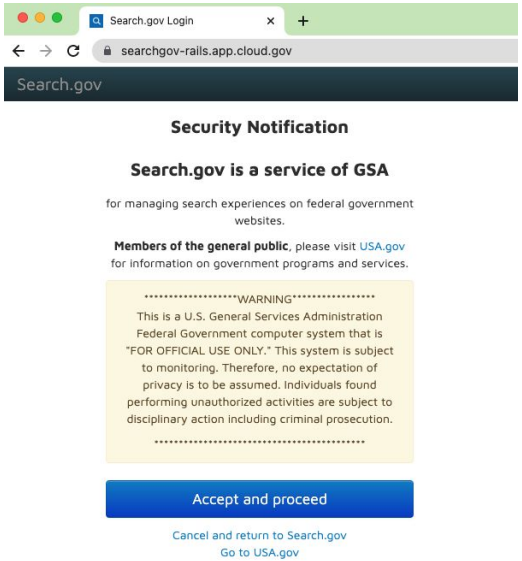
DOCKER IMAGE

- Firewall rules were created to allow Cloud.gov to connect with services (e.g., MySQL, Elasticsearch, Redis) in an existing AWS staging environment
- Addressed issues with insufficient disk space, RAM
- Success! 🎉




```
benjamincdahrooge@FCOH2J-G19HLVDF searchgov-paas % cf app searchgov-rails
Showing health and status for app searchgov-rails in org sandbox-gsa / spa

name:              searchgov-rails
requested state:   started
routes:            searchgov-rails.app.cloud.gov
last uploaded:     Thu 30 Jun 09:47:55 EDT 2022
stack:
docker image:      ghcr.io/bendahrooge/search:latest

type:           web
sidecars:
instances:      1/1
memory usage:   512M
     state     since                  cpu    memory          disk
#0   running   2022-06-30T13:48:14Z   2.6%   440.5M of 512M   1.9G of 5G
benjamincdahrooge@FCOH2J-G19HLVDF searchgov-paas %
```

App status reported by `cf app`



Search-gov running on cloud.gov hostname

CIVIC DIGITAL FELLOWSHIP

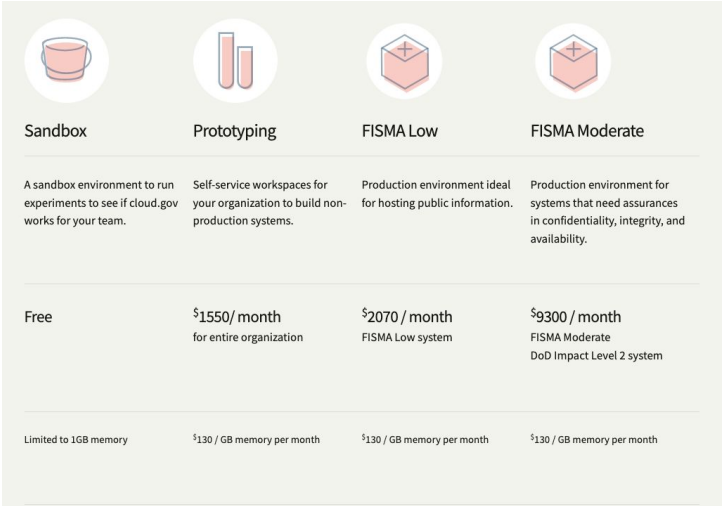
NEXT STEPS

- Test a full scale implementation using service brokers (Elasticsearch, MySQL, Redis) and microservices (i14y, asis)
    - Determine memory usage & instance performance under synthetic load
    - Verify full system functionality using “*Monster Test Plan*”
- Interagency Agreement (IAA)
- Authorization to Operate (ATO)
- If taken, **transition would take roughly one year**
    - Easy to get started building new applications, difficult to transition from complex existing infrastructure

CIVIC DIGITAL FELLOWSHIP

EXPECTED IMPACT

- **More predictable costs** in cloud billing
  - Cloud.gov: memory quota
  - AWS: a la carte resource usage (e.g., vCPU, RAM, disk space, bandwidth, IPs)
- Multi-tenant cloud environments are **more efficient**, saving the government money at scale
  - Decreased overhead
  - Similar to how Search.gov saves agencies money on enterprise search services



Sandbox	Prototyping	FISMA Low	FISMA Moderate
A sandbox environment to run experiments to see if cloud.gov works for your team.	Self-service workspaces for your organization to build non-production systems.	Production environment ideal for hosting public information.	Production environment for systems that need assurances in confidentiality, integrity, and availability.
Free	$1550/ month for entire organization	$2070 / month FISMA Low system	$9300 / month FISMA Moderate DoD Impact Level 2 system
Limited to 1GB memory	$130 / GB memory per month	$130 / GB memory per month	$130 / GB memory per month

Screenshot of Cloud.gov pricing page, August 2022

CIVIC DIGITAL FELLOWSHIP

EXPECTED IMPACT

- Portable architecture, near-instant scaling, faster code updates/deployments
- Easier Authority to Operate (ATO) process and Federal Information Security Management Act (FISMA) compliance, with Cloud.gov taking ownership of some security controls
- Loss of direct control over computing infrastructure, reliance on external party

Most public clouds provide console-level control over your virtual machines without needing physical access

CIVIC DIGITAL FELLOWSHIP

THANK YOU!

Thank you to the entire Search.gov team, especially:
Dawn McCleskey, Program Manager
Amani Farooque, Product Manager
Robert Fink, Lead Developer
James Miller, Developer
Lora Woodford, Developer
Martha Thompson, Developer
Bhavith Katpally, Developer
Dan Swick, Developer

CIVIC DIGITAL FELLOWSHIP