Scalable, Server-side Mapping in Drupal with the Geocluster-Leaflet Stack

Eric Paul (@mpgeek)

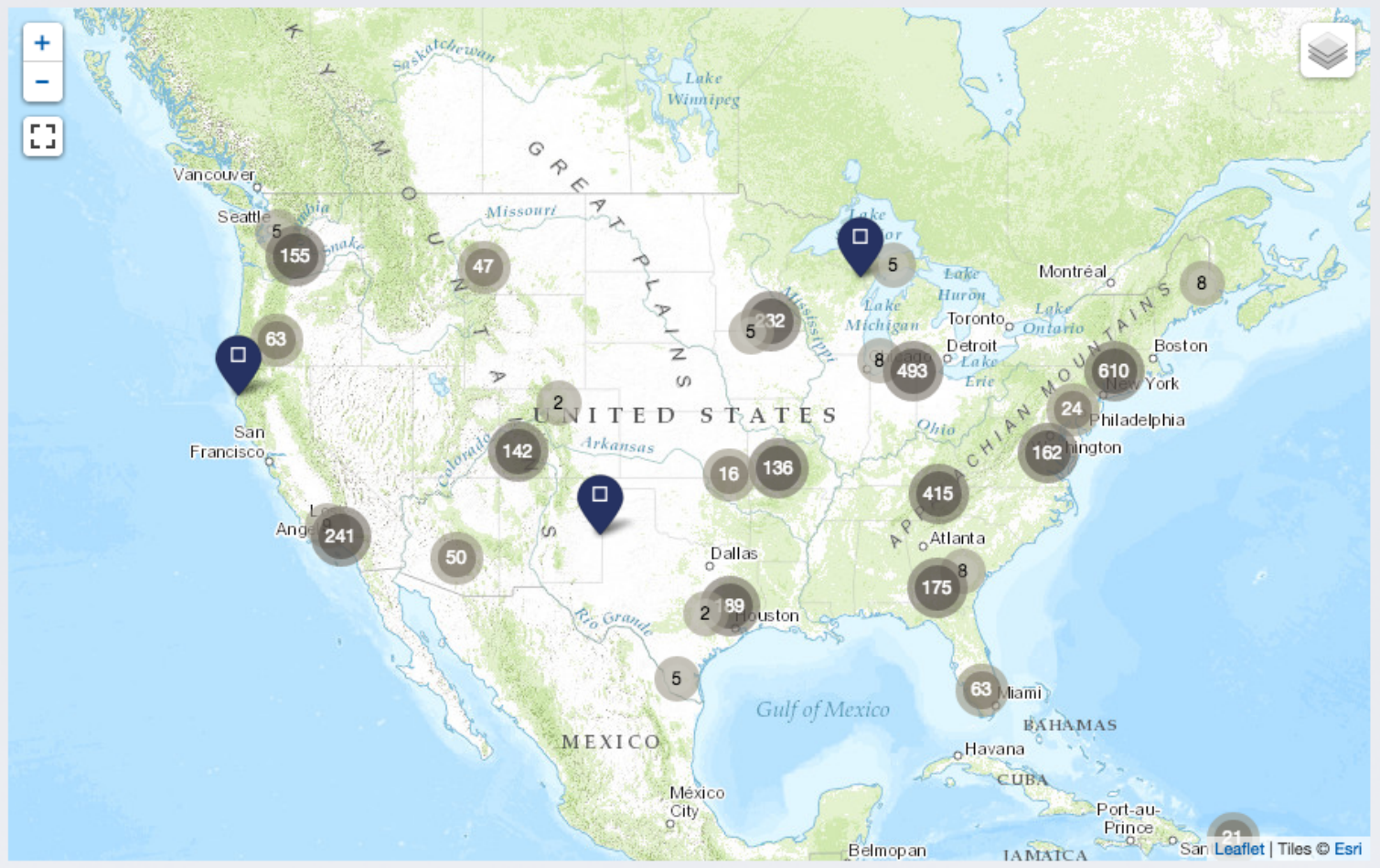
Phase2

Phase2 Technology

DEV Lunch - November 3, 2014

Demo a Thing

http://vistacampus.gov/map



The Recipe

Basic Recipe

- Address Field (location storage)
- Geocoder (geocoding addresses, requires GeoPHP)
- Geofield (geocode storage)
- **Geocluster** (server-side clustering)
- Views
- Views GeoJSON (GeoJSON feeds)
- Leaflet GeoJSON (2.x for Panels support, 1.x for Bean)
- Leaflet Integration (requires Leaflet core library)

But... we need lots of patches.

A Working Model

The client build has been released as GPL2.0

- https://github.com/mpgeek/VistaMap

Patch mania! How about a makefile?

- https://github.com/mpgeek/VistaMap/blob/master/vista_map.make

Key Architectural Feature

Geocluster Keys

- Clustering is performed at the query level by Geocluster
- PHP and JS only see the clusters as single (Views) rows.
- This feature alone is almost entirely responsible for the performance gain.

Key Architectural Feature

Geocluster Keys

- Clustering is performed at the query level by Geocluster
- PHP and JS only see the clusters as single (Views) rows.
- This feature alone is almost entirely responsible for the performance gain.

But How?

Key Architectural Feature

Geocluster Keys

- Clustering is performed at the query level by Geocluster
- PHP and JS only see the clusters as single (Views) rows.
- This feature alone is almost entirely responsible for the performance gain.

But How?

- By geohashing!

Geocluster & Geohash

In a nutshell:

- Geocluster adds a hierarchical, spatial index to geofields based on the Geohash algorithm.
- Each geofield has columns for varying levels of precision (geohash index) created/updated on `entity_save`.
- A query for points/clusters specifies a geohash index and asks for clusters based on that index.

Geocluster & Geohash

In a nutshell:

- Geocluster adds a hierarchical, spatial index to geofields based on the Geohash algorithm.
- Each geofield has columns for varying levels of precision (geohash index) created/updated on `entity_save`.
- A query for points/clusters specifies a geohash index and asks for clusters based on that index.

Notice:

- The clustering information is created when the content is created.
- A request for points and clusters doesn't actually cluster. Rather it's a simple query of a spatial index.

Near-point Clusters vs. Exact-point Clusters

Monolithic Clusters

- Leaflet doesn't discern between points that are near to one another versus multiple points at the same location.
- We needed to create two cluster types, on for each condition.

Near-point Clusters vs. Exact-point Clusters

Monolithic Clusters

- Leaflet doesn't discern between points that are near to one another versus multiple points at the same location.
- We needed to create two cluster types, on for each condition.
- `vista_map.module`, lines 115-155

On-Demand Popups

AJAX!

- We don't load the popup info into the DOM at map-load time (performance tactic).
- We needed to load them on demand and allow them to be cached.

On-Demand Popups

AJAX!

- We don't load the popup info into the DOM at map-load time (performance tactic).
- We needed to load them on demand and allow them to be cached.
- `vista_map.js`, lines 324-404

Current-user Zoom

Focus the Map on the Current-user's Location

- One of the purposes of the map was to emphasize making local connections.
- We wanted to zoom in on the currently logged-in user.

Current-user Zoom

Focus the Map on the Current-user's Location

- One of the purposes of the map was to emphasize making local connections.
- We wanted to zoom in on the currently logged-in user.
- `vista_map.module`, lines 290-351

Limit Geocoder Granularity

Geocode to Center of ZIP-code Only

- One of two data layers needed to geocode only to ZIP-code precision.
- Removing more-specific information and passing abbreviated info only to geocoder.

Limit Geocoder Granularity

Geocode to Center of ZIP-code Only

- One of two data layers needed to geocode only to ZIP-code precision.
- Removing more-specific information and passing abbreviated info only to geocoder.
- `vista_map.module`, lines 12-72

Multiple Data Layers

Implement Data Layering and Panels Support

- OG membership drove layer membership, and source geofield.
- Views necessitated that different source geofields be separate data layers.

Multiple Data Layers

Implement Data Layering and Panels Support

- OG membership drove layer membership, and source geofield.
- Views necessitated that different source geofields be separate data layers.
- Contiributed the 2.x branch of Leaflet GeoJSON for panels support with multiple data layers (https://www.drupal.org/node/2225815)

Scalability Requirement

How big did we need to go?

- Mapping user profiles, about 18k users were migrated
- Originally, it was expected that all users would be map
- Application scale, then is 10^4

Scalability Requirement

How big did we need to go?

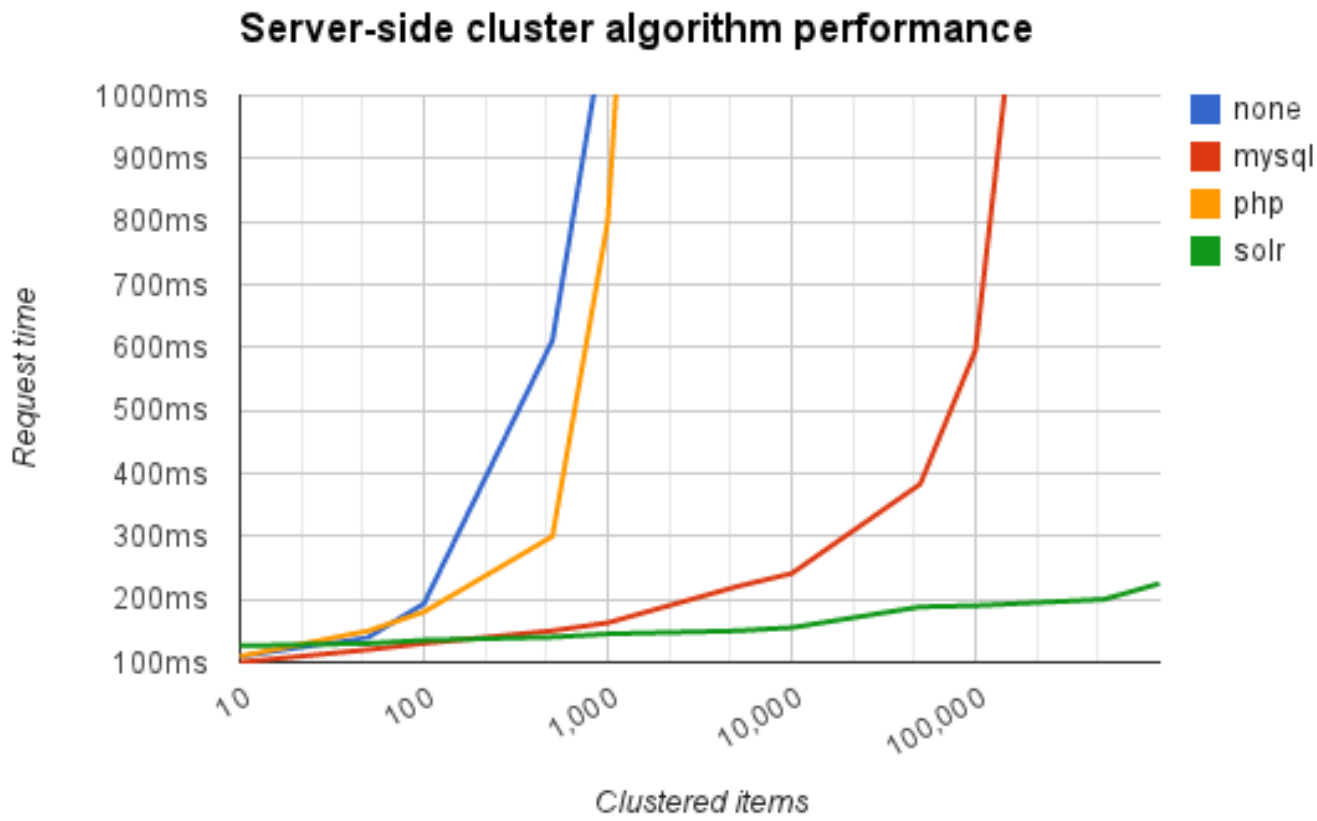
- Mapping user profiles, about 18k users were migrated
- Originally, it was expected that all users would be map
- Application scale, then is 10^4

Geocluster's clustering backend is pluggable

- PHP clustering (post-query clusternig)
- MySQL clustering (query-level clustering)
- Apache solr clustering (alternative query-level clustering)

Scalability Metrics

Cold caches



We implemented MySQL clustering

Possible Improvements

Geocluster

- Progressively enhance with client side clustering below a certain point threshold.
  https://www.drupal.org/node/1914704

Possible Improvements

Geocluster

- Progressively enhance with client side clustering below a certain point threshold.
  https://www.drupal.org/node/1914704

Possible Improvements

Leaflet GeoJSON

- Collapse clusters to a single layer to eliminate layer interference.

Possible Improvements

Leaflet GeoJSON

- Collapse clusters to a single layer to eliminate layer interference.
- Make data feeds cacheable by quantizing bounding box parameters.
  `/$view_url?bbox=$left,$right,$top,$bottom?zoom=$zoom_level`
  - The `bbox` arguments are floating-point numbers that depend on viewport size and zoom. Takes a long time for caches to warm up for non-mobile viewports.
  - https://www.drupal.org/node/1868982