

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

Eric Paul (@mpgeek)



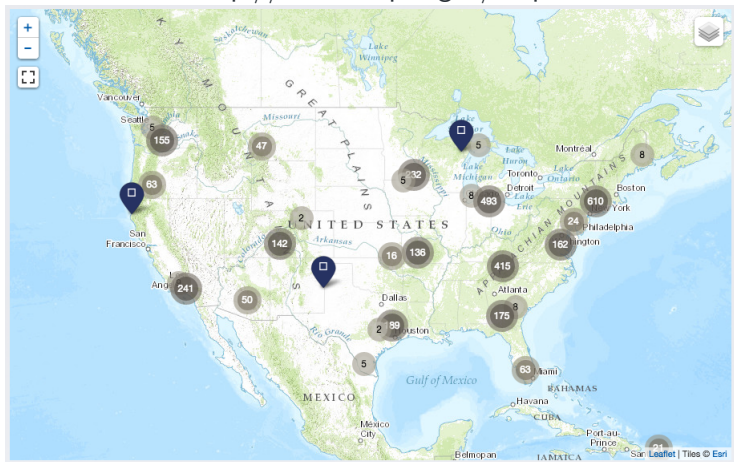
Phase2 Technology

DEV Lunch - November 3, 2014



Demo a Thing

<http://vistacampus.gov/map>



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**.



The client build was released as GPL2.0

- <https://github.com/mpgeek/VistaMap>
- Long and short-version slides + \LaTeX source is there as well

Patch mania! How about a **makefile**?

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



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.



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**!
- See Wikipedia Geohashing article for more.
<http://en.wikipedia.org/wiki/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.



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/updated**.
- A request for points and clusters **doesn't actually cluster**. Rather it's a **simple query** of a spatial index.



How big did we need to go?

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



How big did we need to go?

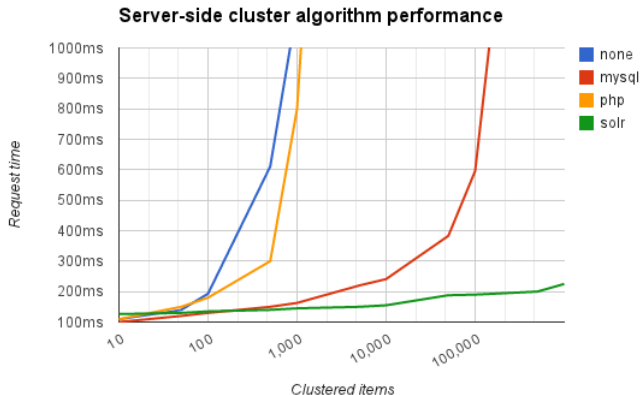
- Mapping user profiles, about 18k users were migrated
- Originally, it was expected that all users would be mapped
- Application scale, then is 10^4

Geocluster's clustering backend is **pluggable**, we get three out of the box:

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



Cold caches



We implemented **MySQL** clustering



Geocluster

- **Progressively enhance** with client side clustering below a certain point threshold.
<https://www.drupal.org/node/1914704>
- **Bounding works:** once zoomed in the cluster load is **much smaller**
- The need for server-side clustering **decreases** as zoom **increases**



Leaflet GeoJSON

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



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>



Things we saw and more resources:

- Map application in production:
<http://www.vistacampus.gov/map>
- Map application Drupal feature:
<https://github.com/mpgeek/Vista-Map>
- Geohash Algorithm:
<http://en.wikipedia.org/wiki/Geohash>
- Geocluster Master's Thesis (by @dasjo):
<http://dasjo.at/thesis>

