

Converting Physical Addresses To Geolocation

James Donn

Senior SE, Splunk

.conf2016

splunk >

A Little About Me

- 4+ Years at Splunk
- Happy Splunk Customer for 4+ years
Harvard and MITRE
- Focused on Network and Systems Management

.conf2016

splunk >

Disclaimer

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 the 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.

Agenda

- History
- Strategy
- Demo

Are Healthcare employees looking at their neighbor's data?

- EPIC data contains physical location data
 - Patients
 - Employees
- EPIC data does not contain latitude or longitude data
- Lat/Lon is need for
 - Mapping data points
 - Calculating Haversine distances
- Possibly multi-millions of locations to map

Challenges

- Where to get the Lat / Lon data?
 - What are the costs?
- How to store customer address data?
 - Indexed data cannot be edited
 - Lookup files might get too large
 - KV Store is just right!
- Epic Data is not yet in Splunk
 - Download and Splunk public school addresses sample data:
 - <http://nces.ed.gov/ccd/pubschuniv.asp>

Additional Challenges

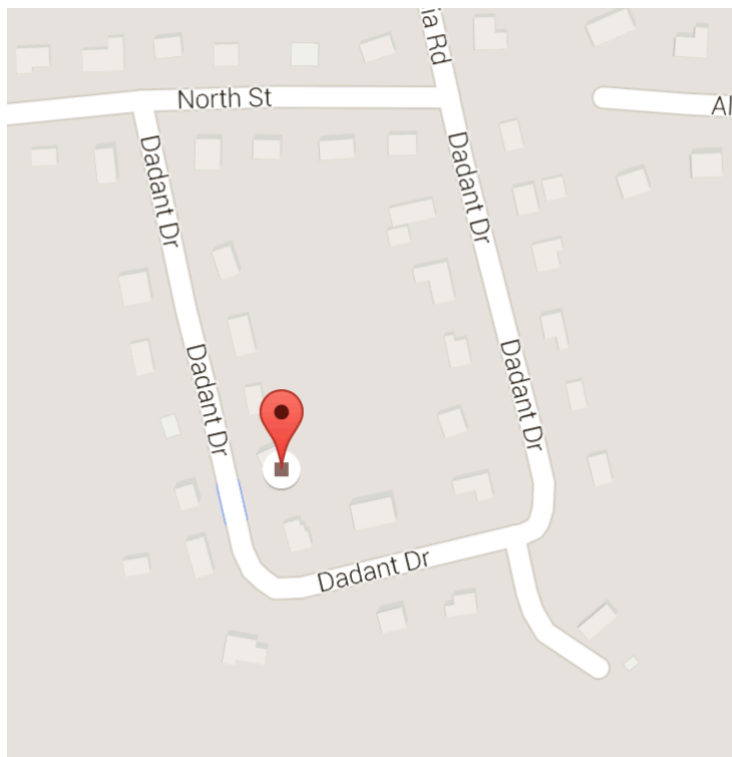
API - Query Results

- How many address are returned per query?
 - One is desired
 - Zero results
 - Multiple Results
- Limitations to API access:
 - Number of queries per second
 - Number of queries per day
 - You can buy upgrades

Epic - Data Quality

- Manually entered data is inaccurate
 - “Formatted name” of the address is not always what is expected
 - Typos
- Address Reconciliation is Required

Sometimes automation gets it **right**



Sometimes automation gets it **wrong**



Lets not repeat this **March 2016** mistake



A Google Maps error led to the
demolition of the wrong house.

“Formatted name” is not always what is expected

Events Patterns Statistics (93,844) Visualization

20 Per Page Format Preview < Prev 1 2 3 4 5

address	CUSTOMER_CITY	CUSTOMER_NUMBER	CUSTOMER_STATE	CUSTOMER_STREET
207 n san marco ave, st augustine, fl 32084	ST AUGUSTINE	207	FL	N SAN MARCO AVE
777 glades rd, boca raton, fl 33434	BOCA RATON	777	FL	GLADES RD
652 3rd st, chipley, fl 32428	CHIPLEY	652	FL	US HIGHWAY 441 N
1408 a highway 83 n, defuniak springs, fl 32433	DEFUNIAK SPRINGS	1408	FL	THIRD ST
601 institution rd, defuniak springs, fl 32433	DEFUNIAK SPRINGS	601	FL	"formatted_address" : "US-441, Okeechobee, FL, USA",
145 park st, defuniak springs, fl 32435	DEFUNIAK SPRINGS	145	FL	"formatted_address" : "652 3rd St, Chipley, FL 32428, USA",
13270 us highway 90 w, defuniak springs, fl 32435	DEFUNIAK SPRINGS	13270	FL	"formatted_address" : "1408 FL-83, Defuniak Springs, FL 32433, USA",
4019 us highway 98e, santa rosa beach, fl 32459	SANTA ROSA BEACH	4019	FL	"formatted_address" : "Defuniak Springs, FL, USA",
645 greenway trl, santa rosa beach, fl 32459	SANTA ROSA BEACH	645	FL	"formatted_address" : "145 S Park St, Defuniak Springs, FL 32435, USA",
10 smolian cir, santa rosa beach, fl 32459	SANTA ROSA BEACH	10	FL	"formatted_address" : "145 S Park St, Defuniak Springs, FL 32435, USA",
15381 us highway 331 s, freeport, il 61032	FREEPORT	15381	FL	"formatted_address" : "DeFuniak Springs Airport, 1931 US-90, Defuniak Springs, FL 32433, USA",
6694 w county highway 30a, freeport, il 61032	FREEPORT	6694	FL	"formatted_address" : "Santa Rosa Beach, FL 32459, USA",
12615 us highway 331 s, freeport, il 61032	FREEPORT	12615	FL	"formatted_address" : "645 Greenway Trail, Santa Rosa Beach, FL 32459, USA",
21893 us highway 331 n, freeport, il 61032	FREEPORT	21893	FL	"formatted_address" : "10 Smolian Cir, Santa Rosa Beach, FL 32459, USA",
164 yellow jacket ave, st marks, fl 32355	ST MARKS	164	FL	"formatted_address" : "Hwy Ave, Freeport, IL 61032, USA",
48 shell island road, daytona beach, fl 32117	DAYTONA BEACH	48	FL	"formatted_address" : "6694 W County Hwy 30A, Santa Rosa Beach, FL 32459, USA",
1025 mason ave # 1031, new Smyrna beach, fl 32168	NEW SMYRNA BEACH	1025	FL	"formatted_address" : "Hwy Ave, Freeport, IL 61032, USA",
565 s lakeview dr unit 110, lake helen, fl 32744	LAKE HELEN	565	FL	"formatted_address" : "Paxton City Park, 22036 US-331, Laurel Hill, FL 32567, USA",

```
jdonn — more — 124x21
"formatted_address" : "207 San Marco Ave, St Augustine, FL 32084, USA",
"formatted_address" : "Glades Rd, Boca Raton, FL 33434, USA",
"formatted_address" : "US-441, Okeechobee, FL, USA",
"formatted_address" : "652 3rd St, Chipley, FL 32428, USA",
"formatted_address" : "1408 FL-83, Defuniak Springs, FL 32433, USA",
"formatted_address" : "Defuniak Springs, FL, USA",
"formatted_address" : "145 S Park St, Defuniak Springs, FL 32435, USA",
"formatted_address" : "DeFuniak Springs Airport, 1931 US-90, Defuniak Springs, FL 32433, USA",
"formatted_address" : "Santa Rosa Beach, FL 32459, USA",
"formatted_address" : "645 Greenway Trail, Santa Rosa Beach, FL 32459, USA",
"formatted_address" : "10 Smolian Cir, Santa Rosa Beach, FL 32459, USA",
"formatted_address" : "Hwy Ave, Freeport, IL 61032, USA",
"formatted_address" : "6694 W County Hwy 30A, Santa Rosa Beach, FL 32459, USA",
"formatted_address" : "Hwy Ave, Freeport, IL 61032, USA",
"formatted_address" : "Paxton City Park, 22036 US-331, Laurel Hill, FL 32567, USA",
"formatted_address" : "164 Yellow Jacket Rd, Sopchoppy, FL 32358, USA",
"formatted_address" : "48 Shell Island Rd, St Marks, FL 32355, USA",
"formatted_address" : "1025 Mason Ave, Daytona Beach, FL 32117, USA",
"formatted_address" : "248 N Samsula Dr, New Smyrna Beach, FL 32168, USA",
"formatted_address" : "565 S Lakeview Dr #110, Lake Helen, FL 32744, USA",
```

Multiple entries are returned

Q New Search Save As Close

```
index="geolocation"
| dedup results{}.formatted_address
| table results{}.formatted_address results{}.geometry.location.lat results{}.geometry.location.lng
| rename results{}.formatted_address AS address, results{}.geometry.location.lat AS lat, results{}.geometry.location.lng AS lon
| eval address=lower(address)
| rex field=address mode=sed "s/, usa/g"
```

✓ 290 events (before 3/26/16 1:47:34.000 PM) Job || ■ → ↓ 📄 🗨 Verbose Mode

Events (290) Patterns Statistics (290) Visualization

20 Per Page Format Preview < Prev 1 2 3 4 5 6 7 8 9 ... Next >

address	lat	lon
301 oak st, sanford, me	43.4481693	-70.81374649999999
301 oak hill rd, sanford, me 04073	43.3751621	-70.70088029999999
301 oak branch ln, sanford, nc 27330	35.521704	-79.17826199999999
301 oakcrest dr, sanford, nc 27332	35.438318	-79.12418699999999
winter springs general services, 1126 e state rd 434, winter springs, fl 32708	28.7017378	-81.2661952
555 e state st, kennett square, pa 19348	39.8488534	-75.70583529999999
chuluota, fl 32766	28.6418949	-81.12365439999999
winter springs, fl	28.6988853	-81.30812259999999
airport rd, sanford, me 04073	43.4011565	-70.71792169999999
1722 airport rd, sanford, nc 27332	35.4409777	-79.18184309999999
sanford seacoast regional airport (sfm), 167 airport rd, sanford, me 04073	43.4010968	-70.71014889999999
840 longwood hills rd, longwood, fl 32750	28.71321009999999	-81.3615708
200 barrett bush ln, sanford, nc 27332	35.4046565	-79.2227451

Several Vendors to Choose From

- Free API access is **Limited**
- Fact – there will be a cost to populate more than 1,000 records/day
- Must decide on one API integration to start with
 - Cost effective
 - Accuracy **MUST** be at the “house” level of granularity
 - Can add others in the future

App Dev Strategy

- Two KV Stores
 - Splunk search populates the base geolocation KV store
 - Second KV Store for Reconciliation
- Populate Lat/Lon in Primary KV Store
 - For each record in geolocation KV store, alert (run script) if there is missing Lat/Lon
- Query geolocation API
 - Good results update the geolocation KV
 - Bad results go into the reconcile KV
- Leverage a record editing dashboard for reconciling records
 - Saved edits delete reconcile record and insert into geolocation KV
 - Ability to select one of the multiple addresses returned by API
 - Ability to manually assign any field value

Script Features

- Is called as an alert for missing Lat/Lon
- Queries Google API per address
- The number of results determines next steps:
 - 0 – Something is wrong, log the error message
 - 1 – Perfect, update the Geolocation KV store
 - 2 or more – Too many results:
 - Send to the Reconcile KV Store
 - Update “in_reconciliation” field in Geolocation KV store

Demo!



.conf2016

splunk >

Load Address DB into Splunk

Physical Address to Geolocation Geolocation KV Status Reconcile KV Status Reconcile Records Search Dashboards Geolocation Records

New Search Save As Close

index="addresses" All time Q

✓ 102,716 events (before 8/1/16 8:33:18.000 PM) No Event Sampling Job Verbose Mode

Events (102,716) Patterns Statistics Visualization

Format Timeline Zoom Out Zoom to Selection Deselect 1 month per column

List Format 20 Per Page Prev 1 2 3 4 5 6 7 8 9 Next

Hide Fields		All Fields		i	Time	Event
Selected Fields				>	3/24/16 10:05:21.000 AM	2014 120201605414 12 1202016 05414 68 9001 DEAF/BLIND FLORIDA SCHOOL FOR THE DEAF AND BLIND SUPERINTEDE 904827 2210 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 2 1 000 2 N N host = LapChopper.home source = Sch14pre.txt sourcetype = tsv
Interesting Fields				>	3/24/16 10:05:21.000 AM	2014 120201601323 12 1202016 01323 68 0019 DEAF/BLIND CAREER EDUCATION AND TRANSITION HIGH SCHOOL 9048272700 1 000 2 09 12 ST AUGUSTINE FL 32084 2762 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 2 host = LapChopper.home source = Sch14pre.txt sourcetype = tsv
				>	3/24/16 10:05:21.000 AM	2014 120201601265 12 1202016 01265 68 0016 DEAF/BLIND BLIND HIGH SCHOOL (FSDB) 9048272700 207 SAN MARCO AVE ST AUGUSTINE FL 32084 2762 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 2 1 09 12 host = LapChopper.home source = Sch14pre.txt sourcetype = tsv
				>	3/24/16 10:05:21.000 AM	2014 120201601257 12 1202016 01257 68 0015 DEAF/BLIND BLIND MIDDLE SCHOOL (FSDB) 9048272700 207 SAN MARCO AVE ST AUGUSTINE FL 32084 2762 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 2 1 06 08 host = LapChopper.home source = Sch14pre.txt sourcetype = tsv
				>	3/24/16 10:05:21.000 AM	2014 120201601248 12 1202016 01248 68 0014 DEAF/BLIND BLIND ELEMENTARY SCHOOL (FSDB) 9048272700 207 SAN MARCO AVE ST AUGUSTINE FL 32084 2762 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 2 1 PK 05 host = LapChopper.home source = Sch14pre.txt sourcetype = tsv
				>	3/24/16 10:05:21.000 AM	2014 120201601247 12 1202016 01247 68 0013 DEAF/BLIND DEAF HIGH SCHOOL (FSDB) 9048272500 207 SAN MARCO AVE ST AUGUSTINE FL 32084 2762 207 N SAN MARCO AVE ST AUGUSTINE FL 32084 2762 2 1 000 2

Alert Drives Script

Update Geolocation

[Searches, reports, and alerts](#) » Update Geolocation

Search

```
| inputlookup geolocation_lookup  
| eval key_id = _key  
| search (cust_lat=DNE OR cust_lon=DNE OR  
location_type=DNE) in_reconciliation=no
```

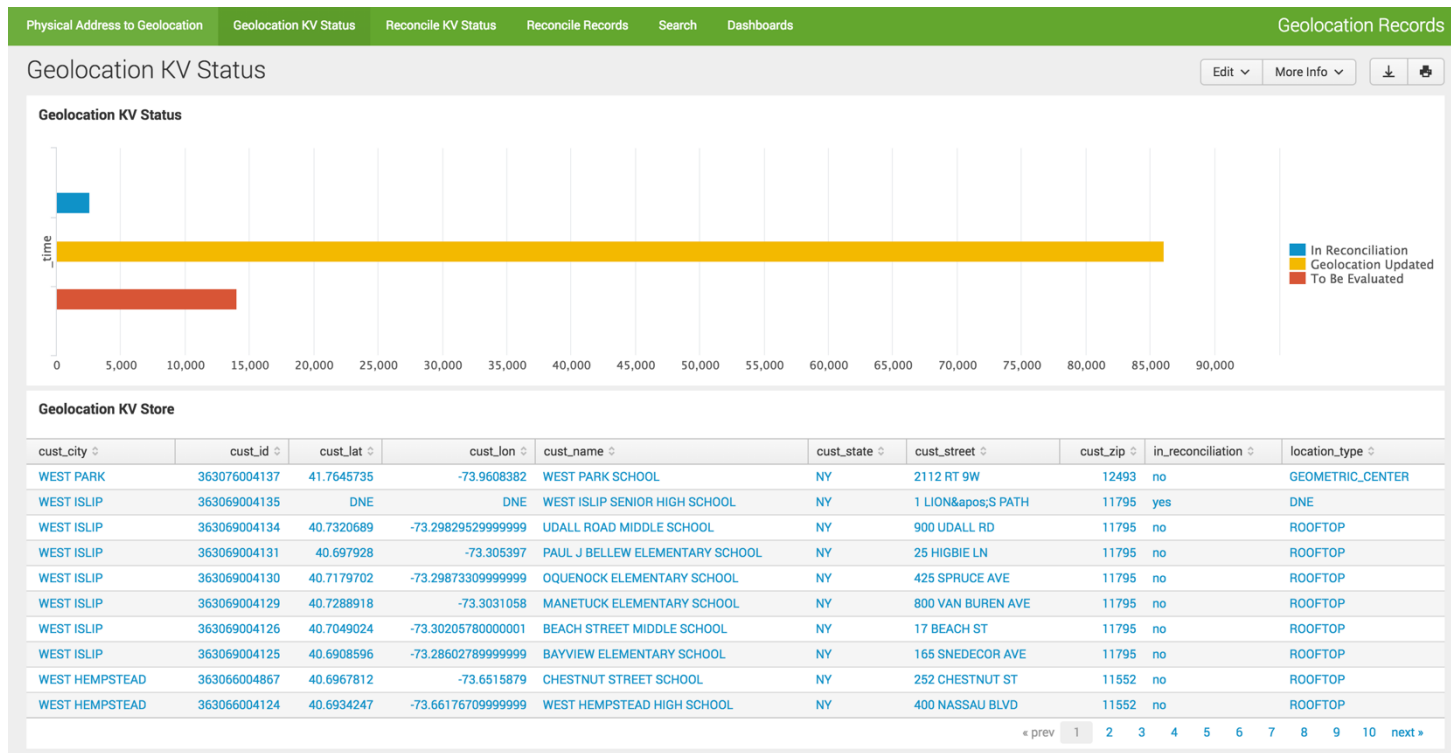
Description

Run as

Owner User

[Learn More](#)

Geolocation KV Status



Reconcile KV Status

Physical Address to Geolocation Geolocation KV Status **Reconcile KV Status** Reconcile Records Search Dashboards Geolocation Records

Reconcile KV Status Edit ▾ More Info ▾ ⬇️ 🖨️

Reconcile KV Status

Number of Addresses per Unique Customer ▾	count ▾
2	1274
3	404
4	284
10	163
6	123
5	91
9	82
8	63
7	50
12	12

◀ prev 1 2 next ▶

Reconcile KV Store

cust_id ▾	cust_lat ▾	cust_lon ▾	cust_name ▾	formatted_address ▾	kv_key ▾	location_type ▾
362511003473	40.5916189	-73.8196686	ROXBURY CENTRAL SCHOOL	New York State Reference Rte 907J, Queens, NY, USA	572b9ae0c8238f672e73a8bd	GEOMETRIC_CENTER
362511003473	40.5736968	-73.8851447	ROXBURY CENTRAL SCHOOL	New York State Reference Rte 901B, New York, NY, USA	572b9ae0c8238f672e73a8bd	GEOMETRIC_CENTER
363069004135	22.3470153	114.1810994	WEST ISLIP SENIOR HIGH SCHOOL	1, Hong Kong	572b9bd8c8238f72d44ab842	GEOMETRIC_CENTER
363069004135	44.6580694	0.0466607	WEST ISLIP SENIOR HIGH SCHOOL	1 Lion, 33580 Le Puy, France	572b9bd8c8238f72d44ab842	RANGE_INTERPOLATED
362511003473	40.5916189	-73.8196686	ROXBURY CENTRAL SCHOOL	New York State Reference Rte 907J, Queens, NY, USA	572b9bd8c8238f72d44ab95d	GEOMETRIC_CENTER
362511003473	40.5736968	-73.8851447	ROXBURY CENTRAL SCHOOL	New York State Reference Rte 901B, New York, NY, USA	572b9bd8c8238f72d44ab95d	GEOMETRIC_CENTER
362511003473	40.5736968	-73.8851447	ROXBURY CENTRAL SCHOOL	New York State Reference Rte 901B, New York, NY, USA	572b9bd8c8238f72d44ab95d	GEOMETRIC_CENTER
362442003339	40.8132669	-72.6942792999999	REMSENBURG-SPEONK ELEMENTARY SCHOOL	11 Mill Rd, Remsenburg-Speonk, NY, USA	572b9bd8c8238f72d44ab9cc	RANGE_INTERPOLATED
362442003339	40.812447	-72.6963619	REMSENBURG-SPEONK ELEMENTARY SCHOOL	11 Old Mill Ln, Remsenburg, NY 11960, USA	572b9bd8c8238f72d44ab9cc	ROOFTOP
362415003324	42.474747	-73.7931301999999	PIETER B COEYMANS SCHOOL	Church St, Coeymans, NY 12045, USA	572b9bd8c8238f72d44ab9db	GEOMETRIC_CENTER

Reconcile Records

Physical Address to Geolocation Geolocation KV Status Reconcile KV Status **Reconcile Records** Search Dashboards Geolocation Records

Reconcile Records Edit ▾ More Info ▾ ⬇️ 🖨️

Filter Down to Customer ID and Then Select the Correct Record

Customer ID: Customer Name:

reconcile_key ▾	kv_key ▾	cust_id ▾	cust_name ▾	formatted_address ▾	cust_lat ▾	cust_lon ▾	location_type ▾
579c14cdc8238f64b8361186	572b9bddc8238f72d44c0826	550216000260	EDEN ELEMENTARY	N Elizabeth St, Eden, WI 53019, USA	43.6947874	-88.36065959999999	GEOMETRIC_CENTER
579c14cdc8238f64b8361187	572b9bddc8238f72d44c0826	550216000260	EDEN ELEMENTARY	210 S Elizabeth St, Eden, WI 53019, USA	43.6927768	-88.36063899999999	RANGE_INTERPOLATED
579cc812c8238f64b8361188	572b9bddc8238f72d44c08ae	550087000113	ALMENA ELEMENTARY	230 Lulu Ave E, Almena, WI 54805, USA	45.41649	-92.033497	ROOFTOP
579cc812c8238f64b8361189	572b9bddc8238f72d44c08ae	550087000113	ALMENA ELEMENTARY	230 Lulu Ave W, Almena, WI 54805, USA	45.4168471	-92.0426705	RANGE_INTERPOLATED
579cc81bc8238f64b836118a	572b9bddc8238f72d44c08c2	550063002586	WILDLANDS CHARTER SCHOOL	E12088 Co Rd K, Augusta, WI 54722, USA	44.6653775	-91.2767904	GEOMETRIC_CENTER
579cc81cc8238f64b836118b	572b9bddc8238f72d44c08c2	550063002586	WILDLANDS CHARTER SCHOOL	County Hwy N, Ludington, WI, USA	44.7857536	-91.1857102	GEOMETRIC_CENTER
579cd9b0c8238f64b836118c	572b9bddc8238f72d44c0958	550003602080	WISCONSIN RESOURCE CENTER	1505 North Dr, Oshkosh, WI 54901, USA	44.0770179	-88.5165806	RANGE_INTERPOLATED
579cd9b0c8238f64b836118d	572b9bddc8238f72d44c0958	550003602080	WISCONSIN RESOURCE CENTER	1505 North St, Neenah, WI 54956, USA	44.19401999999999	-88.49785869999999	RANGE_INTERPOLATED
579cd9b2c8238f64b836118e	572b9bddc8238f72d44c095b	550003503005	GROW ACADEMY	4986 County Hwy A, Oregon, WI 53575, USA	42.8925908	-89.38509309999999	RANGE_INTERPOLATED
579cd9b2c8238f64b836118f	572b9bddc8238f72d44c095b	550003503005	GROW ACADEMY	4986 County Hwy M, Fitchburg, WI 53575, USA	42.9502588	-89.38605799999999	RANGE_INTERPOLATED

« prev 1 2 3 4 5 6 7 8 9 10 next »

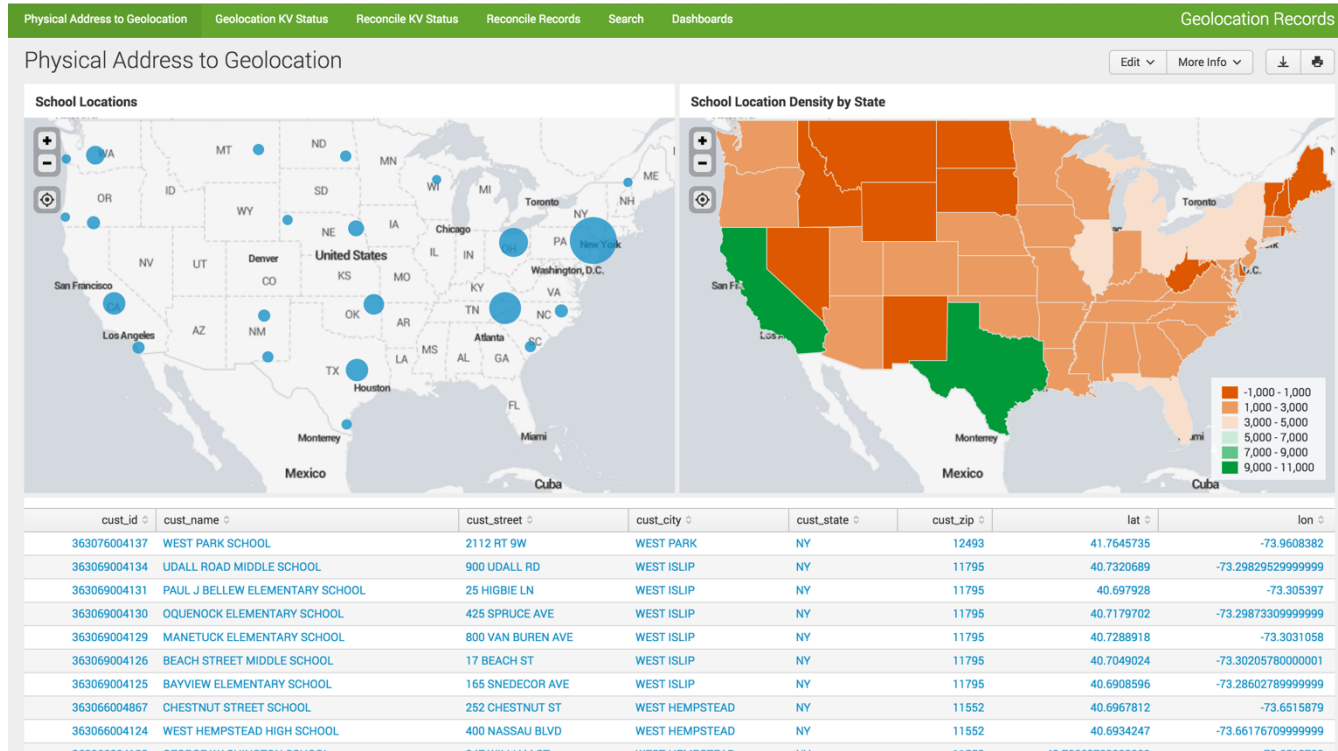
Click on the Record Again to Confirm Update

reconcile_key: 579cc812c8238f64b8361188

reconcile_key ▾	kv_key ▾	cust_id ▾	cust_name ▾	formatted_address ▾	cust_lat ▾	cust_lon ▾	location_type ▾
579cc812c8238f64b8361188	572b9bddc8238f72d44c08ae	550087000113	ALMENA ELEMENTARY	230 Lulu Ave E, Almena, WI 54805, USA	45.41649	-92.033497	ROOFTOP

About Support File a Bug Documentation Privacy Policy © 2005-2016 Splunk Inc. All rights reserved.

Physical Address to Geolocation



What Now?

- Geolocation KV with Lat/Lon drive automated lookups for Haversine calculations
- POC Complete!
- App can be reused for similar use cases
- App can be used to become more familiar with KV stores

THANK YOU

jim@splunk.com

.conf2016