What's Next in Geo For Splunk

FN1735

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans 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 may differ materially. 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, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.

Geoffrey Hendrey
Sr. Principal Engineer

Aditi Nath
Software Development Engineer

splunk> .conf19

Presentation Outline

The value of location

Choropleth Maps in the Dashboard Framework

- The current evolution of geo
- Design overview
- Demo

Geo ‘next’

- Using H3 to spatially index data
- H3 SPL queries for tracking and heat maps

Design Spike

- Combined time and space in responsive UI
- Demonstration on mass transit data

The Rising Importance of Location

splunk> .conf19

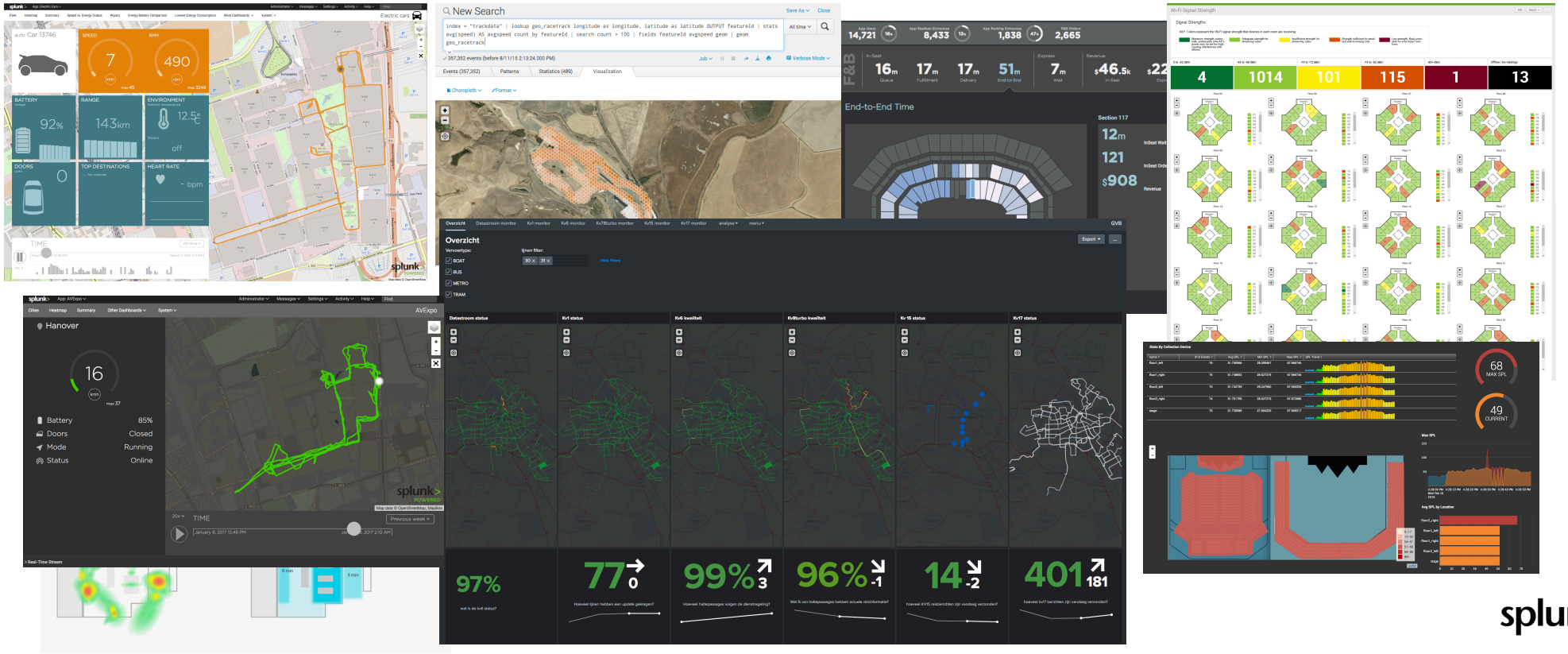
"**Location** will increasingly become one of the most important data types…one could argue that almost all data has a location element to it…"

—Gartner Survey Analysis, 2017

Huge Customer Demand

Elaborate Geo-based solutions built on Splunk

Costly to develop, suggesting value to the customer is high



Technology and Market

Heavily invested in Geo Analytics

Companies like Uber and Lyft have an obvious interest in Geo-analytics

Invested heavily in tools, libraries, and technologies/algorithms

- Kepler
- Deck.gl
- H3

Splunk's competitors have made significant improvements to their geospatial stacks

- Elastic Search
  - Vector maps
  - Geo bounding queries
  - Geo indexing and Geohashing

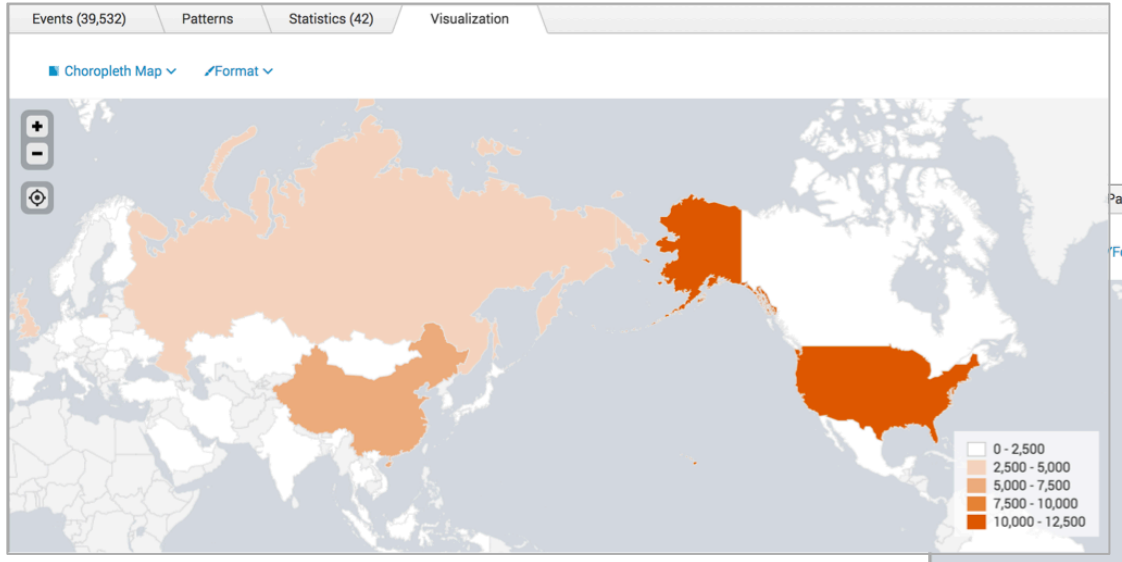
New Components For Geo in the Dashboard Framework

Dashboard Geo Capabilities

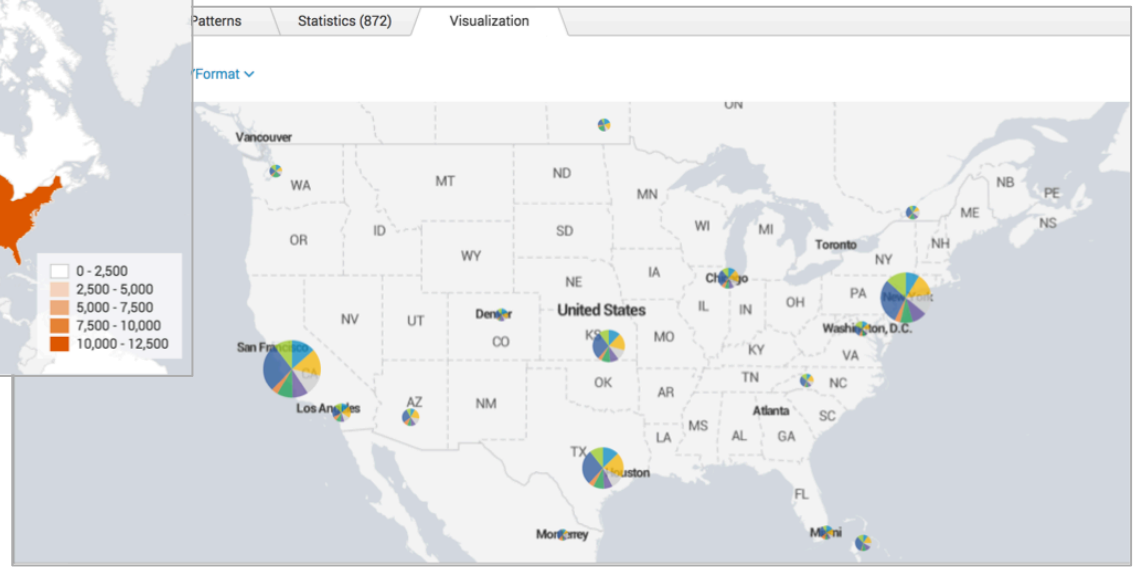
splunk> .conf19

Existing Splunk Enterprise Geo Viz

Choropleth Maps



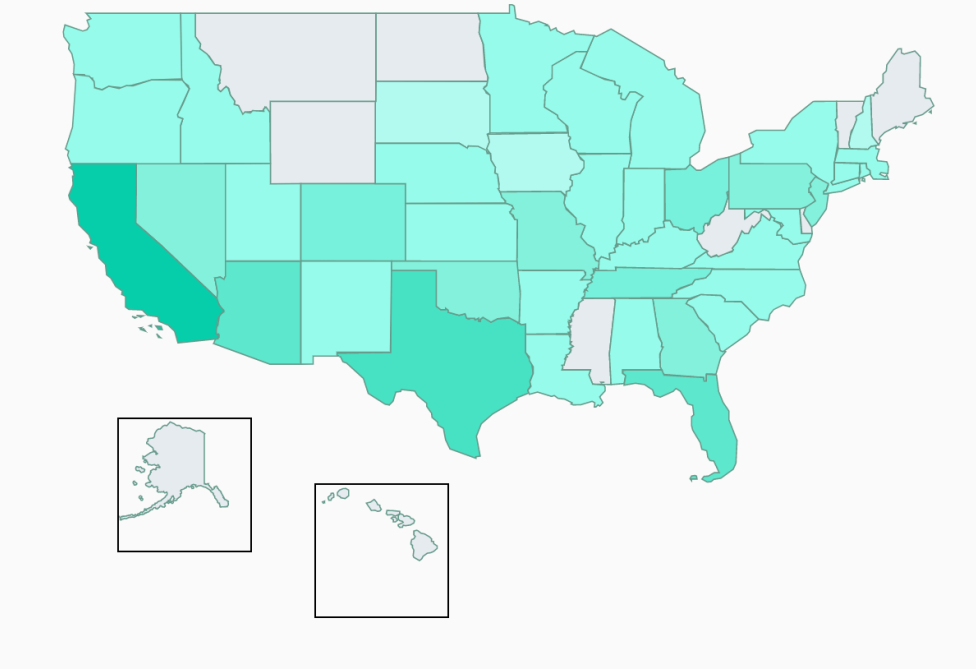
Pie Maps

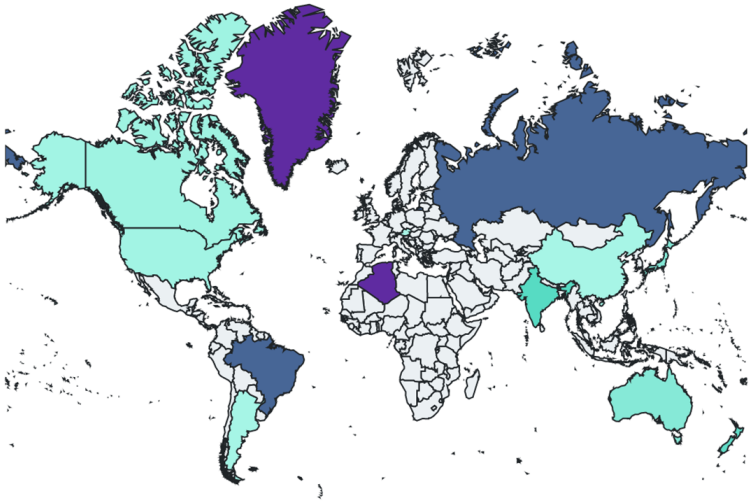


Laying New Foundations

New dashboard components

Introduced Choropleth Dashboard components for geo map in the Dashboard framework



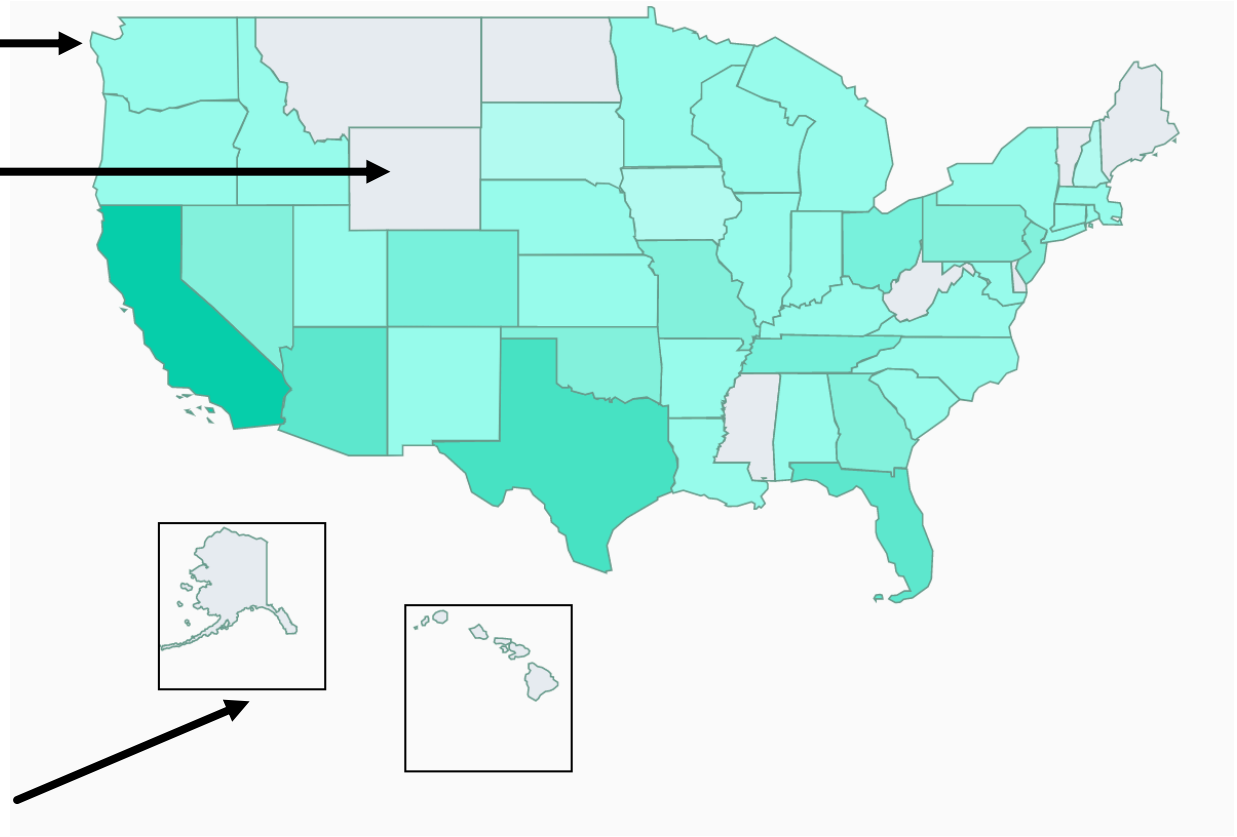


GeoJson and SVG

Vector maps

Loads Map From
Any GeoJson File

Improved and simplified
color gradients



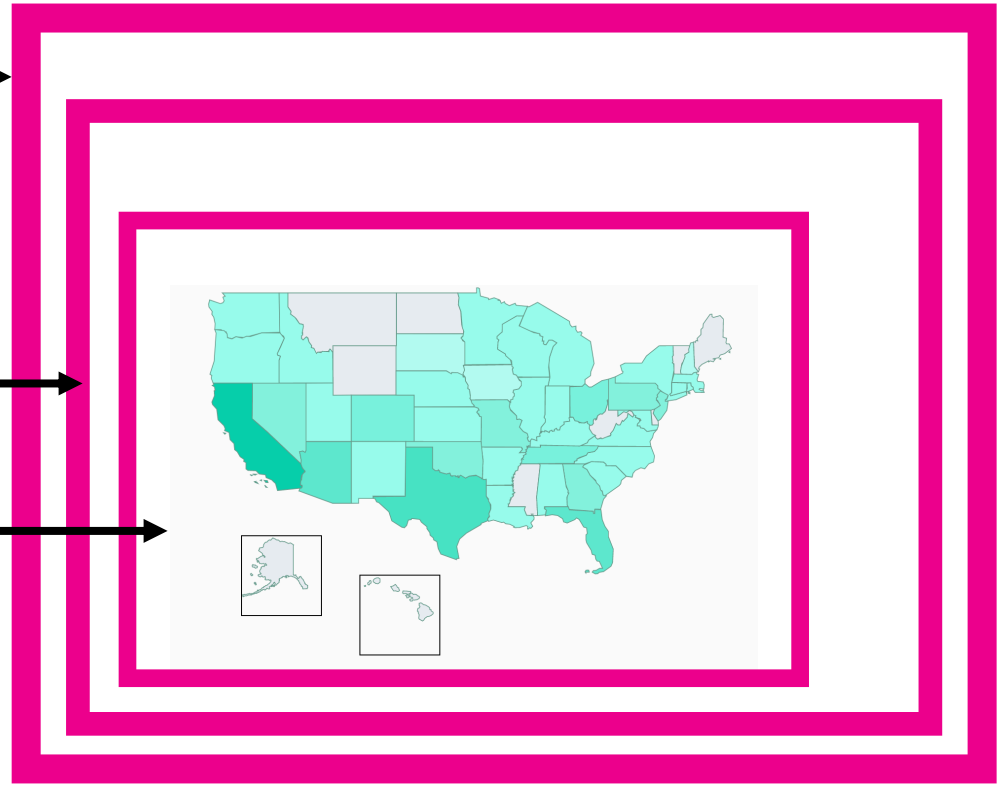
Customizable “inset maps”

Nested Component Design

Dashboard Choropleth
(encodes data values as colors)

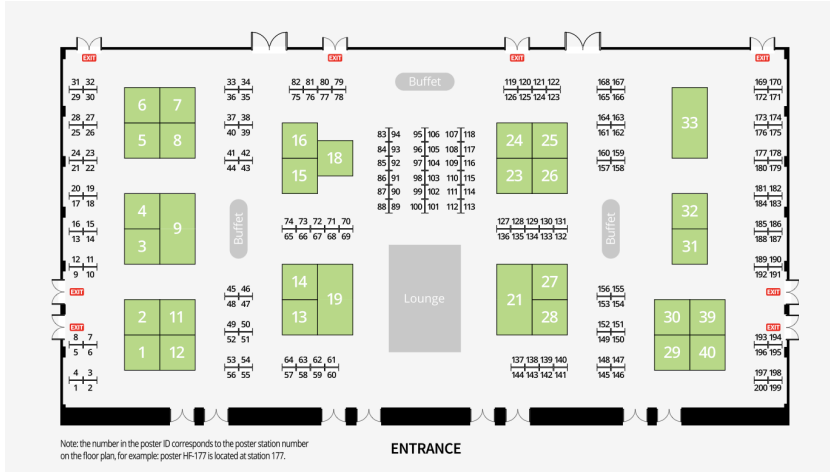
GeoJsonChoropleth
(converts GeoJson to SVG)

SvgChoropleth
(React Visualization)

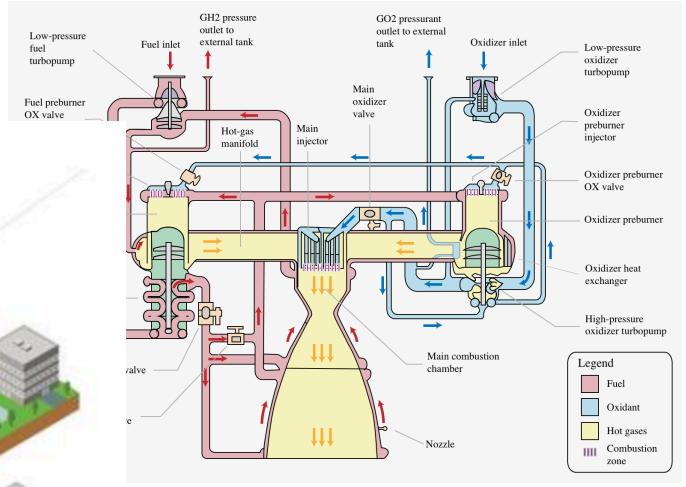


SVG Shape Dashboard Ideas

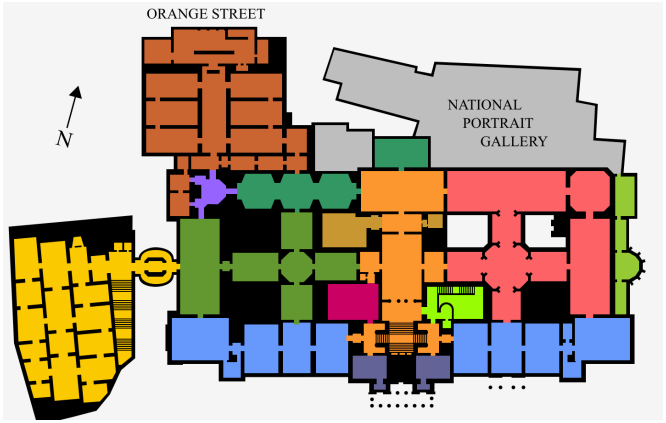
Venue



Process/device



Floorplans



Isometric

Dashboard Framework
Choropleth Code and Config

Demo

Where are We Aiming?

Building a platform for location data

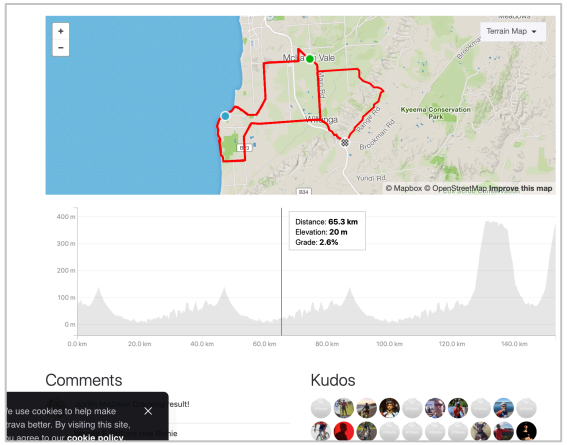
splunk> .conf19

Categorizing ‘next’ GeoSpatial Use Cases

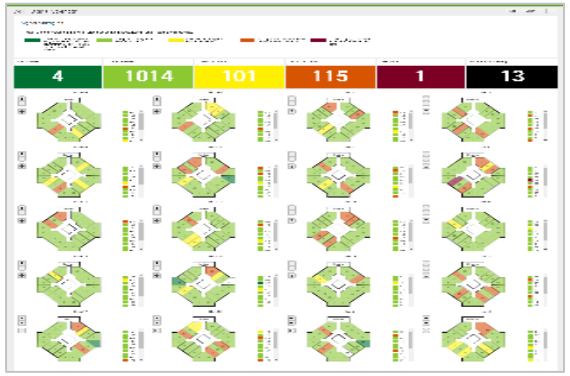
…based on customer conversations

Heat Maps

Moving Point Tracking



Thematic Choropleth



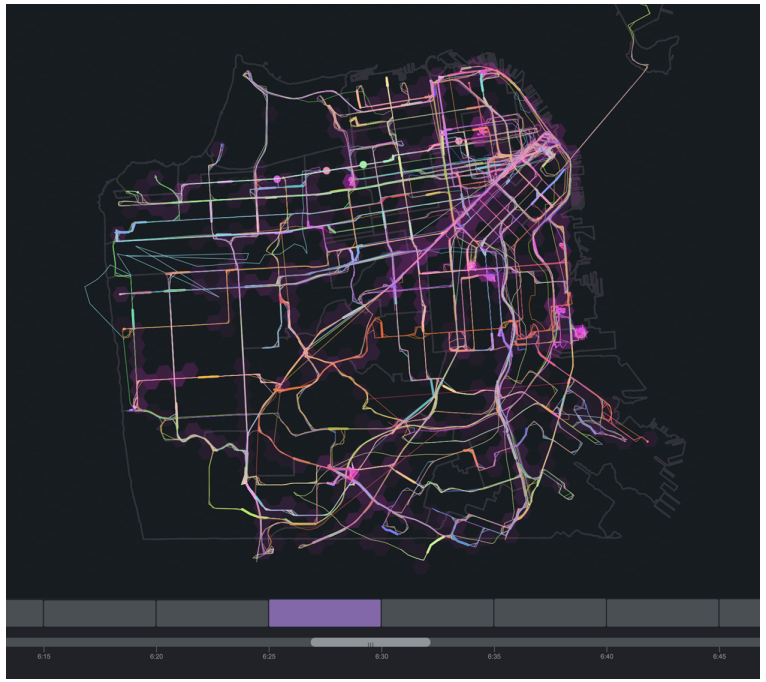
Transit Analytics

splunk> .conf19

Takeaways

Common themes expressed by customers

1. Desire to visualize spatial data over time
2. Desire to investigate, explore, and drill down without SPL
3. Desire for performance
4. Many use cases for moving point data



Demo

Location Data is Tricky

Spatial Data presents unique challenges

Managing temporal event streams is Splunk's bread and butter

But what happens when we combine spatial and temporal?

- 5,000 vehicles
- Nationwide coverage (large spatial extent)
- 10 second location reporting interval

The data-ingest volume itself is not typically an issue.

Query latency is the issue. We need:

- Highly responsive, interactive
- Compute intensive 'transaction' command to mine out vehicle 'tracks'
- Aggregations to make 'heat maps'

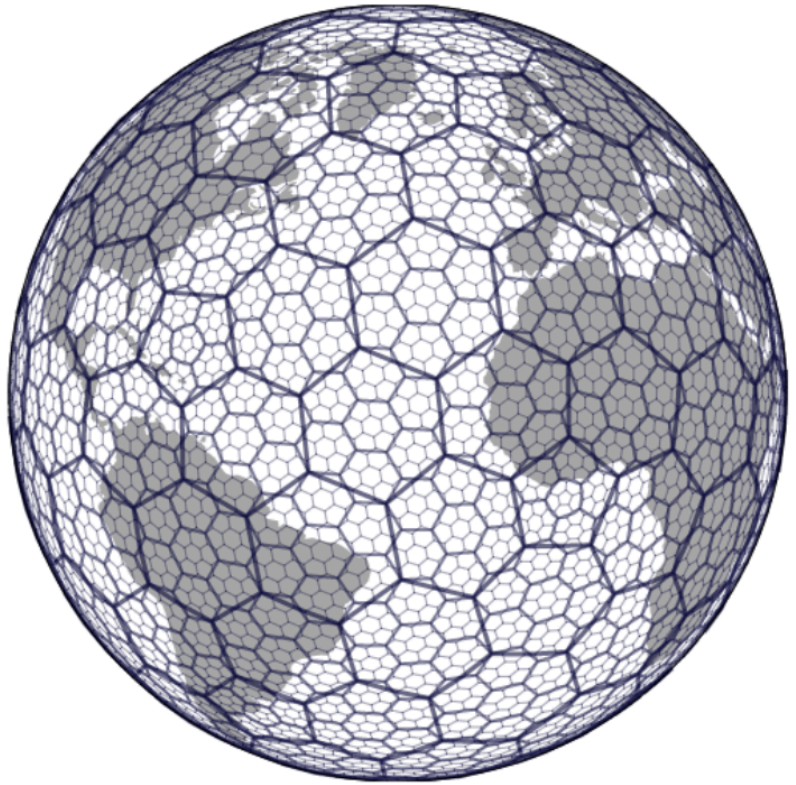
Native H3 Support in Splunk

How Spatial Addressing helps us manage Location Data

“H3 enables users to partition the globe into hexagons for more accurate analysis.”

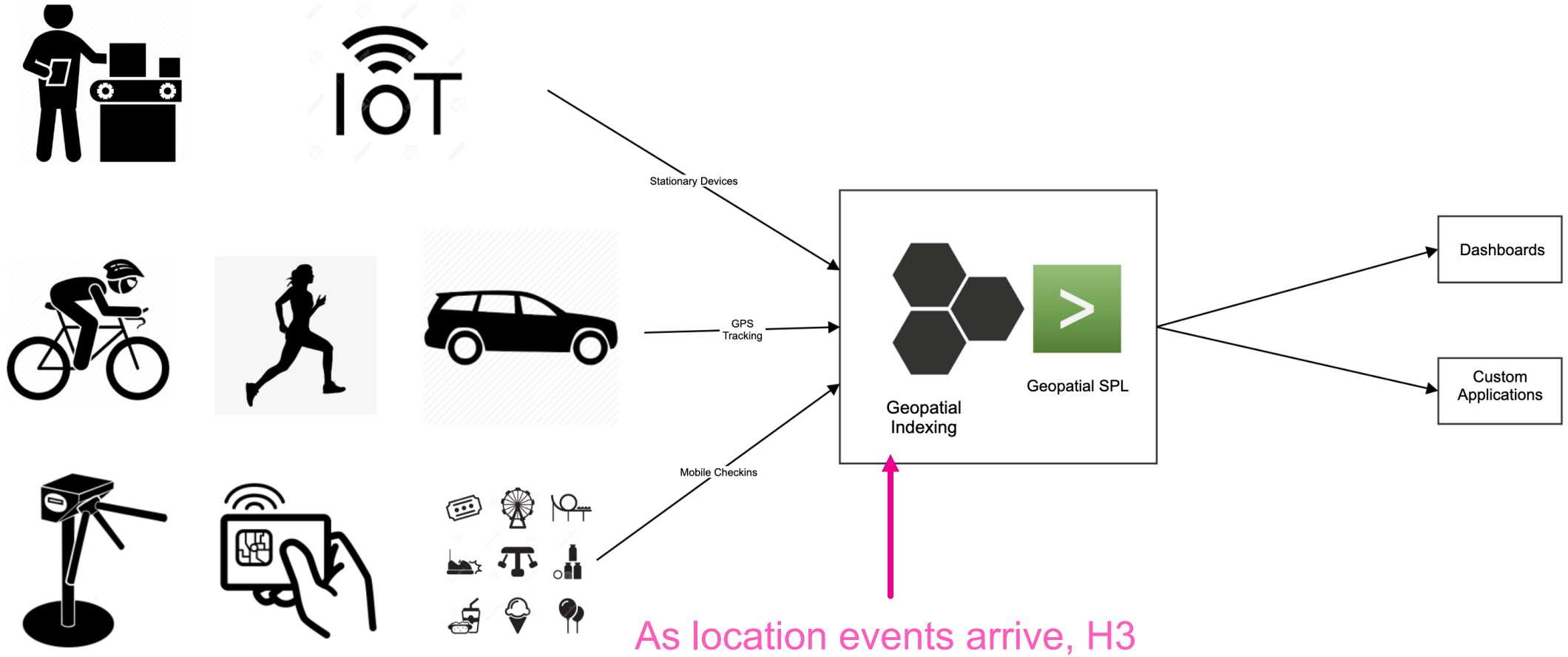
A hierarchy of hexagons

Every location-event falls inside exactly 16 hexagons (one for each ‘zoom level’)



Location Platform

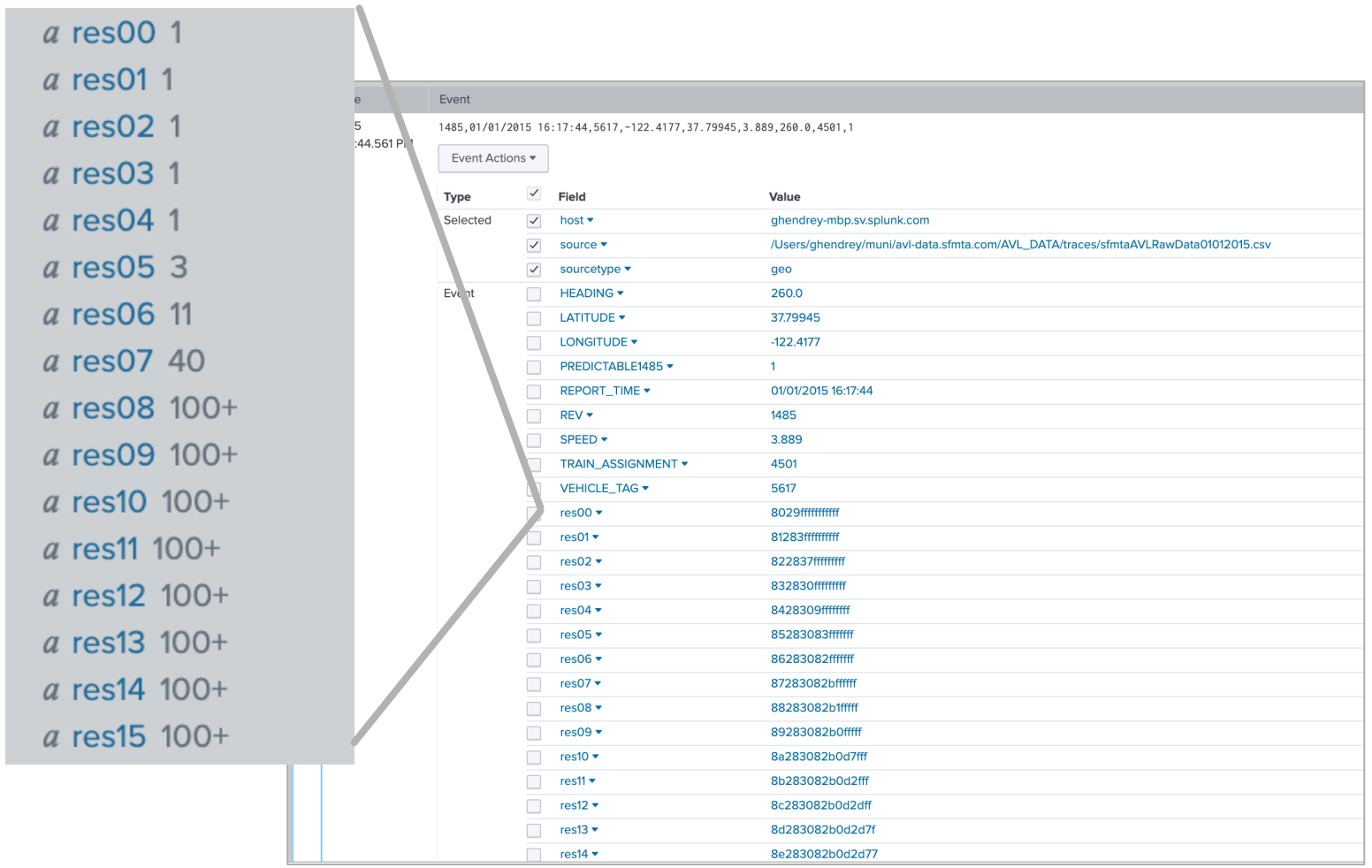
Spatial Indexing and Query



As location events arrive, H3 resolution keys are added

An H3-augmented Location Event

Fields res00-res15 are added automatically



- res00 represents a face of a large icosahedron surrounding entire earth
- res15 represents a very high resolution hexagon covering a small area
- Note in the ‘fields’ counts that the higher the res (resolution), the more unique values for that res

splunk> .conf19

H3 Ingest-Eval

How to spatially index events as they arrive

Enterprise Ingest-eval https://docs.splunk.com/Documentation/Splunk/7.3.1/Data/IngestEval

Available Splunk Enterprise 8.01 Dec 2019

Uses new geohex eval function

props.conf

```
[geo]
TRANSFORMS = h3
DATETIME_CONFIG =
INDEXED_EXTRACTIONS = csv
LINE_BREAKER = ([\r\n]+)
NO_BINARY_CHECK = true
category = Custom
description = ingest eval h3 resNN fields
disabled = false
```

Fields.conf

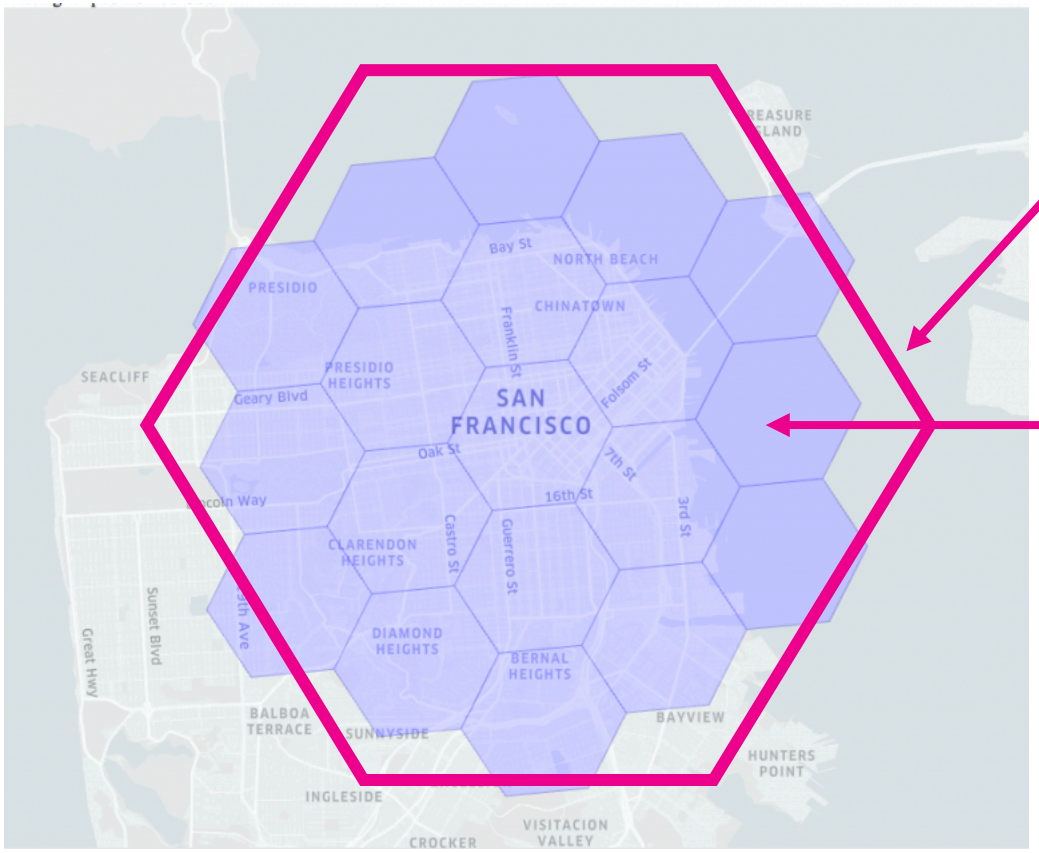
```
[res00]
INDEXED = True
[res01]
INDEXED = True
[res02]
INDEXED = True
[res03]
INDEXED = True
[res04]
INDEXED = True
[res05]
INDEXED = True
[res06]
…
```

Transforms.conf

```
[h3]
INGEST_EVAL = res00=geohex(LATITUDE, LONGITUDE,
0), res01=geohex(LATITUDE, LONGITUDE, 1),
res02=geohex(LATITUDE, LONGITUDE, 2),
res03=geohex(LATITUDE, LONGITUDE, 3),
res04=geohex(LATITUDE, LONGITUDE, 4),
res05=geohex(LATITUDE, LONGITUDE, 5),
res06=geohex(LATITUDE, LONGITUDE, 6),
res07=geohex(LATITUDE, LONGITUDE, 7),
res08=geohex(LATITUDE, LONGITUDE, 8),
res09=geohex(LATITUDE,LONGITUDE, 9),
res10=geohex(LATITUDE, LONGITUDE, 10),
res11=geohex(LATITUDE, LONGITUDE, 11),
res12=geohex(LATITUDE, LONGITUDE, 12),
res13=geohex(LATITUDE, LONGITUDE, 13),
res14=geohex(LATITUDE,LONGITUDE, 14),
res15=geohex(LATITUDE, LONGITUDE, 15)
```

H3 Query Strategies

Accelerating Heat Map queries with spatial selectivity



Spatially Selective Bounding Hexagon

Heat Map “count By” hexagons

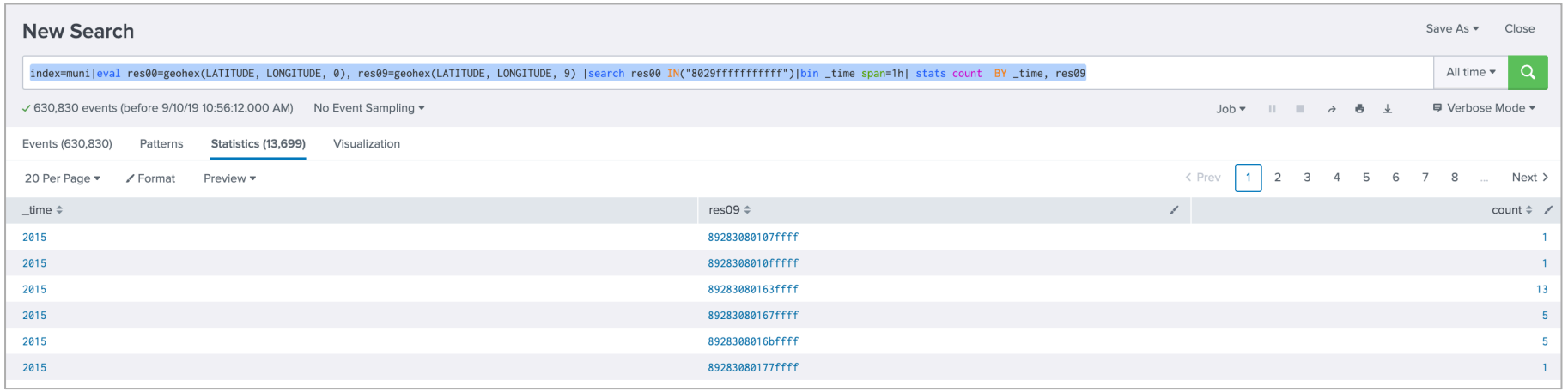
splunk> .conf19

Time Binned Heat Map Queries with H3

A Non-indexed query:

- `index=muni|eval res00=geohex(LATITUDE, LONGITUDE, 0), res09=geohex(LATITUDE, LONGITUDE, 9) |search res00 IN("8029ffffffffffff")|bin _time span=1h| stats count BY _time, res09`

"This search has completed and has returned **13,699** results by scanning **630,830** events in **80.552** seconds"



splunk> .conf19

Comparison: H3 Indexes vs. Non-indexed

Indexes are the key to performance

Time-evolving hexagon heat map query

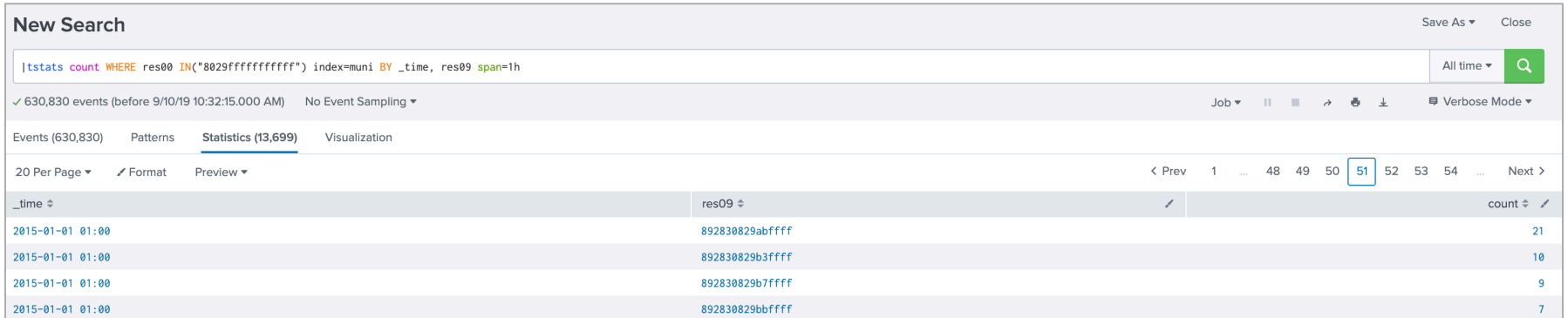
```
|tstats count WHERE res00 IN("8029ffffffffff") index=muni BY _time, res09 span=1h
```

Count location events by hexagon at any resolution (zoom level)

"This search has completed and has returned **13,699** results by scanning **630,830** events in ***0.337 seconds***"

80.552 seconds without H3 indexes vs 0.337 seconds with H3 indexes

239X performance win for H3 indexing



splunk> .conf19

Performance Analysis (I)

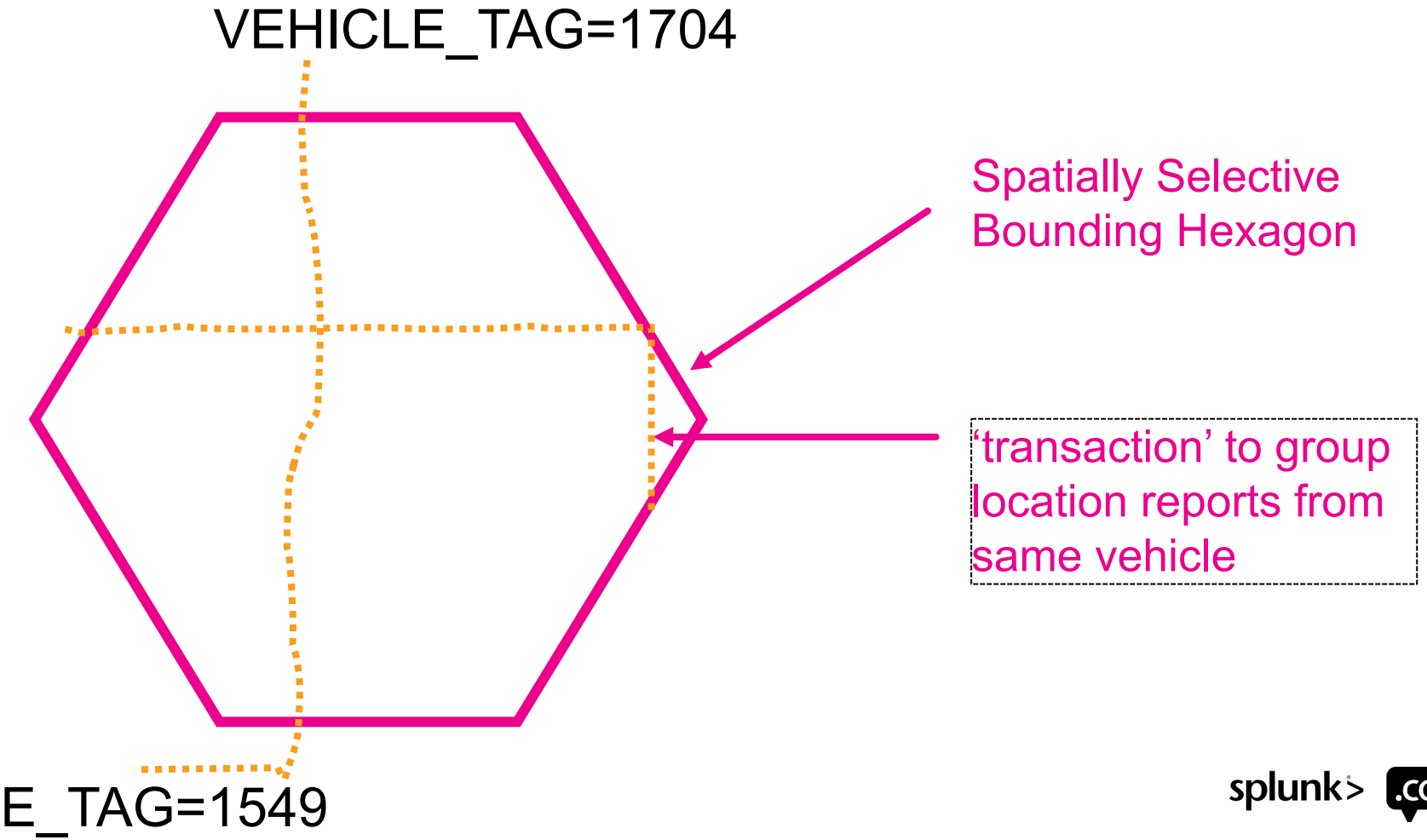
Why is tstats 239x faster with H3 indexes?

Tstats in general is so much faster than stats

- The “trick” with H3 is to turn spatial areas into keywords that are indexed in tsidx
- When events arrive, ingest-eval is simply inserting keywords into the location events
- Splunk is very good at searching and performing stats on keywords
- https://answers.splunk.com/answers/186938/what-is-tstats-and-why-is-so-much-faster-than-stat.html

H3 Query Strategies for Geo Tracking

Accelerating tracking queries with spatial selectivity



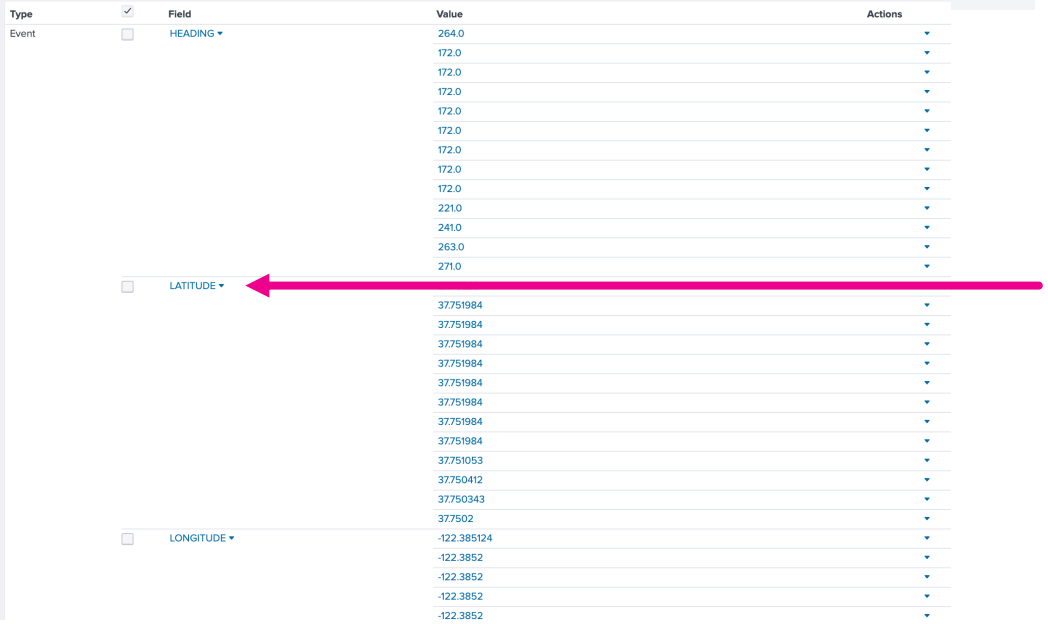
Location Tracking Queries with H3

Using H3 with transaction to collect ‘tracks’

- Get Tracks in desired area

```
index=muni res05=85283083fffffff sourcetype=geo
earliest="01/01/2015:12:00:00" latest="01/01/2015:13:00:00"|fields - _raw,
_res* |fields + SPEED,HEADING,LATITUDE,LONGITUDE, REPORT_TIME, VEHICLE_TAG
| transaction VEHICLE_TAG mvlist="SPEED,HEADING,LATITUDE,LONGITUDE,
REPORT_TIME" maxspan=1h
```

One whole track consolidated to single event



Multivalued “MV” fields are used to store vectors of LATITUDE, LONGITIDE, etc.

Pro Tip: slash query time by 5x by removing fields with ‘fields –’

splunk> .conf19

Performance Analysis (II)

Transaction performance?

Transaction is generally slow because it must grind through events and group them

Grinding is minimized by

- 1: specifying a finite time range
- 2: specifying a finite spatial range (H3 cell)
- 4: removing as many unnecessary fields as possible before transacting

On the client, using binary searching to locate sub-portion of vector corresponding to slider position

The more spatial data you have, the more benefit you will get from H3

However, it will not be practical to gather all tracks over all time quickly

Key Takeaways

For geo performance

1. H3 Indexes allow efficient processing of data, by area
2. Time-binned statistics over huge spatial areas gain hundreds-times performance benefit
3. Performance benefits of spatial indexing increase, the wider the extent and volume of your spatial data

Resources

Docs

Ent: https://docs.splunk.com/Documentation/DashApp/0.1.0/DashApp/IntroApp
ITSI: https://docs.splunk.com/Documentation/ITSI/4.3.0/User/BetaFramework
Splunk Investigate: https://docs.splunk.com/Documentation/SplunkInvestigate/Current/Use/DashOverview
Developer: https://developer.splunk.com/scs/docs/dashviz

NPM

https://www.npmjs.com/package/@splunk/dashboard-core

GitHub with Examples

SCS/Investigate Examples - https://github.com/splunk/conf2019-dashboard-scs
Enterprise Examples - https://github.com/splunk/dashboard-conf19-examples

Dashboard Sessions

FN1815 (Tues 1230-115) - The New Dashboarding & Content Export Experience in Splunk: A single experience across Enterprise, SCP, ITSI and more!

DEV 2165 (Tues 415-5) –Deep Dive on The New Dashboarding & Content Export Experience

DEV1141 (Weds 145-230) – Building Applications with Splunk UI and Splunk React Visualizations

FN1933 (Weds 1230-115) – Unleash your Inner Picasso – Splunk's New Dashboard Framework

FN1735 (Weds 1230-115) – What's next for Geo in Splunk

DEV2171 (Thurs 1030-115) – Build your own custom data visualization on dashboard

We want to hear from you!

Reaching the team

Dashboard Usage & Feedback Survey

Visualizations – bit.ly/scs-visualizations
Dashboards – bit.ly/scs-dashboards
Inputs – bit.ly/scs-inputs

Booths

Foundations & Platform > Splunk Enterprise
Developer (Dev Zone) > Dashboards
Developer (Dev Zone) > Visualizations Library

Email

Core Dashboard & Splunk Investigate - dashboards@splunk.com
Core Viz & Splunk Investigate - visualizations@splunk.com
ITSI Experience – itsi-beta-gt-feedback@splunk.com
Enterprise Experience – dashboardsbeta@splunk.com

Q&A

Geoffrey Hendrey | Sr Principal Engineer
Aditi Nath | Software Development Engineer

splunk> .conf19

.conf19

splunk>

Thank You!

Go to the .conf19 mobile app to

RATE THIS SESSION