


DB<sub>4</sub>IoT

# Tracking Down ITS Data Errors

Eimar Boesjes, CEO  
Moonshadow Mobile, Inc.  
[eimar@moonshadow.com](mailto:eimar@moonshadow.com)

Moonshadow






DB<sub>4</sub>IoT

**Let's start with a Simple Question:  
How long does it take my bus to run this route?**





DB<sub>4</sub>IoT

The answer is not so simple because of . . . traffic

Traffic is affected by:

Day Type

Time of Day

Construction

Accidents

Holidays

Events



# DB<sub>4</sub>IoT

To get detailed information on run times  
we often put riders on buses to time a route.

The Drawbacks:  
Expensive  
Time Consuming  
Is the data representative?  
You can't go back in time



# DB<sub>4</sub>IoT

**We already collect detailed information on run times  
in our CAD/AVL systems.**

**The data is free**

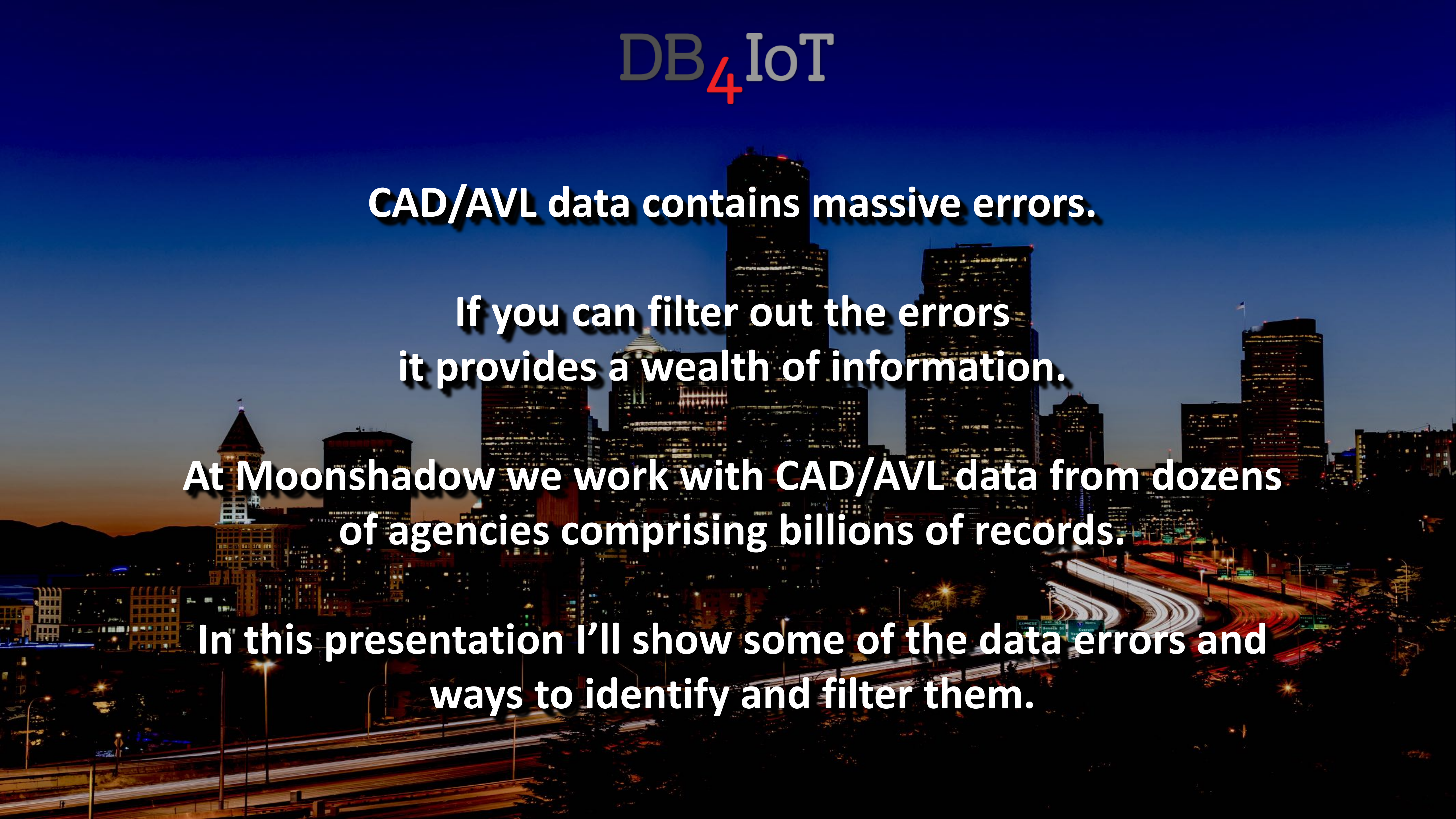
**The data is representative**

**You can go back in time**

**You can do before, during and after comparisons**

**But . . .**



A nighttime photograph of a city skyline, likely Denver, Colorado, featuring several illuminated skyscrapers and a multi-lane highway in the foreground with light trails from traffic. The sky is a deep blue, and the city lights are warm and vibrant.

DB<sub>4</sub>IoT

**CAD/AVL data contains massive errors.**

**If you can filter out the errors  
it provides a wealth of information.**

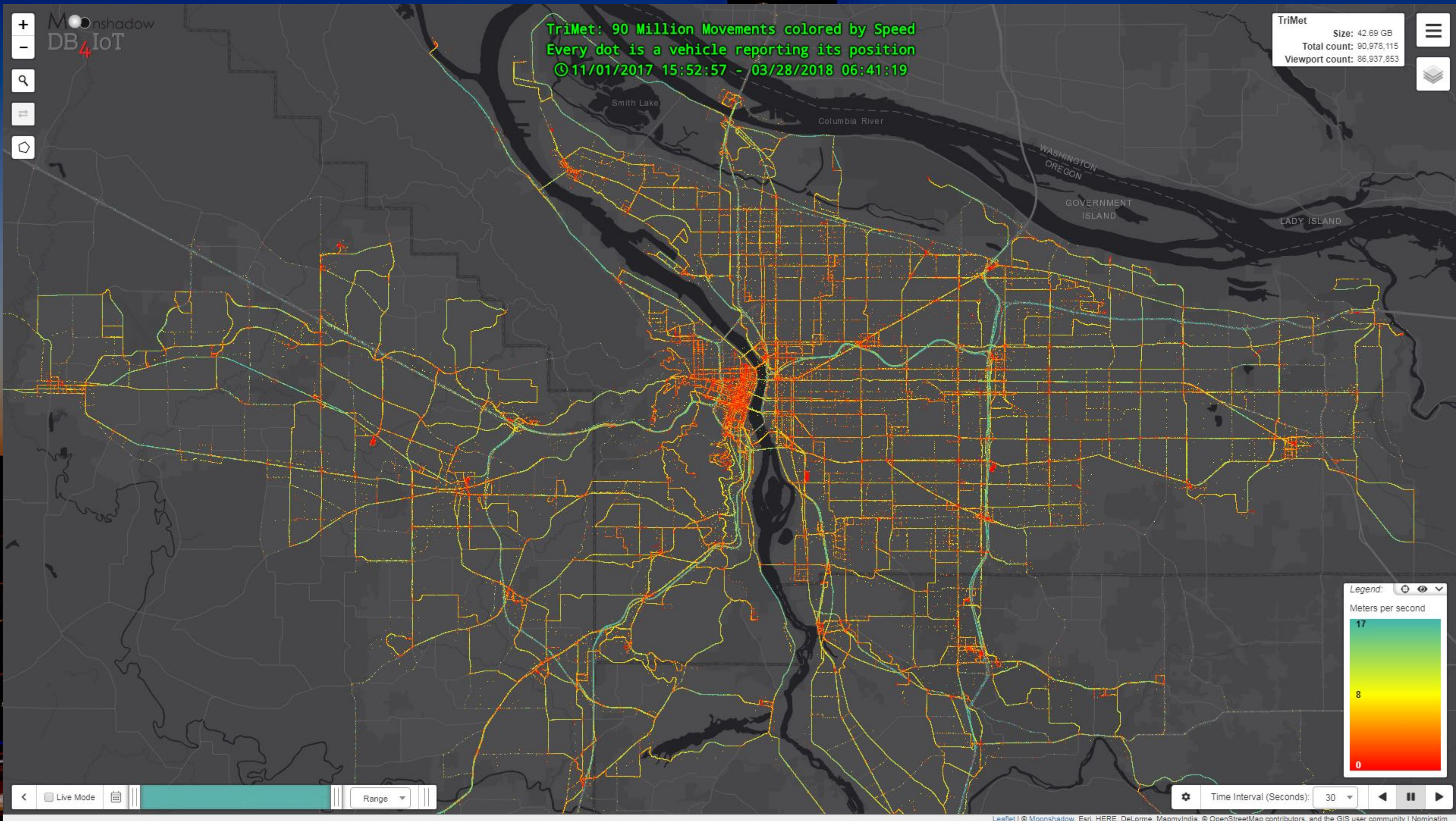
**At Moonshadow we work with CAD/AVL data from dozens  
of agencies comprising billions of records.**

**In this presentation I'll show some of the data errors and  
ways to identify and filter them.**



# DB<sub>4</sub>IoT

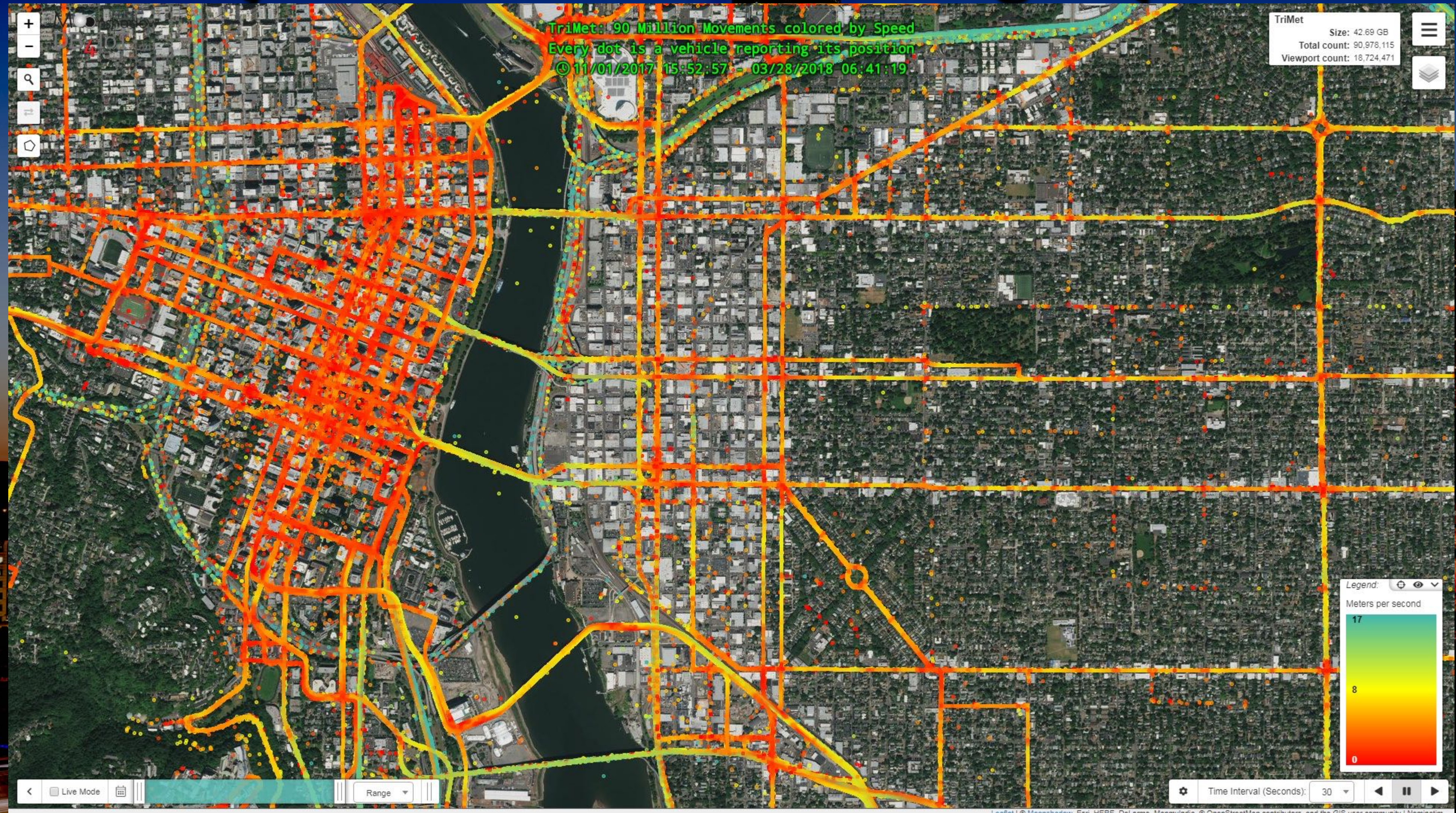
## What does CAD/AVL data look like?





# DB<sub>4</sub>IoT

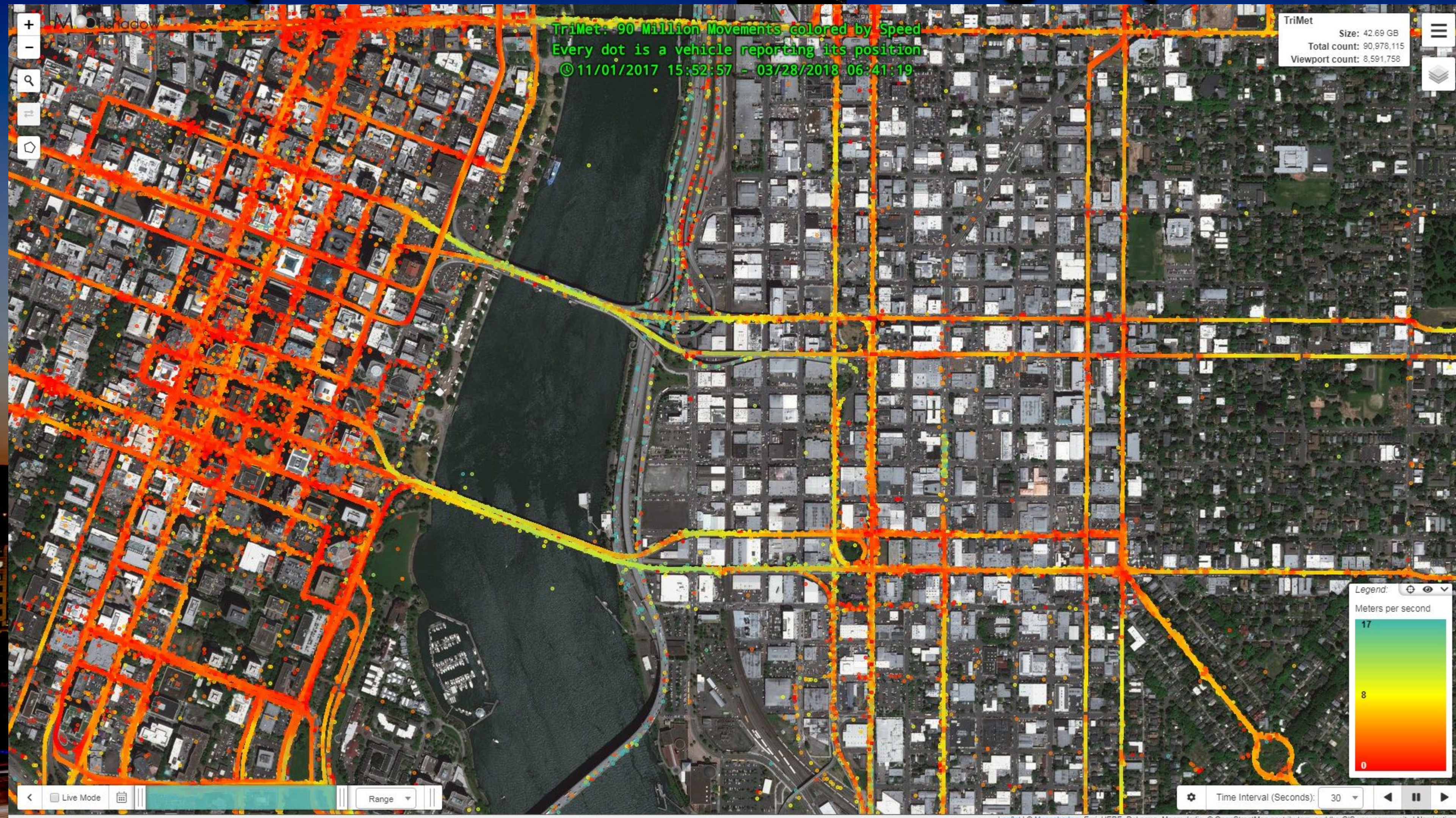
Every dot is a vehicle reporting its position





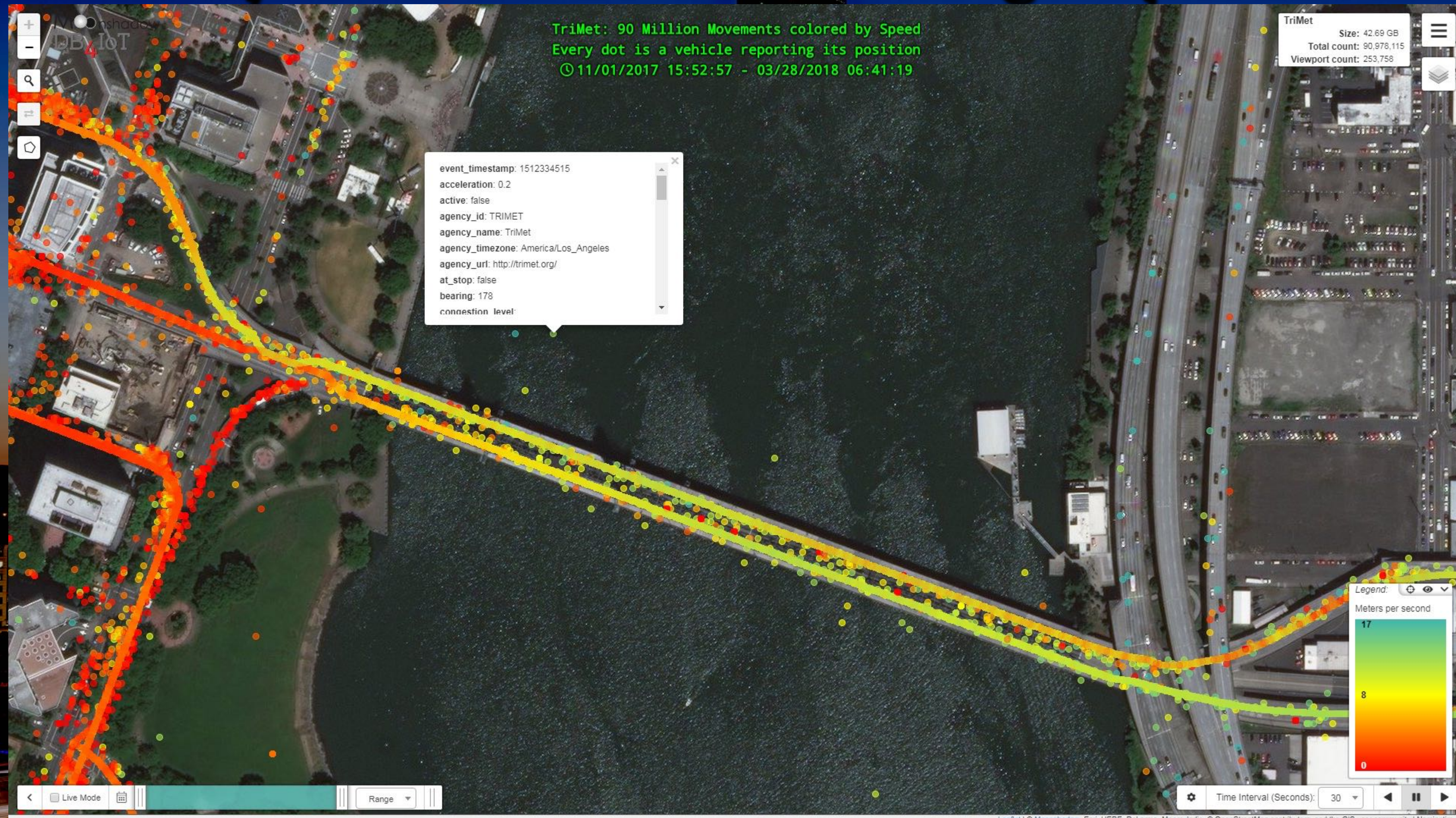
# DB<sub>4</sub>IoT

Every dot is a vehicle reporting its position





## Every dot is a vehicle reporting its position

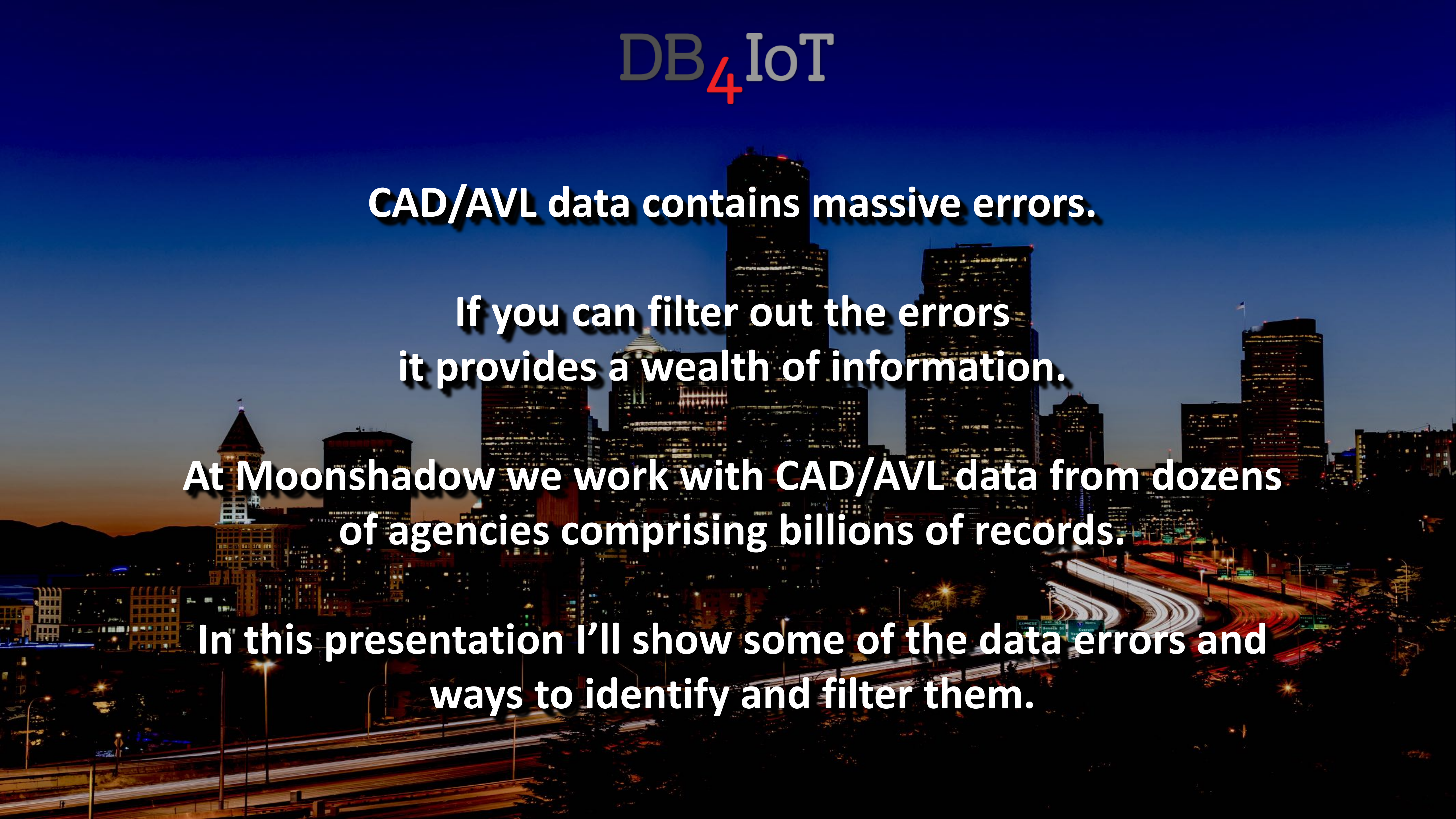




## GPS Errors: Some buses are in the Willamette River







DB<sub>4</sub>IoT

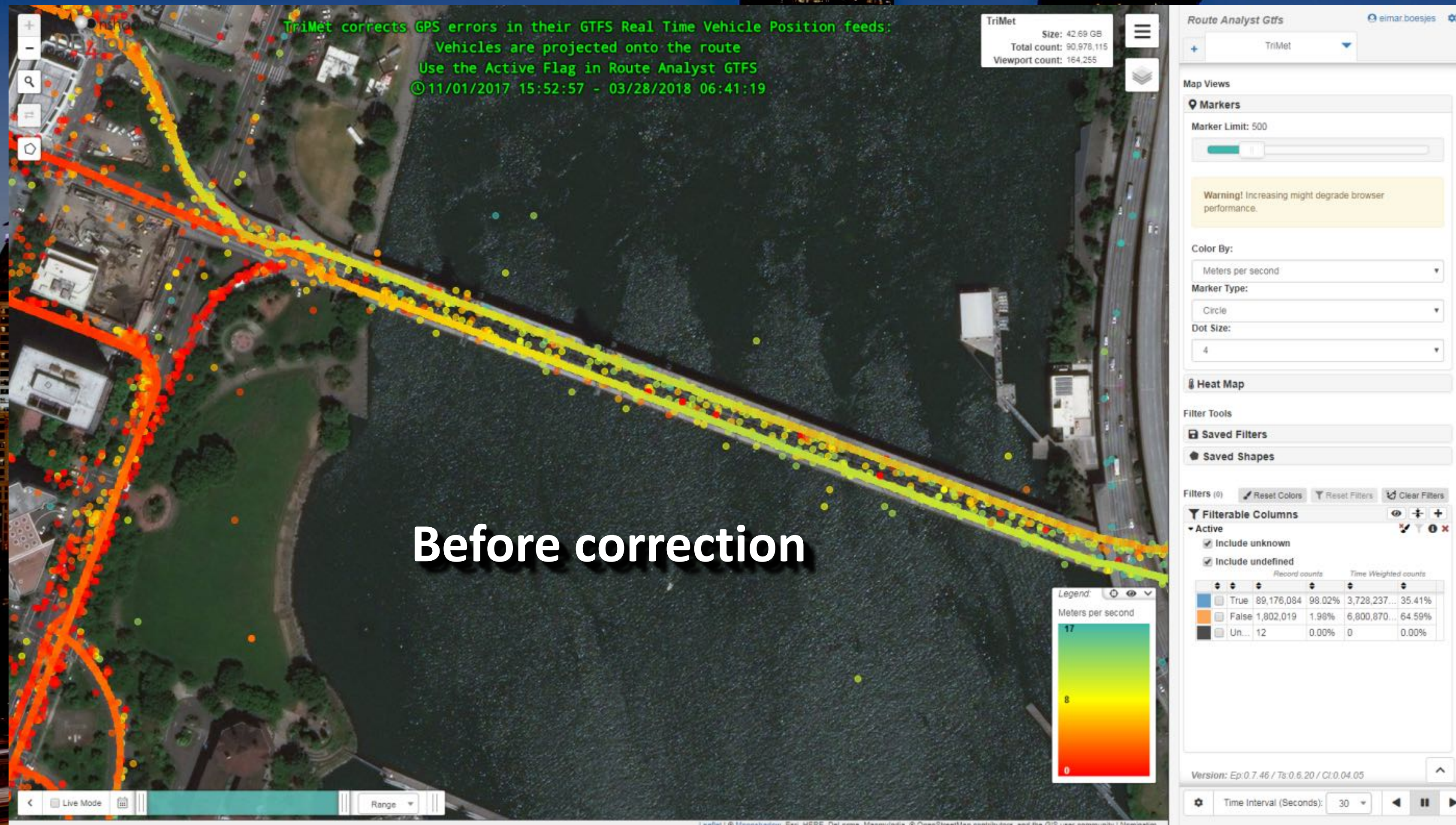
**GPS Errors:**

**Unless you correct GPS errors  
You can't use the GPS data for distance  
Therefore you can't use the data for speed  
You can still use the data for travel times**



# DB<sub>4</sub>IoT

## TriMet corrects GPS Errors in their GTFS vehicle position feeds





# DB<sub>4</sub>IoT

## TriMet corrects GPS Errors in their GTFS vehicle position feeds

TriMet corrects GPS errors in their GTFS Real Time Vehicle Position feeds:  
Vehicles are projected onto the route  
Use the Active Flag in Route Analyst GTFS  
© 11/01/2017 15:52:57 - 03/28/2018 06:41:19

TriMet  
Size: 42.89 GB  
Total count: 89,176,084  
Viewport count: 162,046

After correction

Legend  
Meters per second  
17  
8  
0

Route Analyst Gtfs  
eimar.boesjes

TriMet

Map Views

Markers  
Marker Limit: 500

Warning! Increasing might degrade browser performance.

Color By:  
Meters per second

Marker Type:  
Circle

Dot Size:  
4

Heat Map

Filter Tools

Saved Filters

Saved Shapes

Filters (1) [Reset Colors] [Reset Filters] [Clear Filters]

Filterable Columns

Active

Include unknown

Include undefined

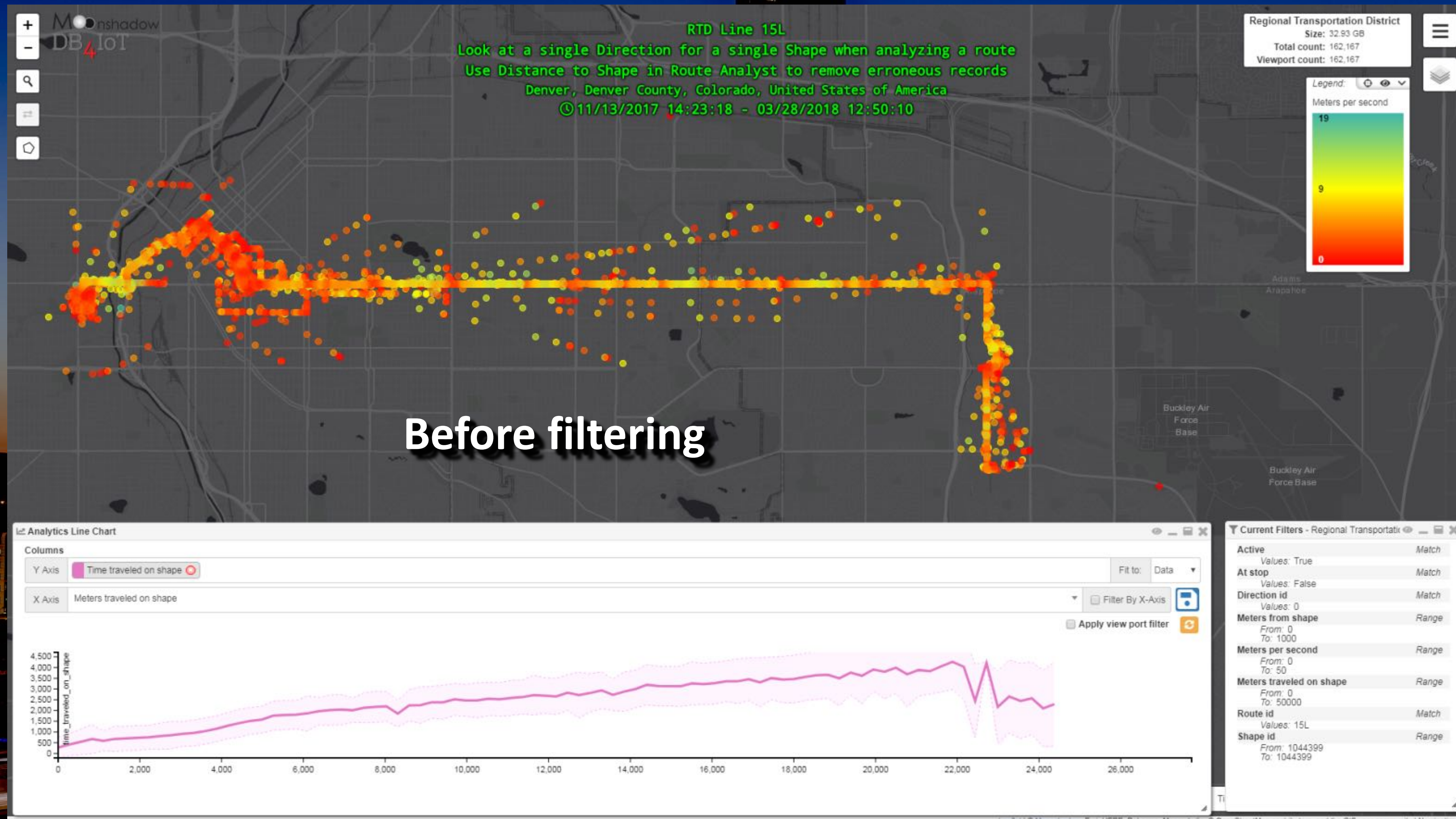
	Record counts	Time Weighted counts
<input checked="" type="checkbox"/> True	89,176,084	100% 3,728,237...
<input type="checkbox"/> False	0	0.00% 0 0.00%
<input type="checkbox"/> Un...	0	0.00% 0 0.00%

Version: Ep: 0.7.46 / Ts: 0.6.20 / Cl: 0.04.05

Time Interval (Seconds): 30

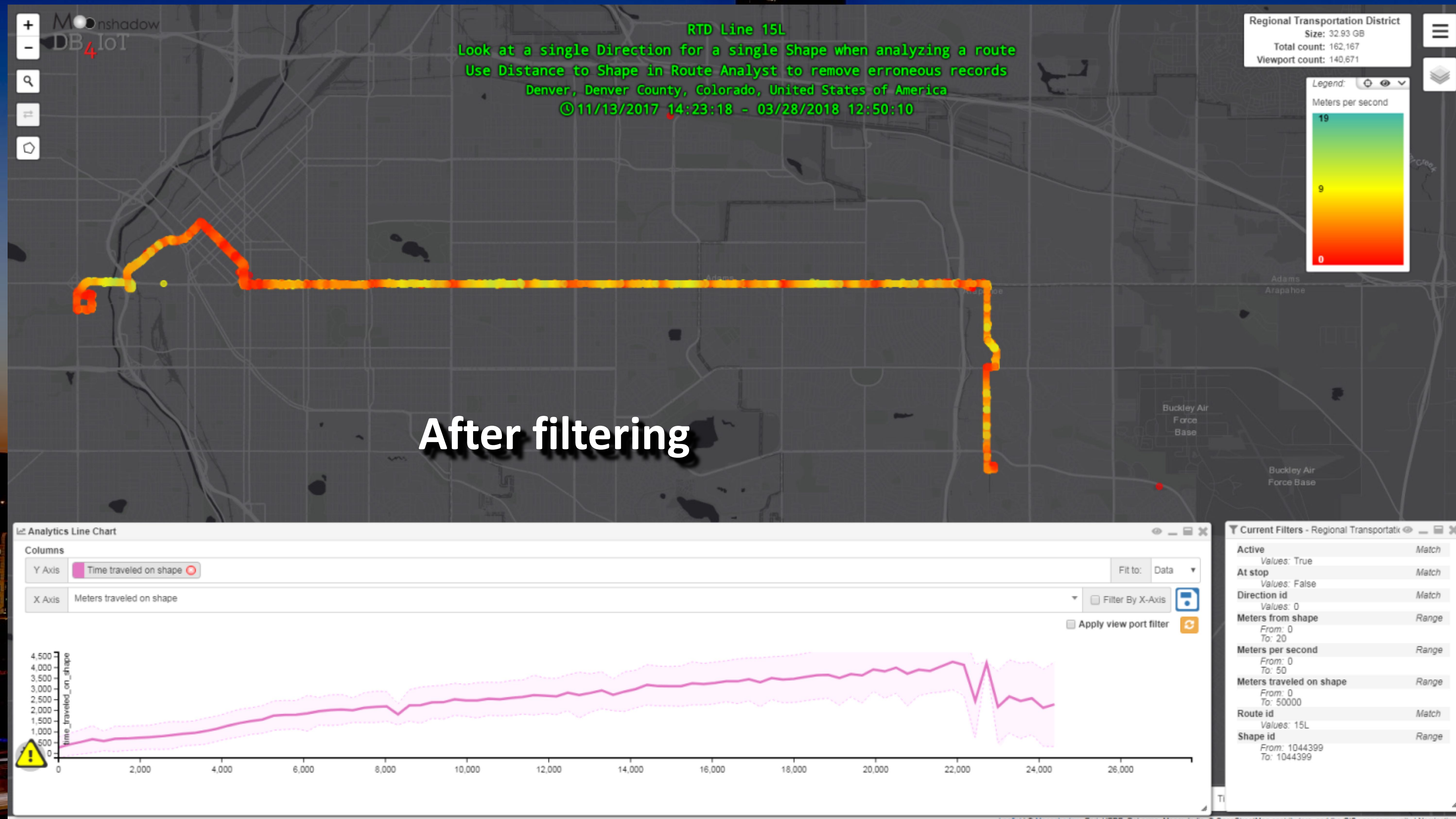


## Erroneous records can be filtered out with distance to shape





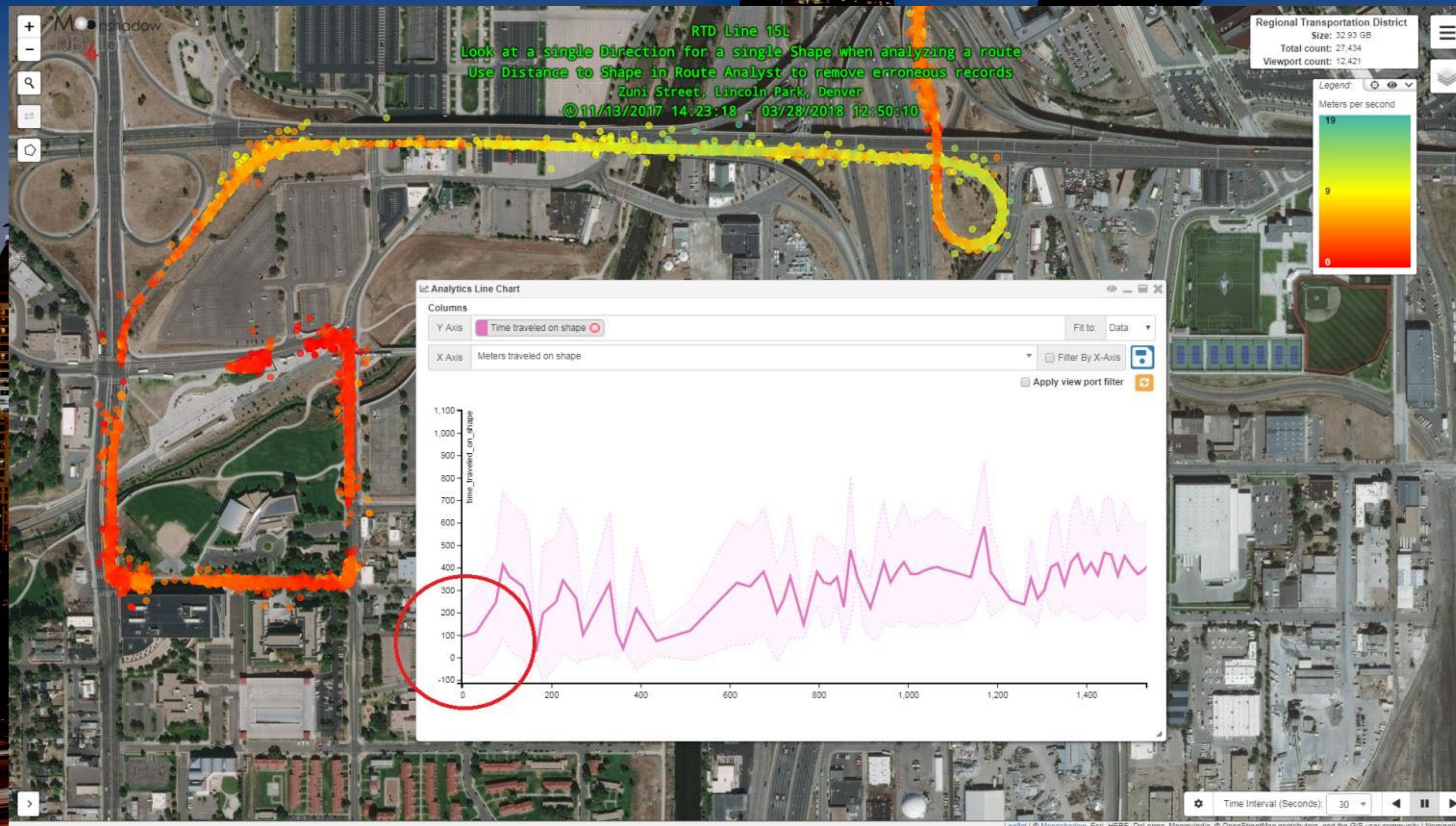
## Erroneous records can be filtered out with distance to shape





# DB<sub>4</sub>IoT

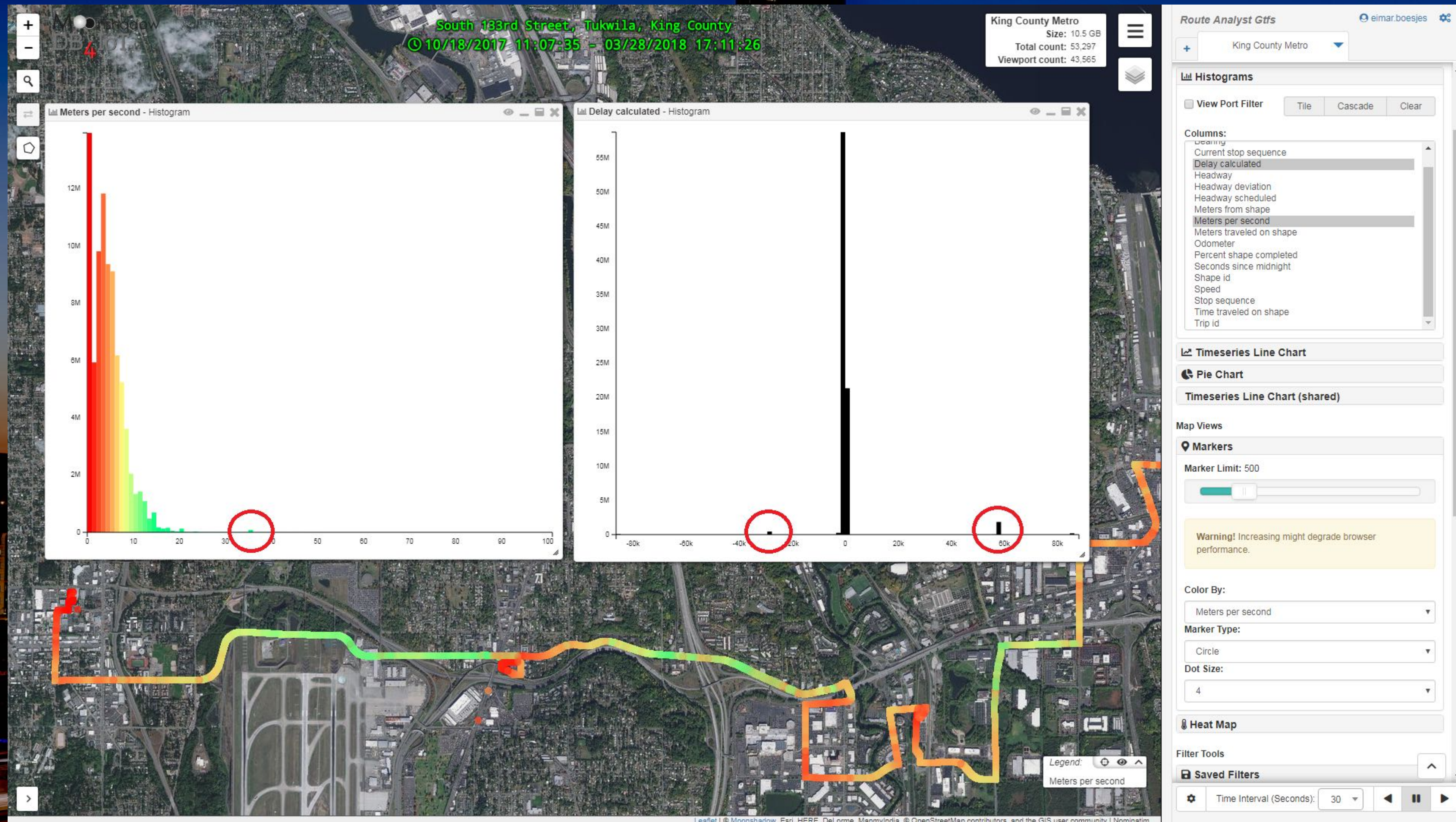
Route Logging is started too early  
As a result the start time is not zero





# DB<sub>4</sub>IoT

## Histograms show outliers to filter out

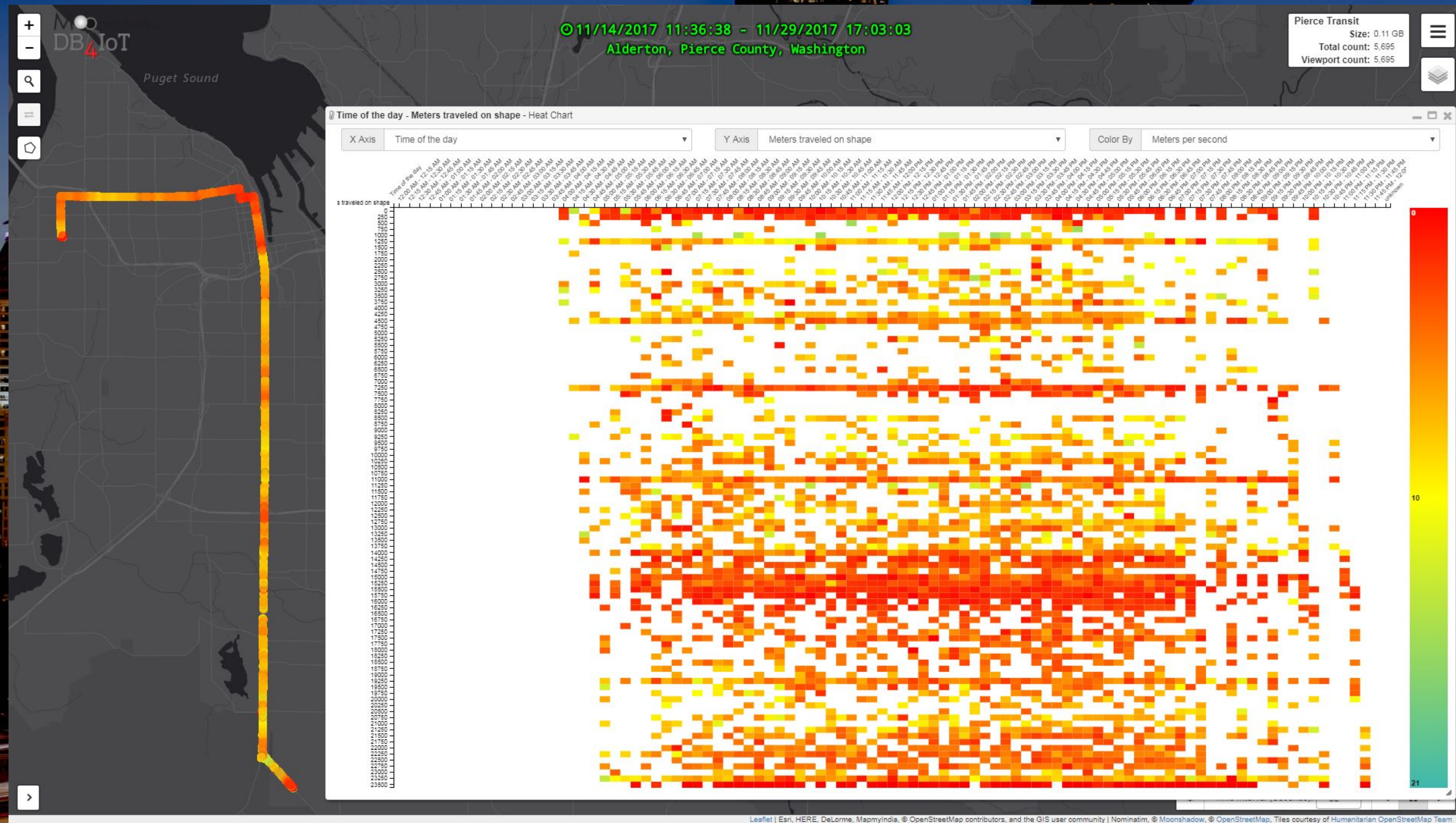




# DB<sub>4</sub>IoT

## White Space = No Data

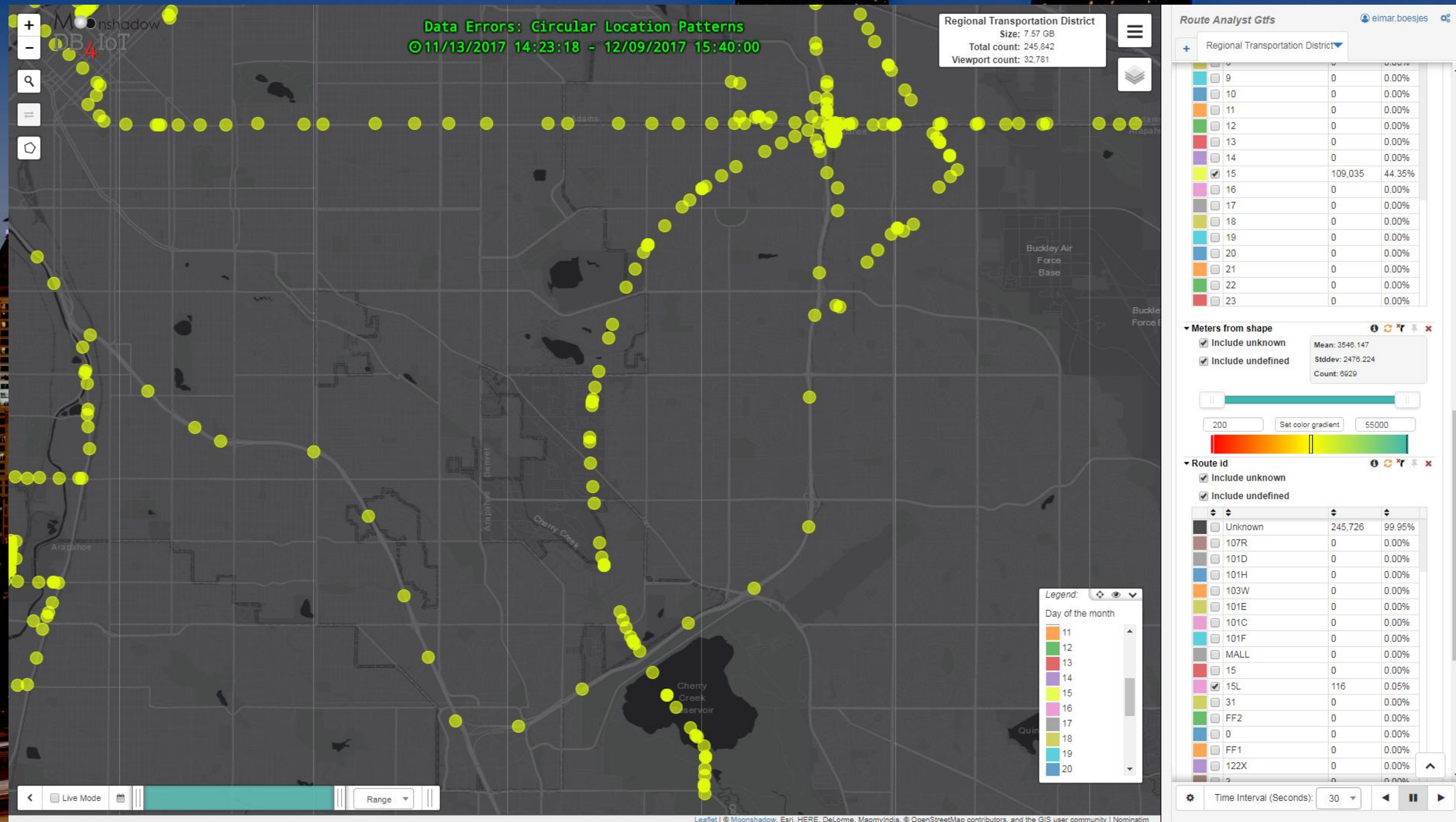
# CAD/AVL Systems don't report 30 seconds after service stops





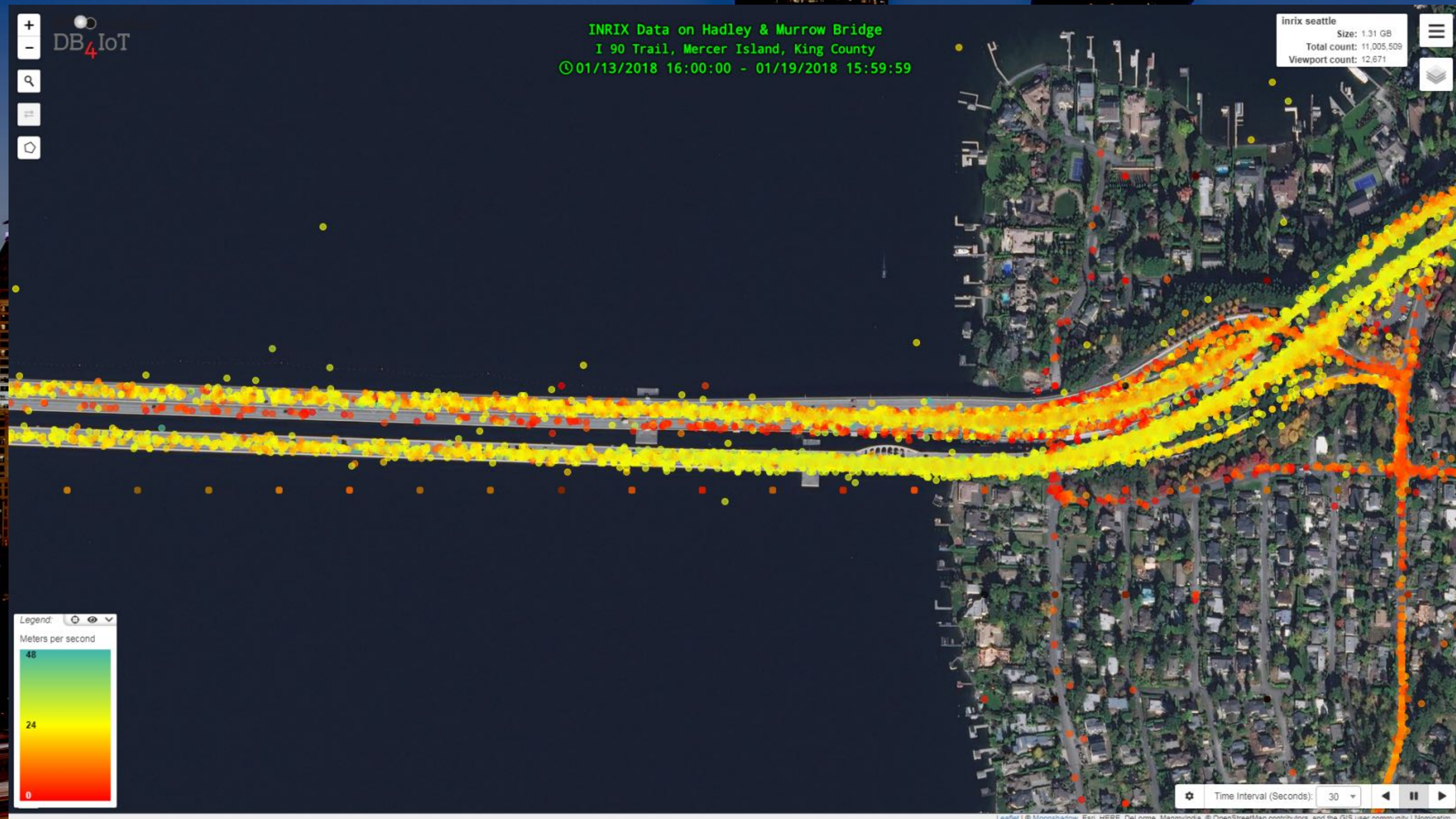
## Data Errors: Circular Patterns

Patterns become clear only when the data is mapped





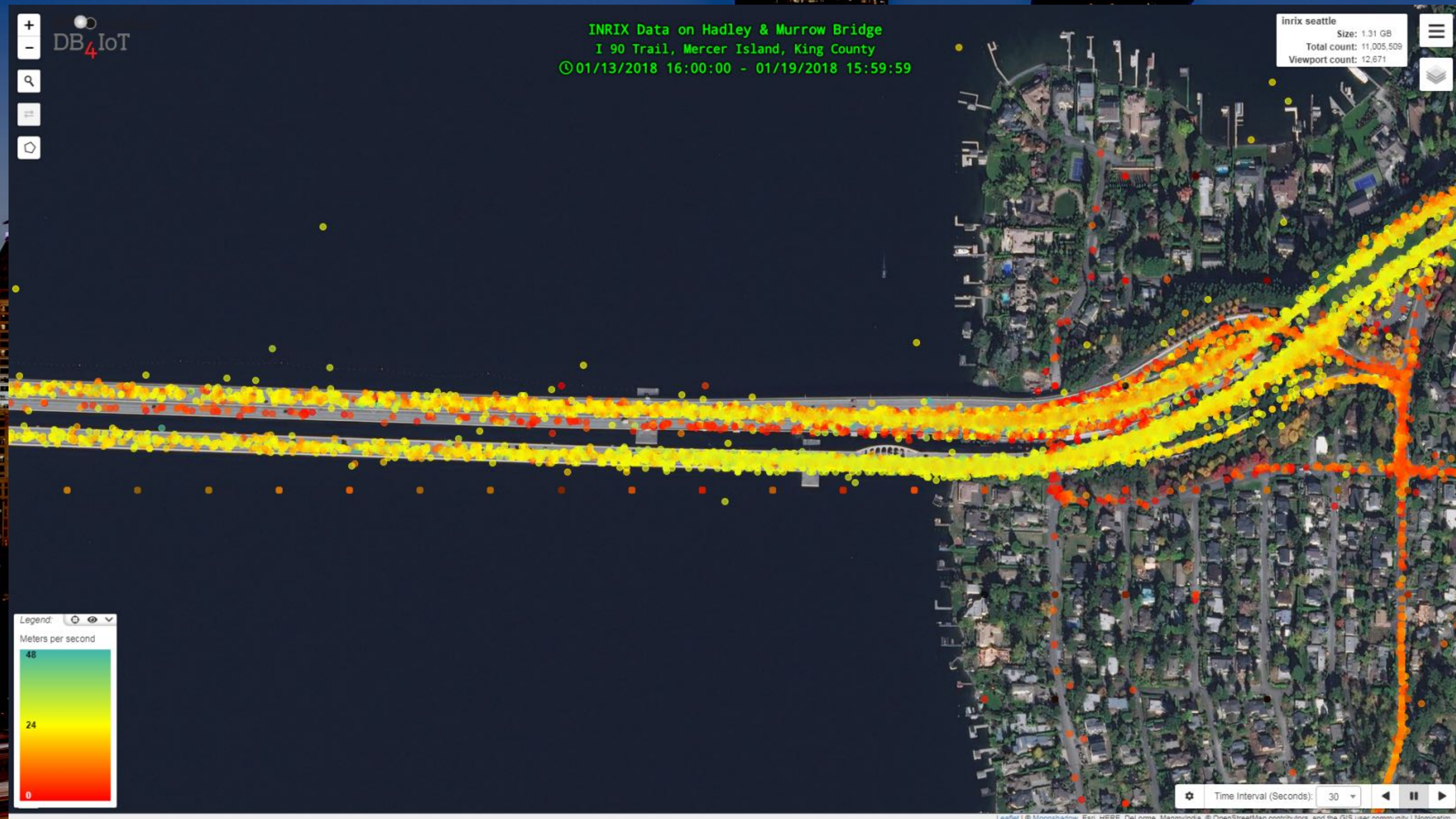
## INRIX Data shows traffic is slow on the Westbound left lane Are these data errors?





# DB<sub>4</sub>IoT

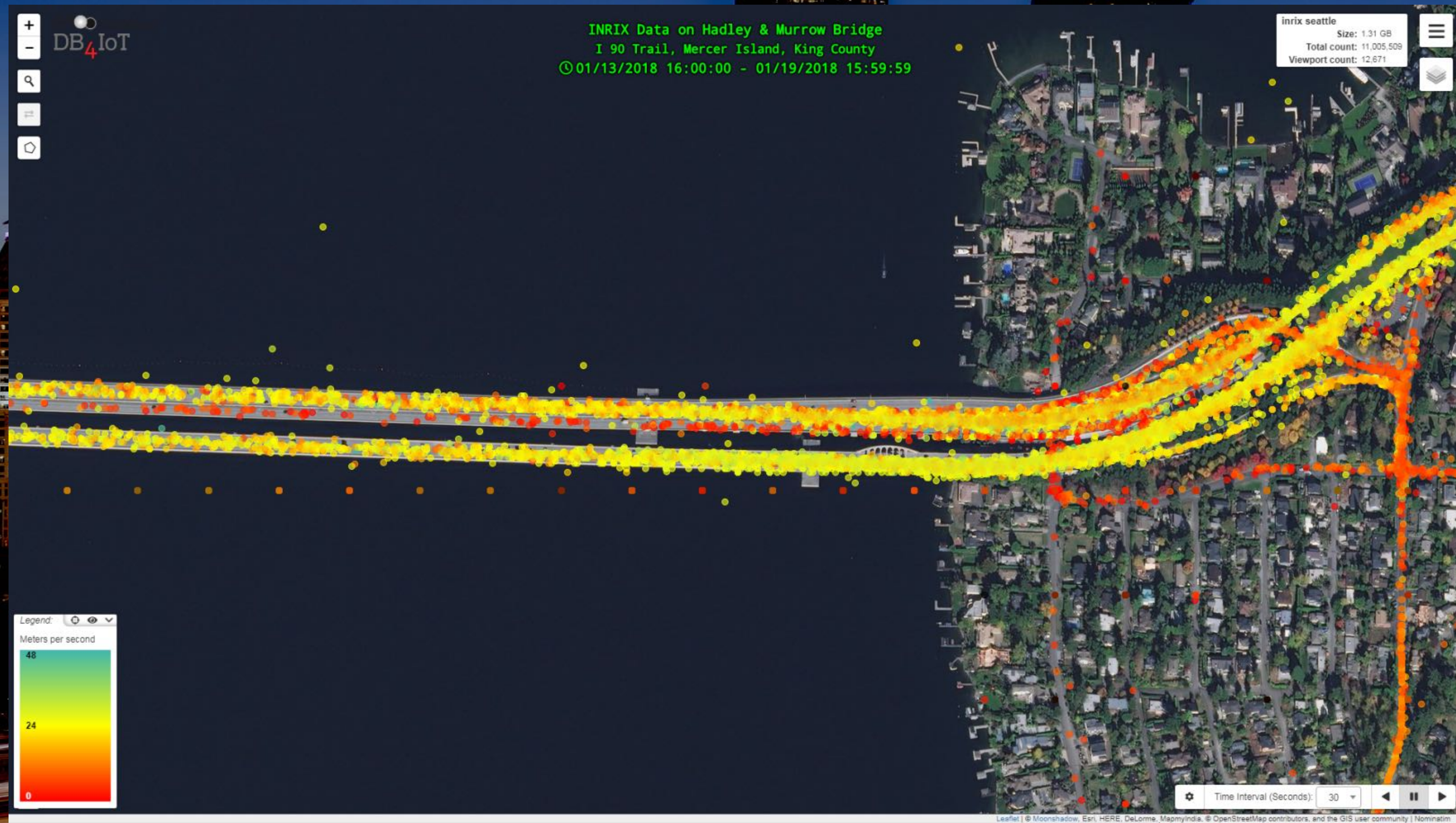
The Westbound left lane is under construction and we're seeing construction traffic.





# DB<sub>4</sub>IoT

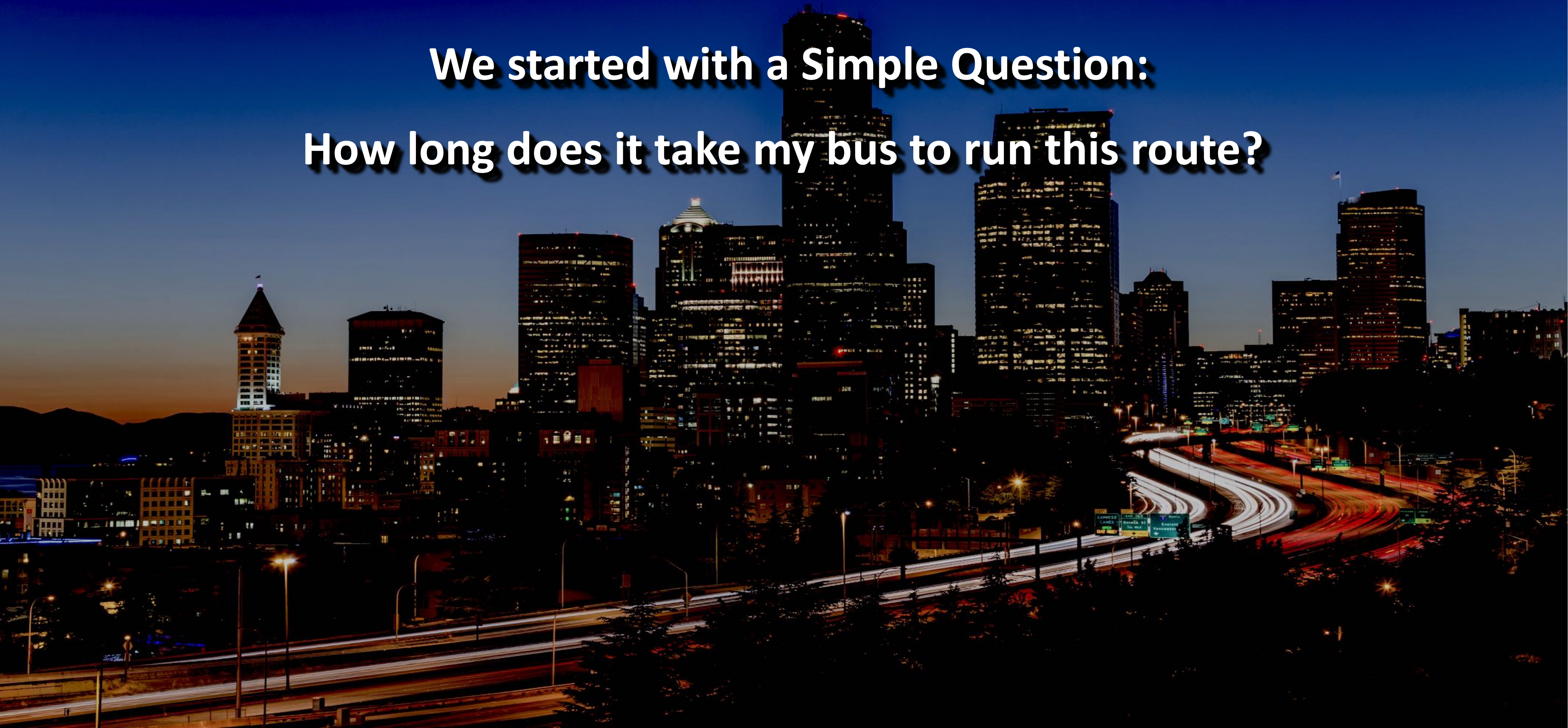
## When analyzing data always know the local situation



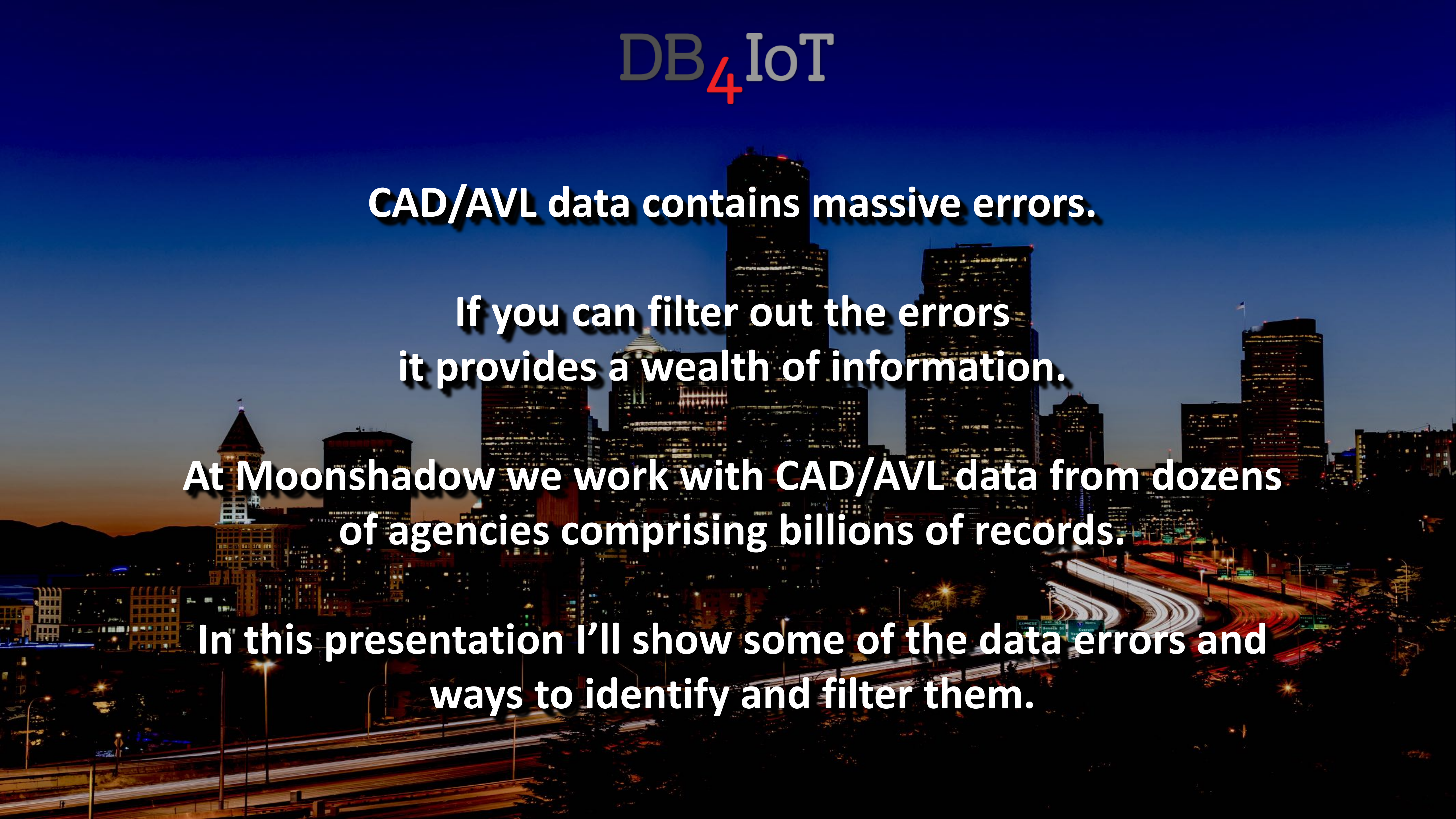


DB<sub>4</sub>IoT

**We started with a Simple Question:  
How long does it take my bus to run this route?**





A nighttime photograph of a city skyline with numerous illuminated skyscrapers. In the foreground, a multi-lane highway is visible with long-exposure light trails from cars and buses, creating streaks of white and red light. The sky is a deep blue, suggesting dusk or dawn.

DB<sub>4</sub>IoT

**We know how long it takes to run a route:**

**At any time of the day  
on weekdays or weekends  
before, during and after construction.**

**We can see when and where delays are incurred,  
buses are speeding to catch up  
or traffic is slowing down service.**



**CAD/AVL Data can provide a wealth of information**

**However,  
Drawing conclusions from CAD/AVL data requires:**

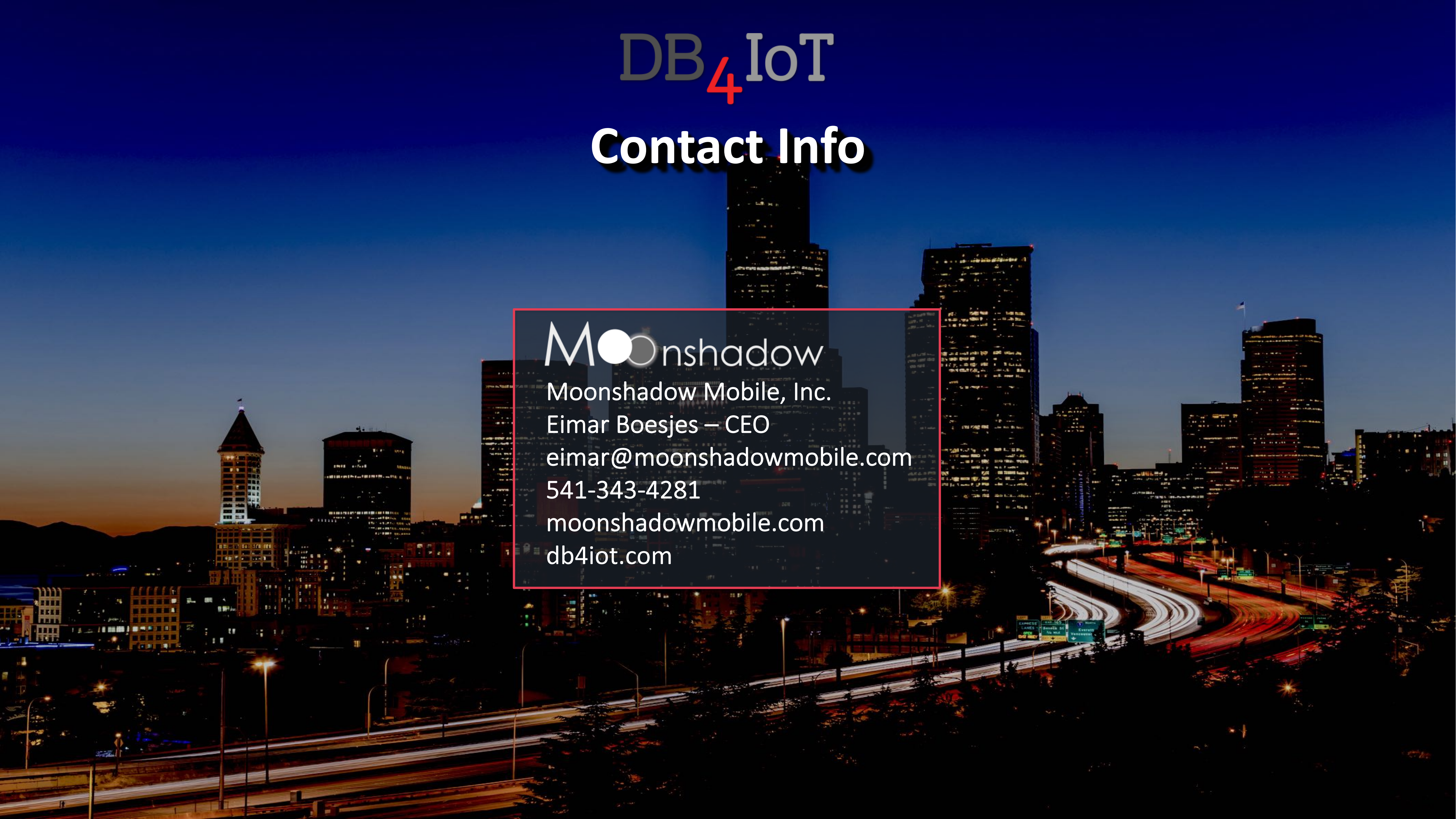
**Public transit expertise**

**Big data expertise**

**Knowledge of the local situation**

**Data visualization to identify errors**





DB<sub>4</sub>IoT

## Contact Info

Moonshadow

Moonshadow Mobile, Inc.

Eimar Boesjes – CEO

[eimar@moonshadowmobile.com](mailto:eimar@moonshadowmobile.com)

541-343-4281

[moonshadowmobile.com](http://moonshadowmobile.com)

[db4iot.com](http://db4iot.com)