Splunking the Tube

June 2022

splunk> turn data into doing®

Forward-Looking Statements

This presentation may contain forward-looking statements regarding future events, plans or the expected financial performance of our company, including our expectations regarding our products, technology, strategy, customers, markets, acquisitions and investments. These statements reflect management's current expectations, estimates and assumptions based on the information currently available to us. These forward-looking statements are not guarantees of future performance and involve significant risks, uncertainties and other factors that may cause our actual results, performance or achievements to be materially different from results, performance or achievements expressed or implied by the forward-looking statements contained in this presentation.

For additional information about factors that could cause actual results to differ materially from those described in the forward-looking statements made in this presentation, please refer to our periodic reports and other filings with the SEC, including the risk factors identified in our most recent quarterly reports on Form 10-Q and annual reports on Form 10-K, copies of which may be obtained by visiting the Splunk Investor Relations website at www.investors.splunk.com or the SEC's website at www.sec.gov. The forward-looking statements made in this presentation are made as of the time and date of this presentation. If reviewed after the initial presentation, even if made available by us, on our website or otherwise, it may not contain current or accurate information. We disclaim any obligation to update or revise any forward-looking statement based on new information, future events or otherwise, except as required by applicable law.

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. We undertake no obligation either to develop the features or functionalities described, in beta or in preview (used interchangeably), or to include any such feature or functionality in a future release.

John Murdoch

June 2022

Sales Engineer with Splunk from Feb 2019

Why did I do this?

Material I prepped for .conf a few times…

Show what Splunk can do with an open, dynamic data set:

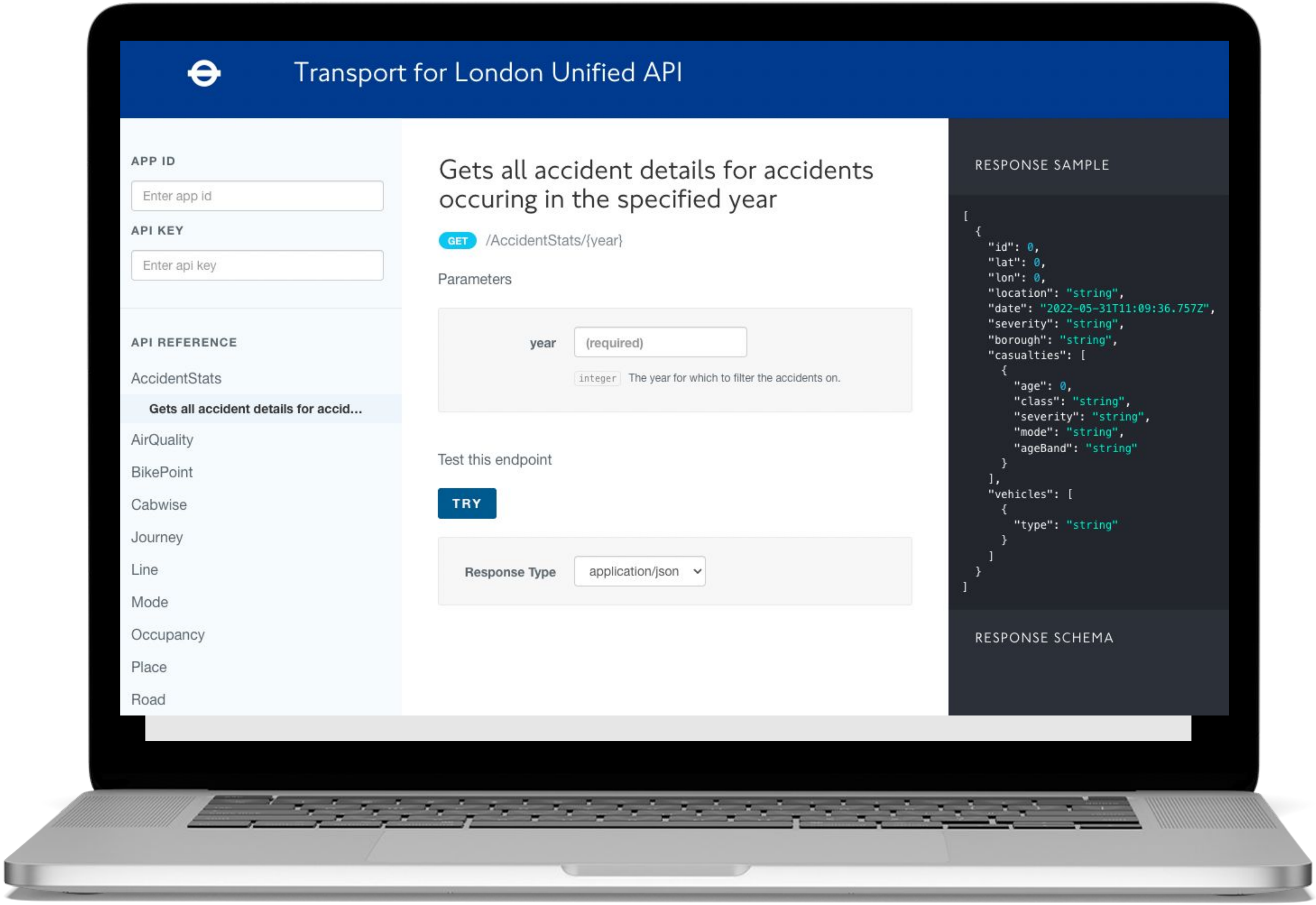
- Simple, graphical, recognisable, explainable, use-case-laden dataset
- Open data source that anyone can interrogate
- Temporal / seasonal data for those into statistical analysis / ML
- Opportunities to demonstrate core Splunk functionality:
  - GDI, Scaling, Search, Analysis, Dashboarding, Correlation
  - e.g. where is/was my train? = arrivals dataset + route dataset
- Extensible scope
  - station crowding, buses & boats, roads, connecting services (national rail, airports, taxi)
- Dependencies to explore
  - tube service = line status + vehicle arrival + station ops + route status
- Splunk office / expo showcases
  - Showcase value of Splunk to visitors, clients, students and employees without needless explaining

TfL API

Swagger Interface

First up was to explore the API, try out the various REST API end-points. Also to ascertain if the data is good as-is, or needs preparation.

1) Use curl to save JSON
2) Upload into Splunk
3) Analyse data



Gets all accident details for accidents occuring in the specified year

GET /AccidentStats/{year}

Parameters

year (required)

integer The year for which to filter the accidents on.

Test this endpoint

Response Type application/json

```
[
  {
    "id": 0,
    "lat": 0,
    "lon": 0,
    "location": "string",
    "date": "2022-05-31T11:09:36.757Z",
    "severity": "string",
    "borough": "string",
    "casualties": [
      {
        "age": 0,
        "class": "string",
        "severity": "string",
        "mode": "string",
        "ageBand": "string"
      }
    ],
    "vehicles": [
      {
        "type": "string"
      }
    ]
  }
]
```

Data onboarding experience

More than just pointing Splunk at a URL

How to best ingest the data to Splunk

- API Keys, headers, cookies, etc - not difficult for TfL data, but API key required
- Data parsing (JSON, XML, CSV, etc) and data nesting (sequenced, mapped)

Make the data user-friendly (and Splunk-friendly)

- 4MB payload with disparate, nested datasets → many 375-byte events with essential context
    - **End-user benefits:** readable, succinct data for the end-user
    - **Splunk benefits:** filtering with SPL results in performance gains
- The only time where schema-at-read is probably not the best option

Use Splunk to explore the data, look for abnormalities, further enrichment use-cases:

- **Missing:** Previous stops, including origin, is unrecorded. Essential for root-cause analysis.
- **Incomplete:** Invalid routes due to many branches. However plotted elsewhere using Lat / Lon
- **Duplicate:** Platform changes result in data duplications due to TTL

REST Output

e.g Waterloo & City

20Kb for a 2 station route!

Contains datasets on:

- Overall Line metadata
- Absolute coordinates of the line(s)
- List of Stations
  - Transport connections lists
- Stop Points per branch / direction
  - Connecting Lines
  - Station coordinates, metadata
- Possible route permutations

```
{
    "lineId": "waterloo-city",
    "lineName": "Waterloo & City",
    "direction": "all",
    "isOutboundOnly": false,
    "mode": "tube",
    "lineStrings": [
        "[[[-0.088899,51.513356],[-0.11478,51.503299]]]",
        "[[[-0.11478,51.503299],[-0.088899,51.513356]]]"
    ],
    "stations": [
       "modes": [ ],
       "lines": [{ }]
    ],
    "stopPointSequences": [{
            "stopPoint": [{
                    "id": "940GZZLUBNK",
                    "name": "Bank Underground Station",
                    "lat": 51.513356,
                    "lon": -0.088899,
                    "lines": []
                   }, {
                  ……
     }],
     "orderedLineRoutes": [
        {
            "name": "Bank  &harr;  Waterloo ",
            "naptanIds": [
                "940GZZLUBNK",
                "940GZZLUWLO"
            ],
            "serviceType": "Regular"
        }, {
       }
     ]
}
```

All of this “defines” a route. Needs consolidating and refactoring!

TfL Technical Add-on

Add-on Builder to the rescue

splunk> turn data into doing®

Splunk Output

Any route station

~350-400 bytes / stop / route

Contains the following

- Record timestamp
- LineId, Name, Direction
- Stop sequence, detail, coordinates
- Related stop information:
  - Origin
  - Destination
  - Previous

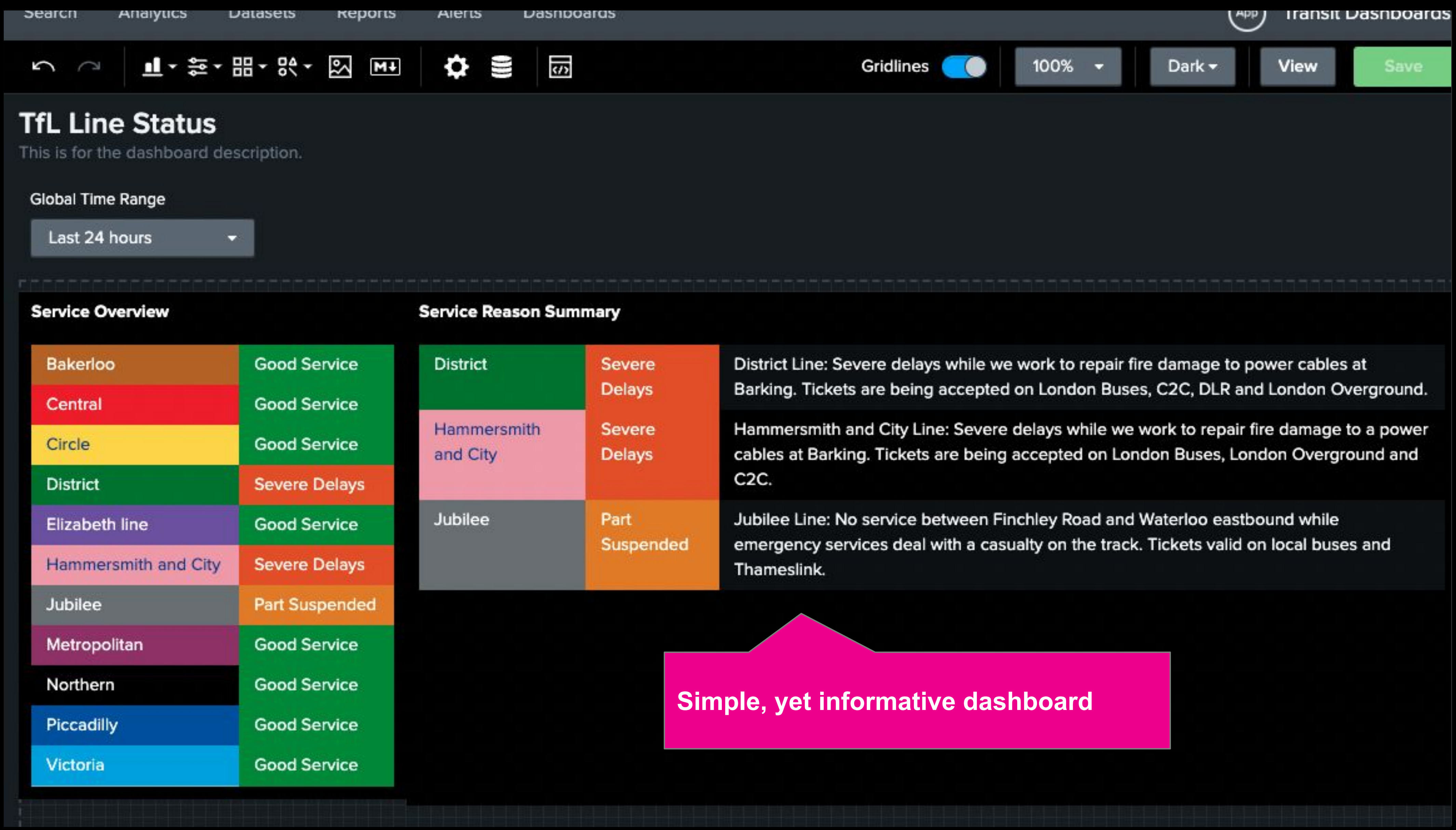
```
{
    "timestamp": 1653913766.372975,
    "routeId": "910GPADTON-910GRDNGSTN",
    "lineId": "elizabeth",
    "lineName": "Elizabeth line",
    "direction": "outbound",
    "destinationNaptanId": "910GRDNGSTN",
    "originNaptanId": "910GPADTON",
    "stationNum": 15,
    "naptanId": "910GRDNGSTN",
    "stationName": "Reading Rail Station",
    "latitude": 51.458786,
    "longitude": -0.971863,
    "prevNaptanId": "910GTWYFORD"
}
```

This is far easier to query / stitch together / correlate

Working with the data

We’re going to need a demonstration…

splunk> turn data into doing®



TfL Line Status

This is for the dashboard description.

Global Time Range: Last 24 hours

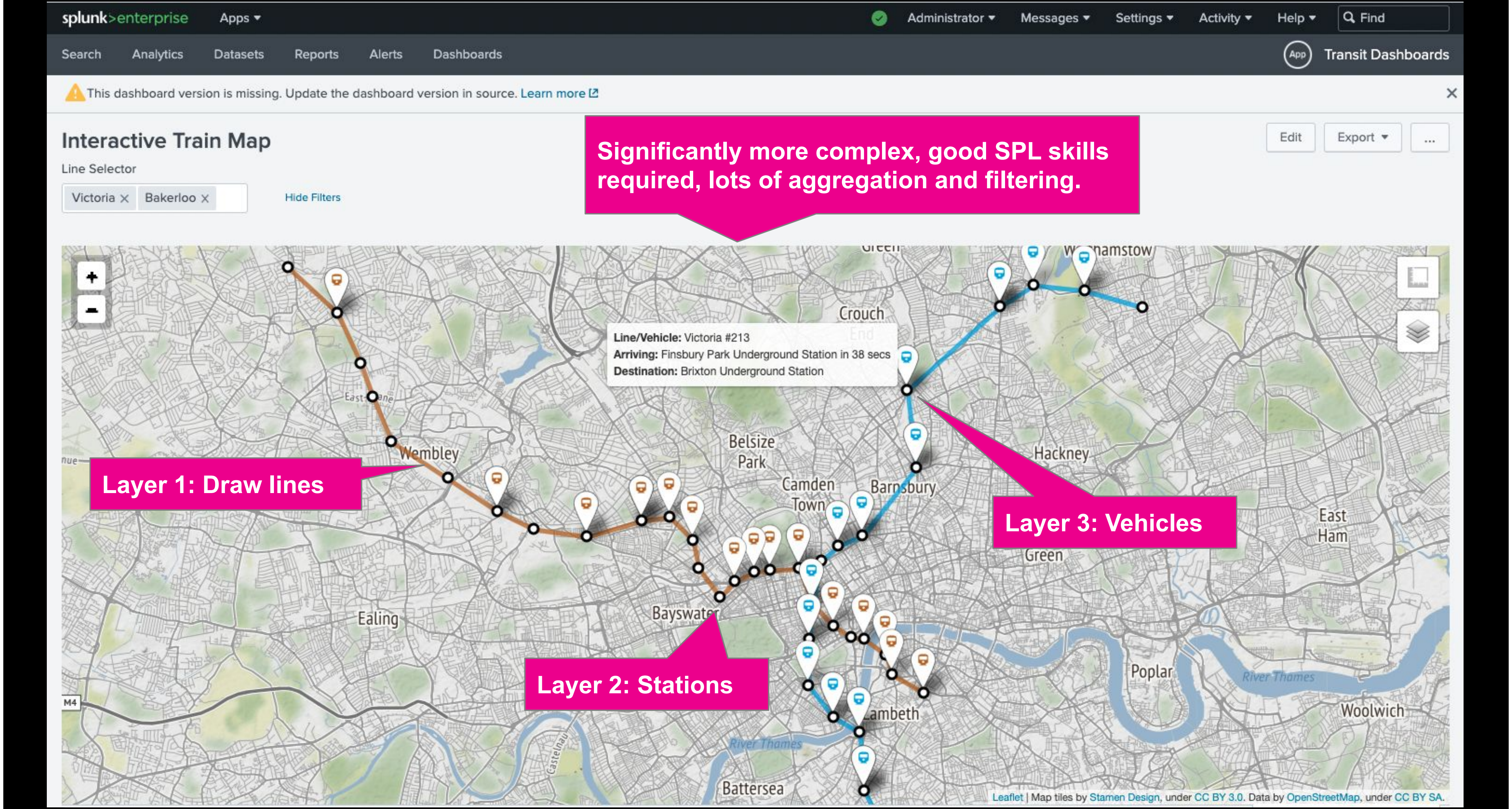
Service Overview

Bakerloo	Good Service
Central	Good Service
Circle	Good Service
District	Severe Delays
Elizabeth line	Good Service
Hammersmith and City	Severe Delays
Jubilee	Part Suspended
Metropolitan	Good Service
Northern	Good Service
Piccadilly	Good Service
Victoria	Good Service

Service Reason Summary

District	Severe Delays	District Line: Severe delays while we work to repair fire damage to power cables at Barking. Tickets are being accepted on London Buses, C2C, DLR and London Overground.
Hammersmith and City	Severe Delays	Hammersmith and City Line: Severe delays while we work to repair fire damage to a power cables at Barking. Tickets are being accepted on London Buses, London Overground and C2C.
Jubilee	Part Suspended	Jubilee Line: No service between Finchley Road and Waterloo eastbound while emergency services deal with a casualty on the track. Tickets valid on local buses and Thameslink.

Simple, yet informative dashboard



This dashboard version is missing. Update the dashboard version in source. Learn more

Interactive Train Map

Line Selector

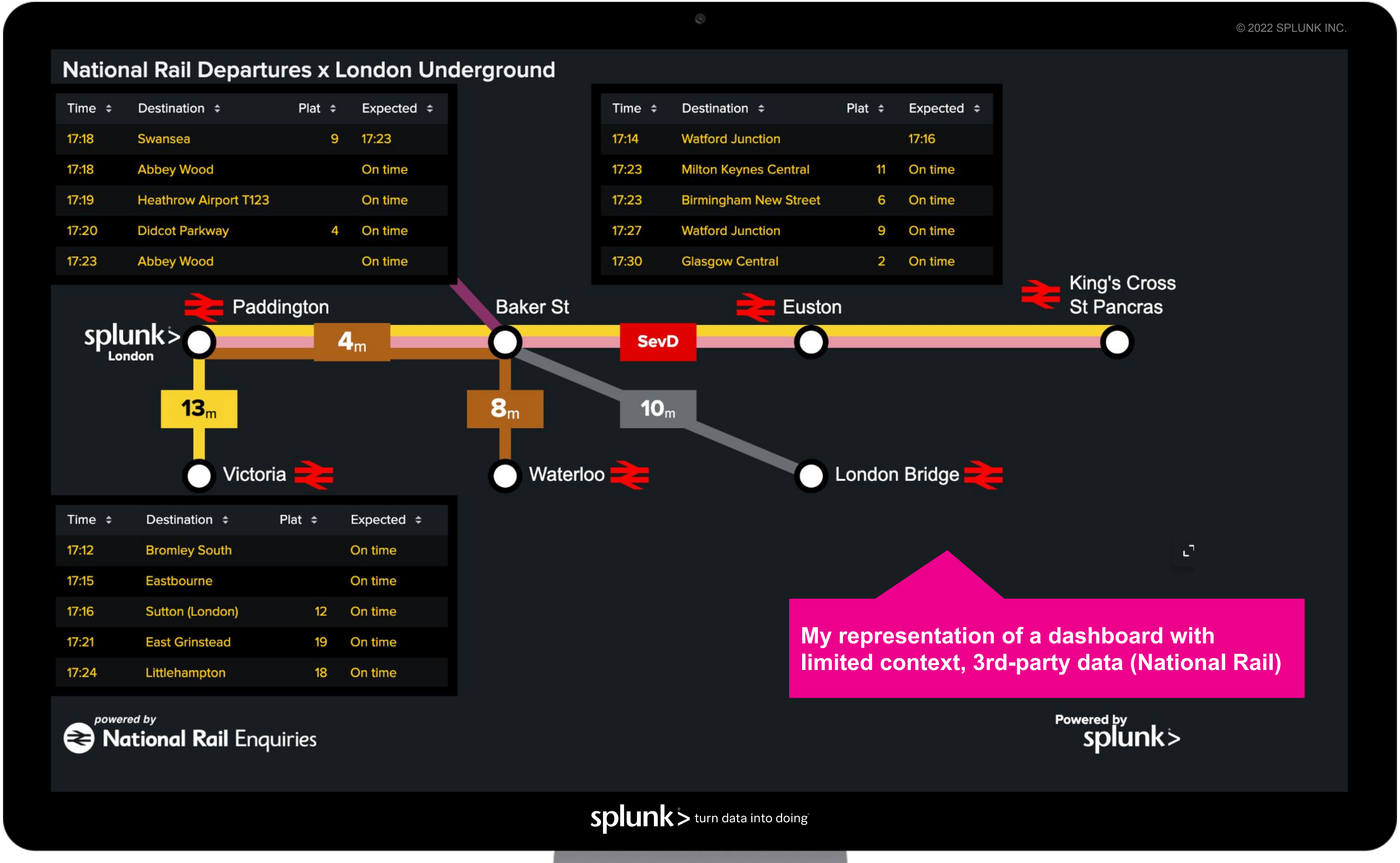
Line/Vehicle: Victoria #213
Arriving: Finsbury Park Underground Station in 38 secs
Destination: Brixton Underground Station

Significantly more complex, good SPL skills required, lots of aggregation and filtering.

Layer 1: Draw lines

Layer 2: Stations

Layer 3: Vehicles



National Rail Departures x London Underground

Time	Destination	Plat	Expected
17:18	Swansea	9	17:23
17:18	Abbey Wood		On time
17:19	Heathrow Airport T123		On time
17:20	Didcot Parkway	4	On time
17:23	Abbey Wood		On time

Time	Destination	Plat	Expected
17:14	Watford Junction		17:16
17:23	Milton Keynes Central	11	On time
17:23	Birmingham New Street	6	On time
17:27	Watford Junction	9	On time
17:30	Glasgow Central	2	On time

Time	Destination	Plat	Expected
17:12	Bromley South		On time
17:15	Eastbourne		On time
17:16	Sutton (London)	12	On time
17:21	East Grinstead	19	On time
17:24	Littlehampton	18	On time

powered by National Rail Enquiries

Powered by splunk>

splunk> turn data into doing

My representation of a dashboard with limited context, 3rd-party data (National Rail)

Future plans - Further development

Optional subtitle

Get this into the hands of Splunk gurus with better examples:

- Sandbox environment with a days data (e.g. Monday 3am-Tuesday 3am)
- ReactJS dashboards
  - Classic looks classic
  - Dashboard studio is not fully-realized (complex drilldowns, comaps, pop-ups)
- More services:
  - Station crowding, buses, bicycles, charging points
  - National Rail, Airports
- Leverage Splunk ITSI to look from a service perspective
  - Comparing two similar, but interlinked services. Could we show Central vs Elizabeth line?
  - Showing cause and effect: e.g. line disruption on line stations, other lines & stations
  - Predictive analysis (peak usage, contributing factors, etc)

Future plans - Splunk Product

Optional subtitle

Endeavour to get this working alongside:

- Self-service to Splunk staff & partners through our dedicated platform
  - Learning - rich data playground
  - Showcasing
- Explore new Dashboard Studio capabilities in future Splunk iterations
  - Graphical customisation improvements
  - Drilldown improvements
- Splunk Cloud Developer Edition
- Combine with wider Splunk Observability Suite
  - Code performance
  - API response monitoring

Thank You!

splunk> turn data into doing®