WORKING WEEK 2025
AND Locate25 THE NATIONAL GEOSPATIAL CONFERENCE
Collaboration, Innovation and Resilience: Championing a Digital Generation
FIG Geospatial Council of Australia
Brisbane, Australia 6–10 April

Uniting Data Science and GIS:

Spatial Analysis with Databricks & ArcGIS

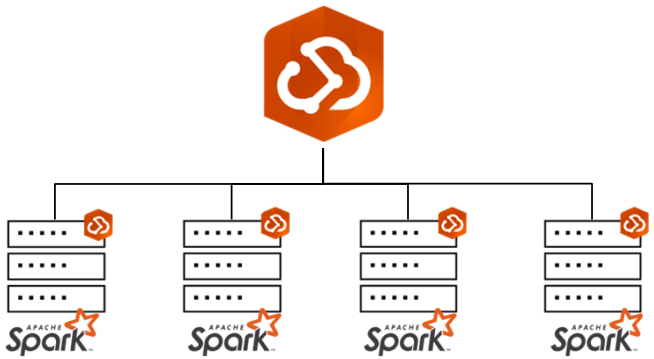
Simon Jackson | Esri Australia | Technology Strategist

ORGANISED BY FIG Geospatial Council of Australia
PLATINUM SPONSORS Australian Government CHCNAV esri THE SCIENCE OF WHERE Leica Geosystems Meter Surveyors Australia

- Advanced Spatial Analytics in Databricks
- Use cases & examples
- Surfacing results from Databricks into GIS

ArcGIS GeoAnalytics Engine

ArcGIS GeoAnalytics Engine



- Spark native library that extends dataframes with spatial processing and analytics
- Over 200 spatial Functions & Tools
- Analyse data from data lakes, warehouses as well as ArcGIS systems

Supported Environments

Microsoft Fabric

Azure Synapse Analytics

Amazon EMR

Google Cloud Dataproc

databricks

Your own Cluster

Data Sources

DATA FORMATS & SYSTEMS SUPPORTED BY SPARK

Generic Files
- CSV file
- Hive table
- JSON file
- Parquet file
- XML file
- Zip file
- ORC files

Data Lakes & Databases
- Amazon S3
- Azure Blob Storage
- Azure Data Lake Storage
- Azure Cosmos DB
- Cassandra
- Couchbase
- ElasticSearch
- Google Cloud Storage
- MongoDB
- Neo4j
- Oracle
- Redis
- Riak Time Series
- SQL databases using JDBC
- Hadoop HDFS

Data Warehouses
- Amazon Redshift
- Azure Synapse Analytics
- Google BigQuery
- Snowflake

ADDS SUPPORT FOR:

Spatial Formats
- Shapefile (Read/Write)
- Feature Service (Read/Write)
- File GeoDatabase (Read)
- Vector Tiles (Write)
- GeoJSON (Read/Write)
- GeoParquet (Read/Write)
- *Use ArcGIS API for Python to write into additional spatial formats.*

Spatial Functions

CONSTRUCTORS

- ST_AsBinary
- ST_AsText
- ST_AsGeoJSON
- ST_AsEsriJSON
- ST_AsShape
- ST_GeomFromBinary
- ST_GeomFromText
- ST_GeomFromGeoJSON
- ST_GeomFromEsriJSON
- ST_GeomFromShape
- ST_PointFromBinary
- ST_PointFromText
- ST_PointFromGeoJSON
- ST_PointFromShape
- ST_MPointFromBinary
- ST_MPointFromText
- ST_MPointFromEsriJSON
- ST_MPointFromGeoJSON
- ST_MPointFromShape
- ST_LineFromBinary
- ST_LineFromText
- ST_LineFromEsriJSON
- ST_LineFromGeoJSON
- ST_LineFromShape
- ST_PolyFromBinary
- ST_PolyFromText
- ST_PolyFromEsriJSON
- ST_PolyFromGeoJSON
- ST_PolyFromShape
- ST_Point
- ST_MultiPoint
- ST_LineString
- ST_MultiLineString
- ST_Polygon
- ST_MultiPolygon

ACCESSORS

- ST_GeometryType
- ST_Area
- ST_Azimuth
- ST_GeodesicArea
- ST_Centroid
- ST_PointOnSurface
- ST_CoordDim
- ST_Dimension
  ST_Distance
- ST_GeodesicDistance
- ST_GeodesicLength
- ST_GeometryN
- ST_Is3D
- ST_IsClosed
- ST_IsEmpty
- ST_IsMeasured
- ST_Length
- ST_GeodesicLength
- ST_MaxM
- ST_MaxX
- ST_MaxY
- ST_MaxZ
- ST_MinM
- ST_MinX
- ST_MinY
- ST_MinZ
- ST_NumGeometries
- ST_NumInteriorRings
- ST_NumPoints
- ST_PointN
- ST_IsRing
- ST_SRID
- ST_StartPoint
- ST_EndPoint
- ST_X, ST_Y, ST_Z, ST_M

OPERATIONS

- ST_Aggr_ConvexHull
- ST_Aggr_Intersection
- ST_Aggr_Linestring
- ST_Aggr_Union
- ST_Aggr_MeanCenter
- ST_Aggr_StdevEllipse
- ST_Bbox_Intersects
- ST_Boundary
- ST_Buffer
- ST_GeodesicBuffer
- ST_Cast
- ST_Centerline
  ST_ClosestPoint
- ST_ConvexHull
- ST_Densify
- ST_Difference
- ST_Envelope
- ST_ExteriorRing
- ST_Flip
- ST_InteriorRingN
- ST_Rotate
- ST_Scale
- ST_Shear
- ST_Split
- ST_SymDifference
- ST_Union
- ST_Segmentize
- ST_Segments
- ST_ShortestLine
- ST_Points
- ST_Geometries
- ST_Generalize
- ST_Simplify
- ST_Transform
- ST_Translate

RELATIONSHIP TESTS

- ST_Contains
- ST_Crosses
- ST_Disjoint
- ST_EnvIntersect
- ST_Equals
- ST_Intersects
- ST_Overlaps
- ST_Relate
- ST_Touches
- ST_Within
- ST_DWithin

TRACK FUNCTIONS

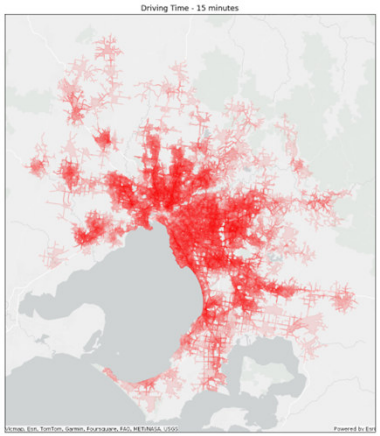
- TRK_After
- TRK_Aggr_CreateTrack
- TRK_Before
- TRK_Between
- TRK_DistanceAlong
- TRK_Duration
- TRK_DurationAlong
- TRK_EndTimestamp
- TRK_IsValid
- TRK_LCSS
- TRK_Length
- TRK_Query
- TRK_Speed
- TRK_SplitByDistance
- TRK_SplitByDistanceGap
- TRK_SplitByDuration
- TRK_SplitByTimeGap
- Trk_StartTimestamp

BINNING

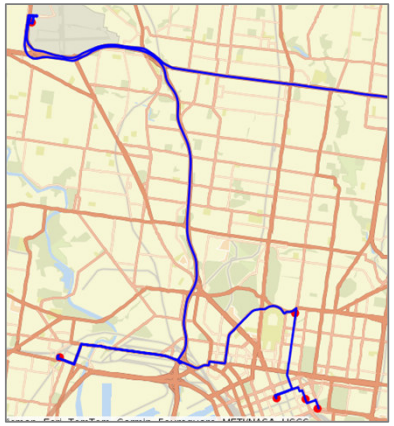
- ST_SquareBin
- ST_SquareBins
- ST_BinGeometry
- ST_BinCenter
- ST_HexBin
- ST_HexBins
- ST_H3Bin
- ST_H3Bins
- ST_BinId

Spatial Analysis Tools

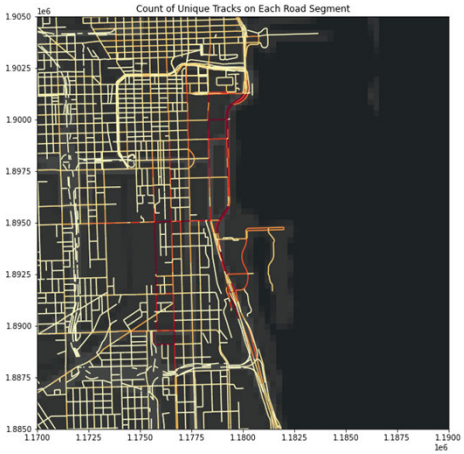
Hotspot analysis of crash data



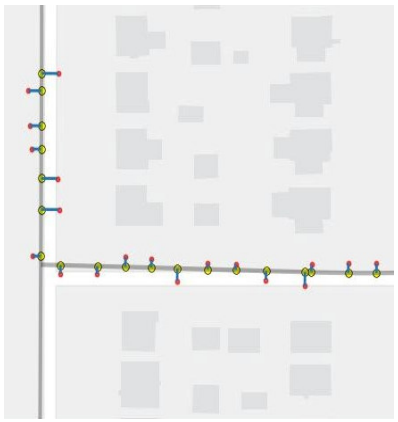
Walk & Drive times to schools



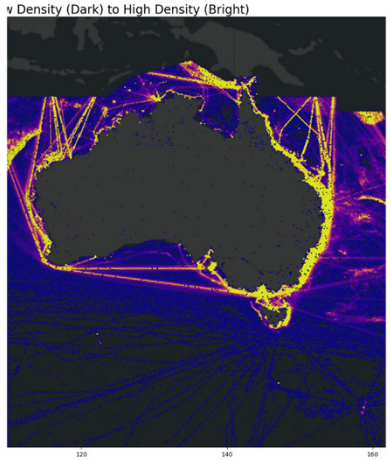
Route optimisation



Counting trips across segments

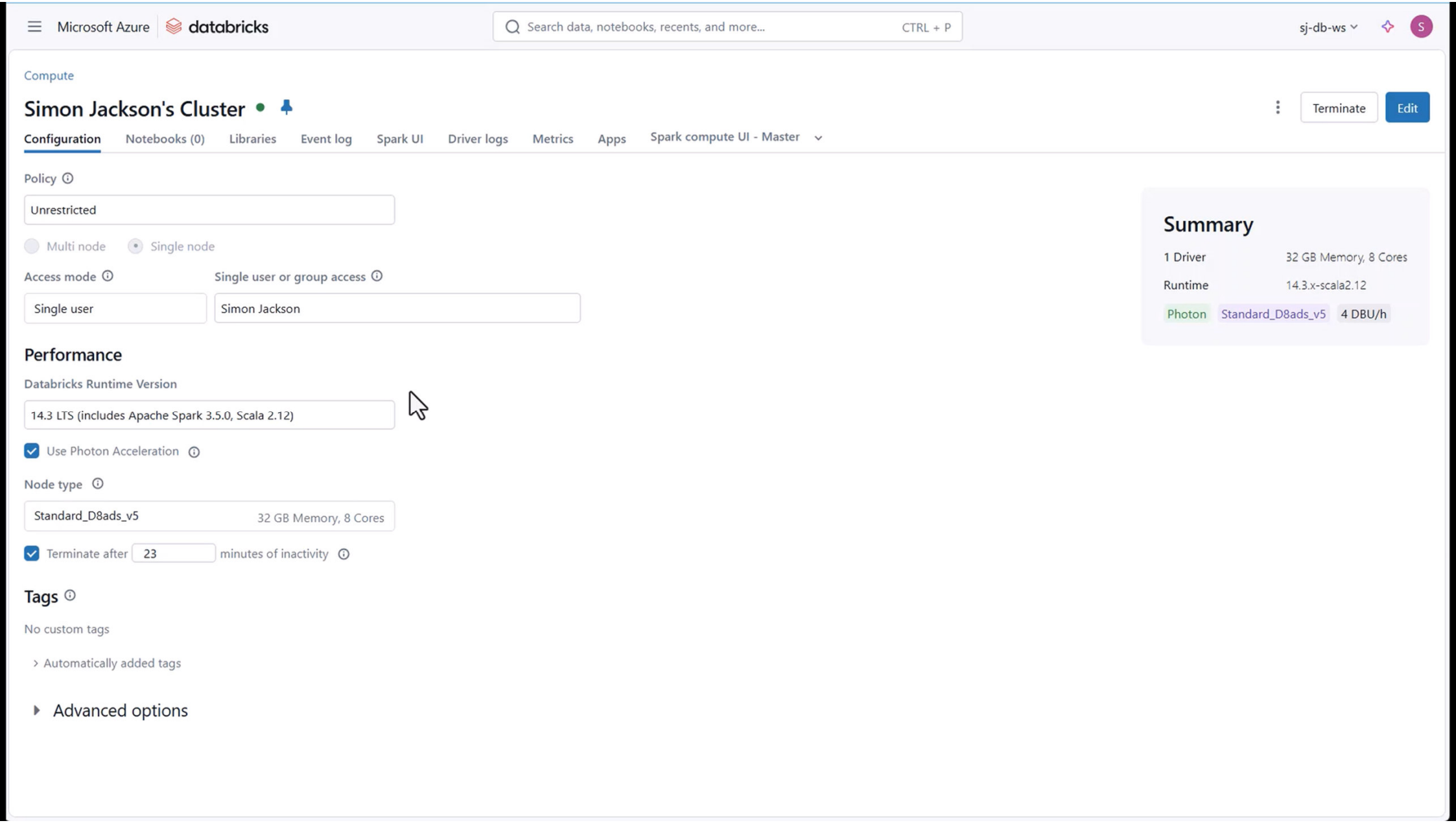


Snapping GPS data to roads



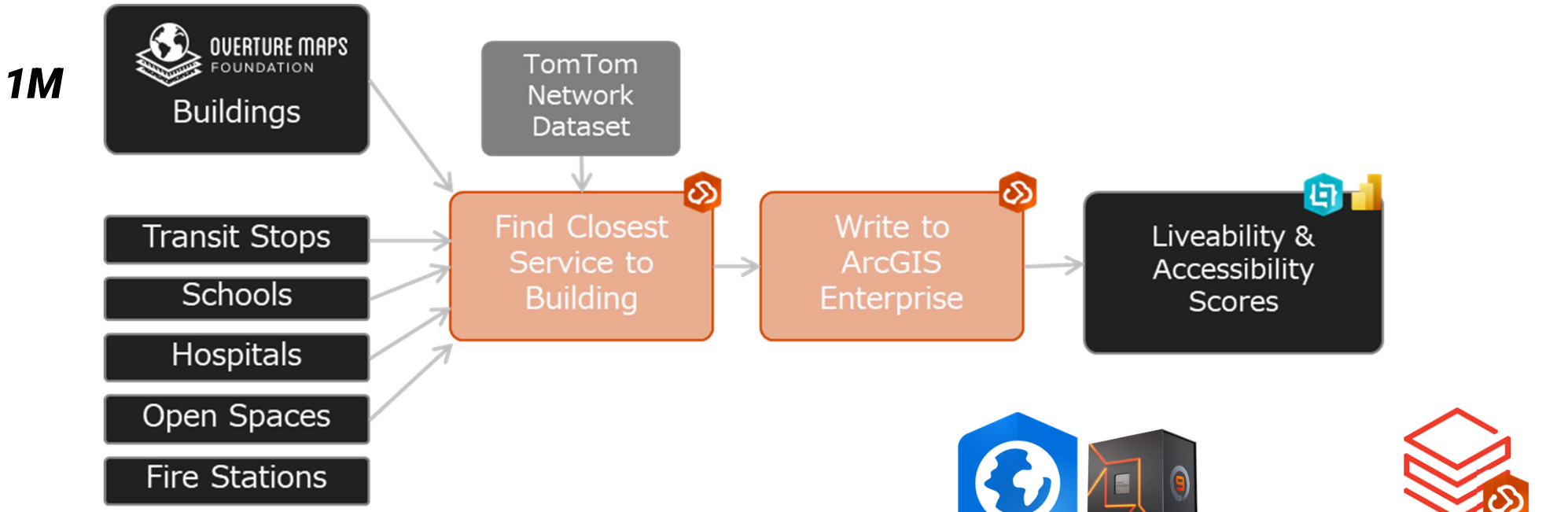
Time-stepped aggregations

- *Aggregate Points*
- *Calculate density*
- *Calculate field*
- *Calculate motion statistics*
- *Clip Layer*
- *Create routes*
- *Create service areas*
- *Detect incidents*
- *Find closest facilities*
- *Find dwell locations*
- *Find hot spots*
- *Find point clusters*
- *Find similar locations*
- *Generate OD matrix*
- *Geocode*
- *Geographically weighted regression*
- *Group by proximity*
- *Nearest neighbors*
- *Overlay*
- *Reconstruct tracks*
- *Reverse geocode*
- *Snap tracks*
- *Spatiotemporal join*
- *Summarize within*
- *Trace proximity events*


Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
sj-db-ws
Compute
Simon Jackson's Cluster
Terminate
Edit
Configuration
Notebooks (0)
Libraries
Event log
Spark UI
Driver logs
Metrics
Apps
Spark compute UI - Master
Policy
Unrestricted
Multi node
Single node
Access mode
Single user or group access
Single user
Simon Jackson
Performance
Databricks Runtime Version
14.3 LTS (includes Apache Spark 3.5.0, Scala 2.12)
Use Photon Acceleration
Node type
Standard_D8ads_v5
32 GB Memory, 8 Cores
Terminate after
23
minutes of inactivity
Tags
No custom tags
Automatically added tags
Advanced options
Summary
1 Driver
32 GB Memory, 8 Cores
Runtime
14.3.x-scala2.12
Photon
Standard_D8ads_v5
4 DBU/h

How accessible are transit stops in Greater Sydney?

ArcGIS brings scalable spatial analytics into Databricks



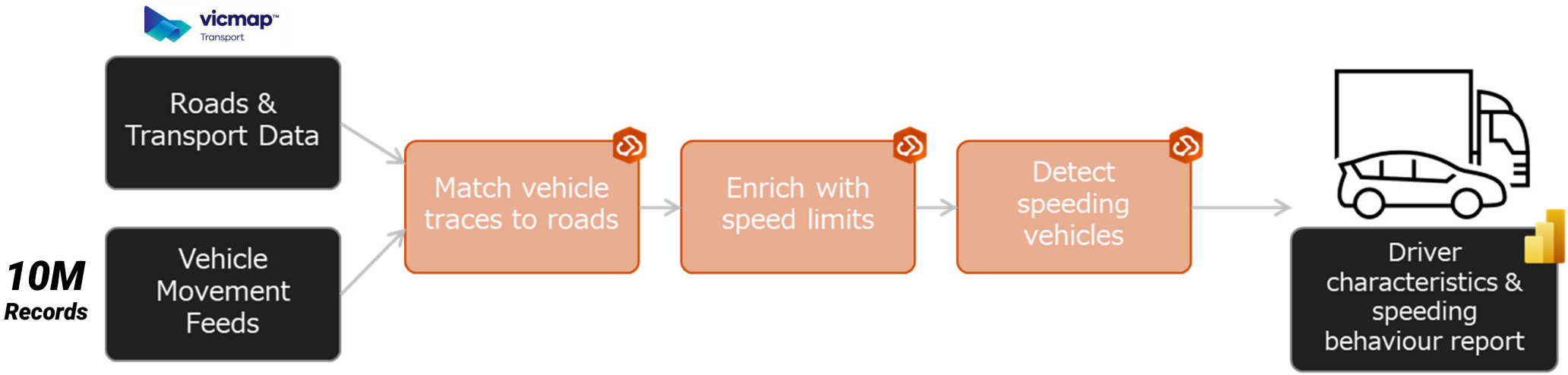
ArcGIS Pro
16 Cores CPU
64GB RAM

11 Hours

Databricks
Single-node
32 Cores CPU
128GB RAM

2 hours

Detecting speeding incidents from VPS Data

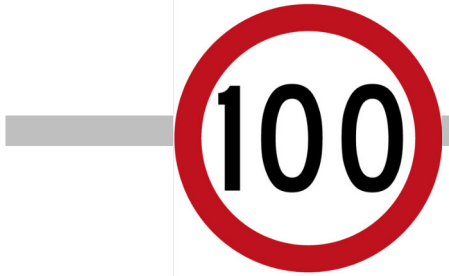


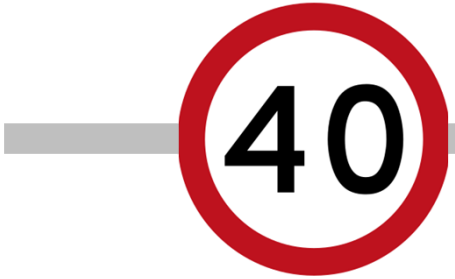
Detecting speeding incidents from VPS Data

Speed & location from Vehicle Sensor

Speed limit data

Incident?

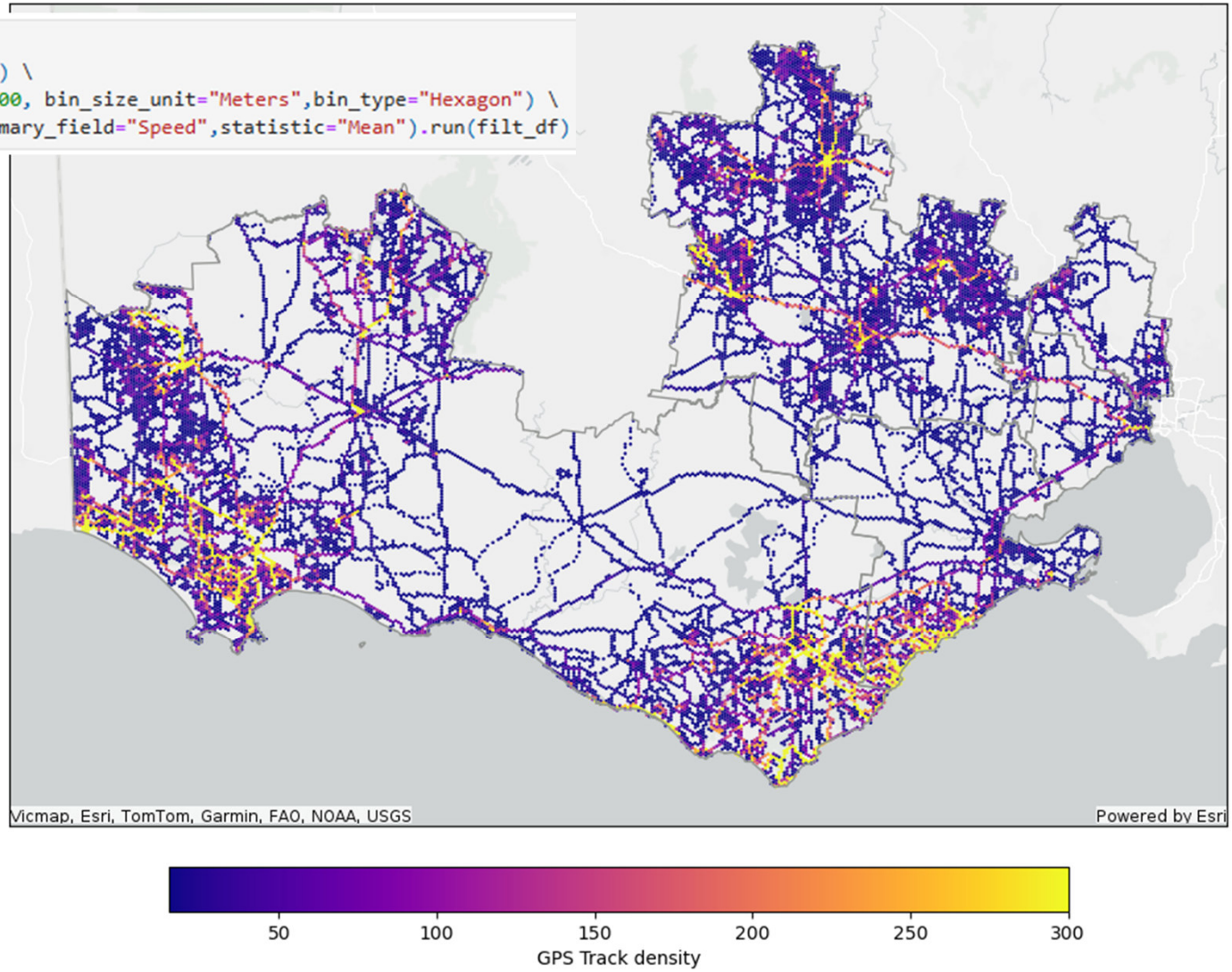




Detecting speeding incidents from VPS Data

```
# Aggregate data into bins
result_bins = AggregatePoints() \
          .setBins(bin_size=1000, bin_size_unit="Meters",bin_type="Hexagon") \
          .addSummaryField(summary_field="Speed",statistic="Mean").run(filt_df)
```

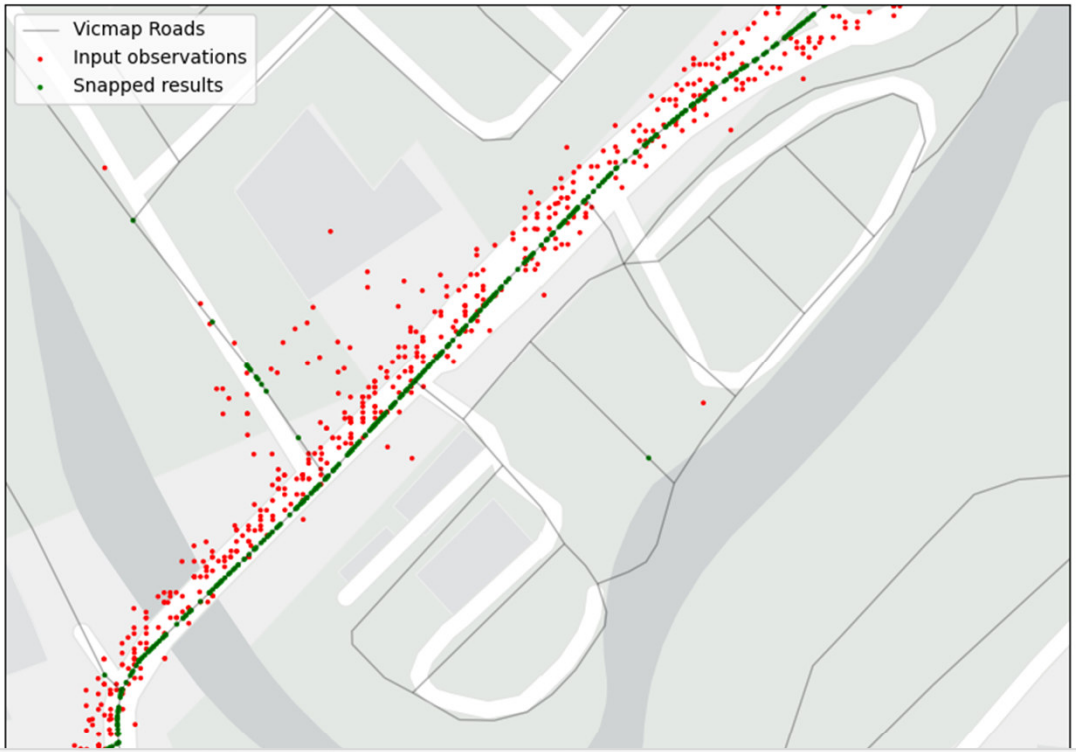
Vehicle movements - 1km aggregation in VicGrid GDA2020



Detecting speeding incidents from VPS Data

Map Matching

- *Uses direction attributes*
- *Cleanses data*
- *Removes GPS Drift*



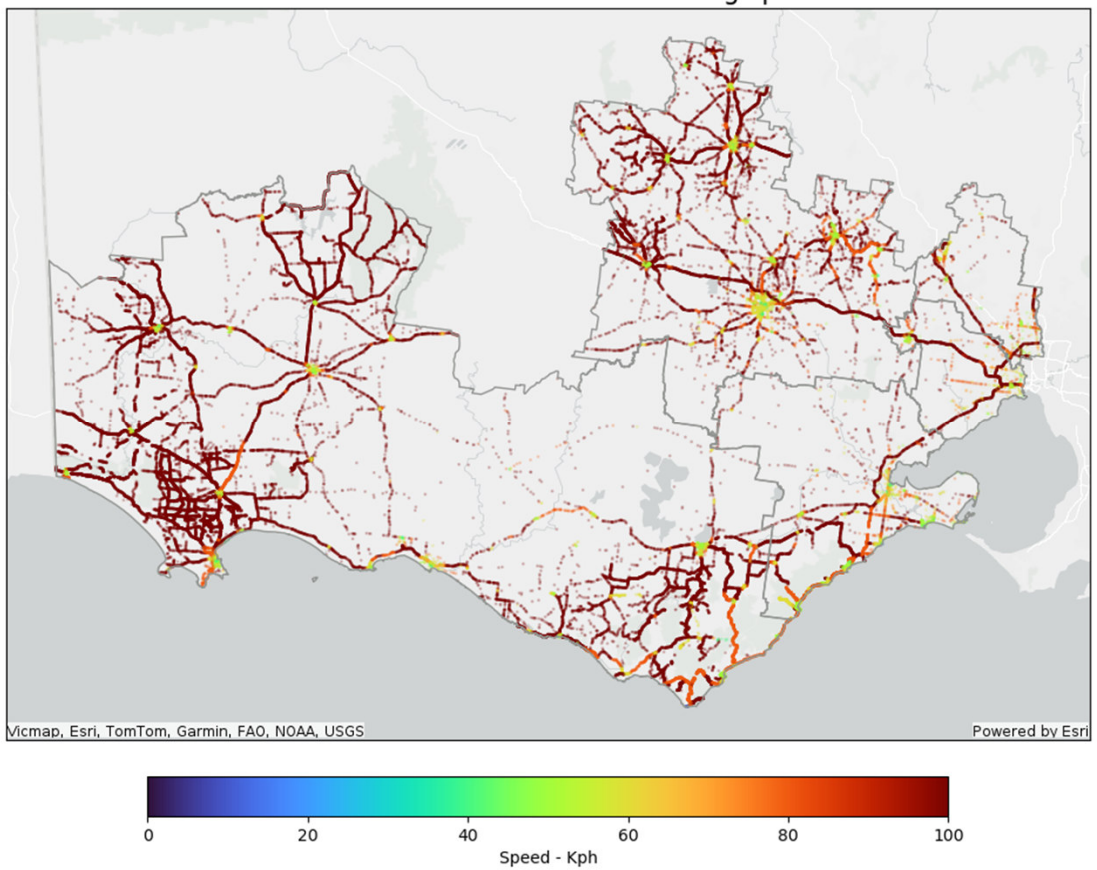
vicmap™ Transport

```
# Snap tracks using the device ID as the track identifier and snap within 100 meters
vps_snap = SnapTracks() \
            .setTrackFields("Designator") \
            .setSearchDistance(search_distance=50, search_distance_unit="Meters") \
            .setDistanceMethod(distance_method="Planar") \
            .setConnectivityFields(from_node="from_ufi", to_node="to_ufi") \
            .setDirectionFieldMatching(direction_field="direction_code", forward_value="F", backward_value="R", both_value="B") \
            .setOutputMode(output_mode="MatchedPoints") \
            .run(filt_df, vmt_df)
```

Detecting speeding incidents from VPS Data

Enrich Vehicle Positioning points with road speed limits

Vehicle Position data enriched with intersecting speed zones data



Enrich Vehicle Position data with underlying Speed Limit from roads

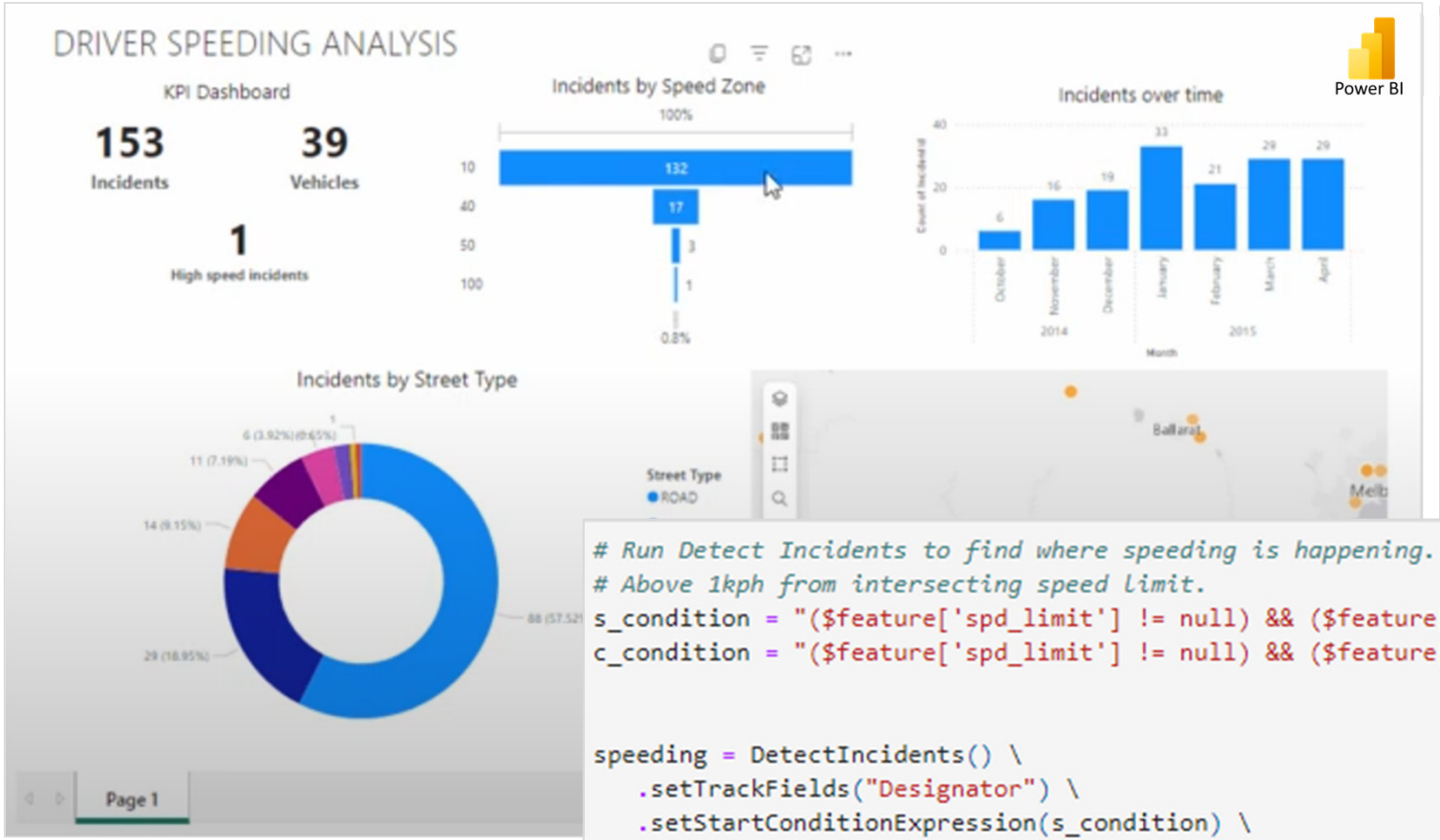
```
# Get Geoscape Roads from Digital Atlas Australia
gs_url = "https://services-ap1.arcgis.com/ypkPEy1AmwPKGNNv/arcgis/rest/services/Nati
gs_df = spark.read.format("feature-service").load(gs_url)\
    .filter("state == 'VIC'")\
    .withColumn("shape", ST.transform("shape", 7899)) # Transform to VicGrid GDA2020
gs_df = gs_df.withColumnRenamed("speed", "spd_limit")
```

```
StatementMeta(, 954eb665-5cd0-447b-89da-a6f4e53a9adc, 81, Finished, Available, Finishe
```

```
# Use nearest neighbour tool to enrich movement data with the nearest roads speed zo
nn_df = NearestNeighbors()\
    .setNumNeighbors(1)\
    .setSearchDistance(5, "Meters")\
    .setResultLayout("long")\
    .run(vps_snap, gs_df)
```

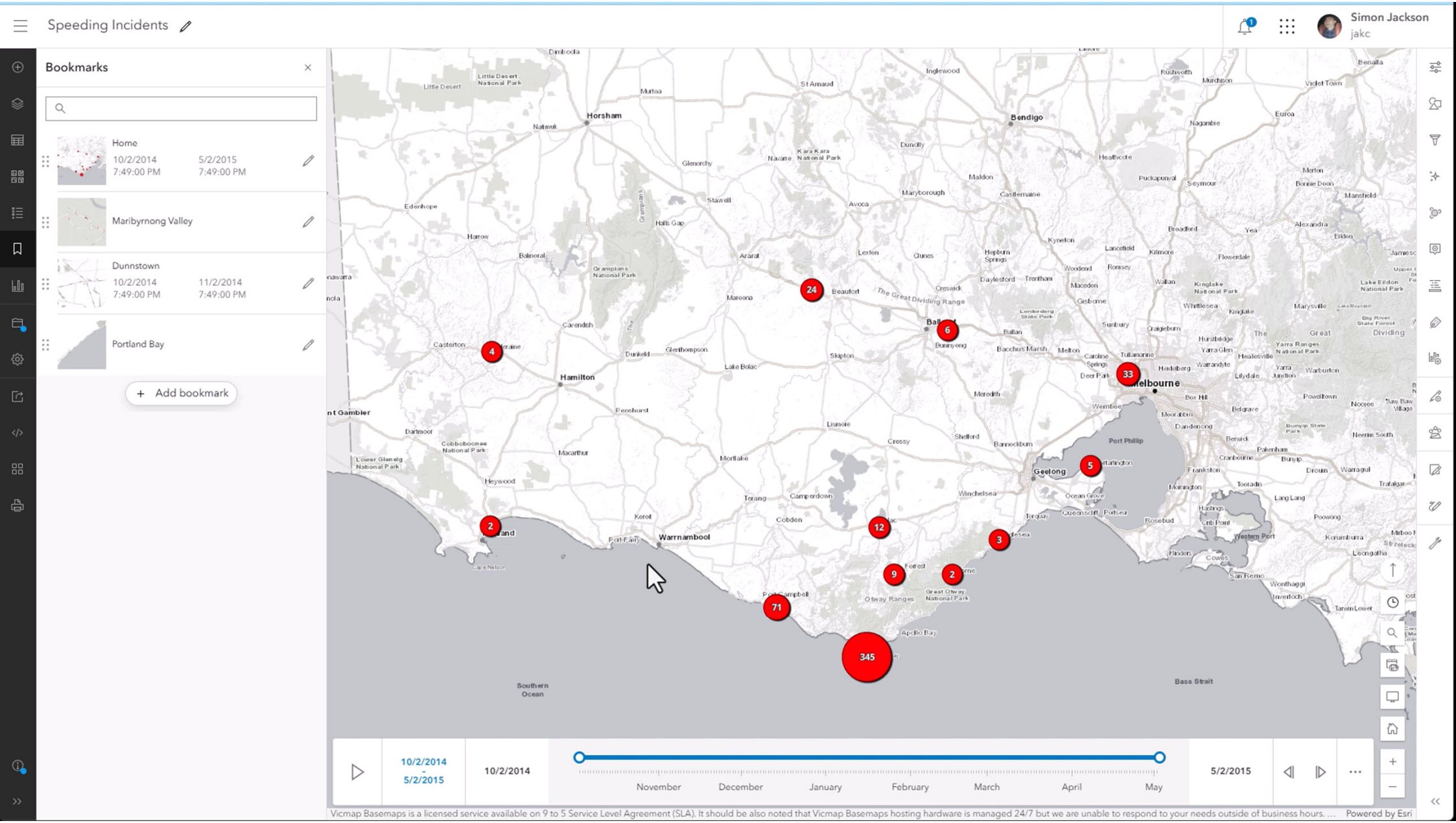
Detecting speeding incidents from Victorian Government VPS Data

Speeding Incidents with an Incident Detector




```
# Run Detect Incidents to find where speeding is happening.
# Above 1kph from intersecting speed limit.
s_condition = "($feature['spd_limit'] != null) && ($feature['speed'] > $feature['spd_limit'] + 5)"
c_condition = "($feature['spd_limit'] != null) && ($feature['speed'] < $feature['spd_limit'] - 1)"


speeding = DetectIncidents() \
    .setTrackFields("Designator") \
    .setStartConditionExpression(s_condition) \
    .setEndConditionExpression(c_condition) \
    .setOutputMode("Incidents") \
    .run(dataframe=nn_df)
```


Speeding Incidents
Simon Jackson
jakc
Bookmarks
Home
10/2/2014 7:49:00 PM
5/2/2015 7:49:00 PM
Maribyrnong Valley
Dunnstown
10/2/2014 7:49:00 PM
11/2/2014 7:49:00 PM
Portland Bay
+ Add bookmark
Horsham
Bendigo
Ballarat
Hamilton
Warrnambool
Geelong
Melbourne
Port Phillip
Bass Strait
Southern Ocean
Great Ocean Road
Great Otway National Park
Grampians National Park
24
6
4
33
5
12
3
2
9
2
71
345
10/2/2014 - 5/2/2015
10/2/2014
November
December
January
February
March
April
May
5/2/2015
Vicmap Basemaps is a licensed service available on 9 to 5 Service Level Agreement (SLA). It should be also noted that Vicmap Basemaps hosting hardware is managed 24/7 but we are unable to respond to your needs outside of business hours. ...
Powered by Esri

Wealth of use cases for Spatial Analytics in Databricks

PREDICTIVE MAINTENANCE
Predict where issues will occur with your assets

MINING
Fuel consumption monitoring across sites, outlier analysis

WATER UTILITY
Work Order Optimisation - proximity based clustering & efficient routing

RAIL
High level reporting of assets and incidents using linear measure values

LAW ENFORCEMENT
Identify patterns in call histories.

INSURANCE
Calculating exposure to cyclone risk / Flood risk modelling

ELECTRIC UTILITY
Demand forecasting from smart meter & weather data

FINANCE
Fraud detection in transactions.
Incident detection using spatiotemporal joins

LOGISTICS EFFICIENCY
Identifying where problems are in supply chain

HEALTH & PARTNERS
Processing human movement data for scenario modelling