

# Single view of ... *Everything*

Streaming 4.4 billion events  
with MongoDB & Apache Kafka



Simon Aubury

Principal Data  
Engineer Architect

# *Where*

Are these guys from?





scio



**Australia**

swann  
insurance



**New Zealand**

swann  
insurance



ASURANSI  
PAROLAMAS



**Asia**





# *Why*

Build a single customer view?

iag



# We've a lot of data

## Context #1

mvyear	mvmake	mvmodel	mvbody
1886	RUDGE	PENNY FARTHING	MBIKE
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1900	MERCEDES		CONVT



# We have a lot of systems

Context #2



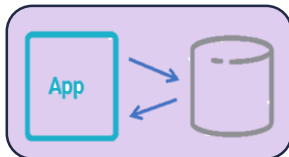
Single View



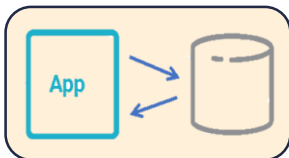
Digital Channels



Vehicles Data



Customer Data



Policy Data



Modelling & Scoring

# We want to tie it together

## Context #3

The collage features three digital interfaces:

- CGU Website:** Displays a green navigation bar with categories like 'Car & Vehicle', 'Home & Contents', 'Travel', 'Lifestyle', and 'Business'. The main content area has a large headline: **BACKING AMBITIOUS AUSSIES FOR OVER 165 YEARS.** Below this, it says 'It's our ambition to help you live yours.' and includes a 'Get a quote' button.
- NRMA Insurance Website:** Features a white header with the NRMA logo and navigation links for 'Insurance', 'Claims', 'Renewals & payments', and 'Contact us'. A prominent orange banner asks, 'Have you been affected by the recent storm?' and provides a link to the emergency page. Below the banner are buttons for 'Get a quote' and 'Retrieve a quote'. Further down, there are three service boxes: 'Loyalty Discount', 'Making a claim is easy', and 'Existing customer'. The bottom section shows a video player with the text 'Help is not something we do. Help is who we are.' and a 'Watch the film' button.
- Mobile App:** A smartphone screen showing a user interface for '2014 TOYOTA AURION MCL16N'. It displays details for 'Comprehensive Car Insurance', including the policy expiry date (13 Mar 2018) and the premium amount (\$751.09). It also shows a 'Home Buildings' section with a policy expiry date (13 Jun 2017) and a premium amount (\$997.98). The bottom navigation bar includes icons for 'Home', 'Policies', 'Claims', and 'Details'.



# Focused Insights

Simon Aubury  
Customer Key

**Profile**

Basic Profile

Title MR  
First Name SIMON  
Middle Names B  
Last Name AUBURY  
Aliases not collected  
Date of Birth [redacted]

Best Contact

Home Phone not collected  
Work Phone not collected  
Mobile Phone [redacted]  
Email [redacted]

Relationships

Same Address [redacted]  
Same Address [redacted]  
Inferred Family [redacted]

Addresses

Map Satellite

3 addresses

LOT 308 60 D [redacted] AVENUE, WEST RYDE NSW 2114

Profile  
Products  
Vehicles  
Customer Health  
Marketing  
Communication History

Activity timeline

2017

Nov Took out CTP insurance on a 2017 HOLDEN BARINA

2016

Nov Took out comprehensive insurance on a 2016 HOLDEN BARINA

Invalid date

Invalid date Lodged a Collision claim on a 2016 HOLDEN BARINA

2019

Jan Took out CTP insurance on a 2009 HOLDEN BARINA

Property Factors

Property Attributes

Land size 590m<sup>2</sup>  
Slope 1.80  
Elevation 36m AHD W  
Aspect 247° S

Nearest pub/bar/club 1.50 mins

Environmental Factors

Drive time to points of interest

Nearest education 2.50 mins  
Nearest emergency 3.30 mins  
Nearest shop 3.30 mins  
Train station 3.30 mins  
Nearest pub/bar/club 1.50 mins  
State NSW  
Country Australia

Peril Risk  
Flood Risk N

Profile  
Environmental Factors  
Property Factors

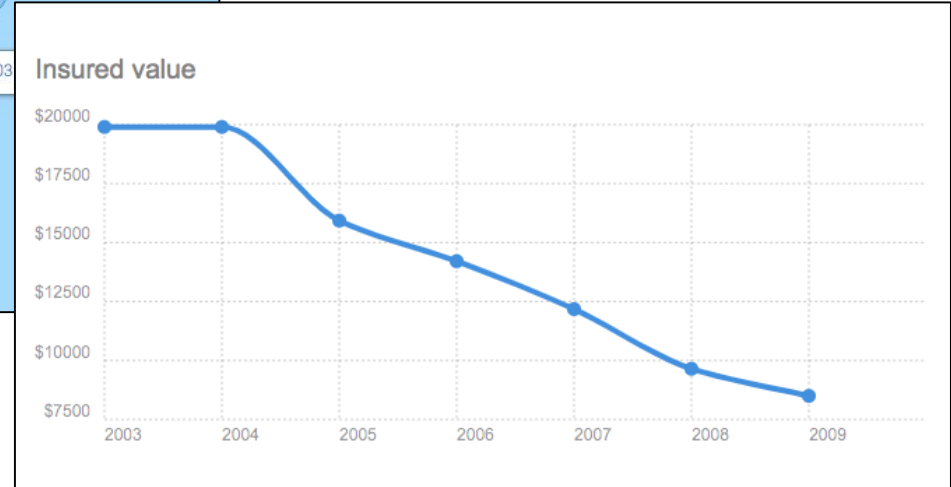
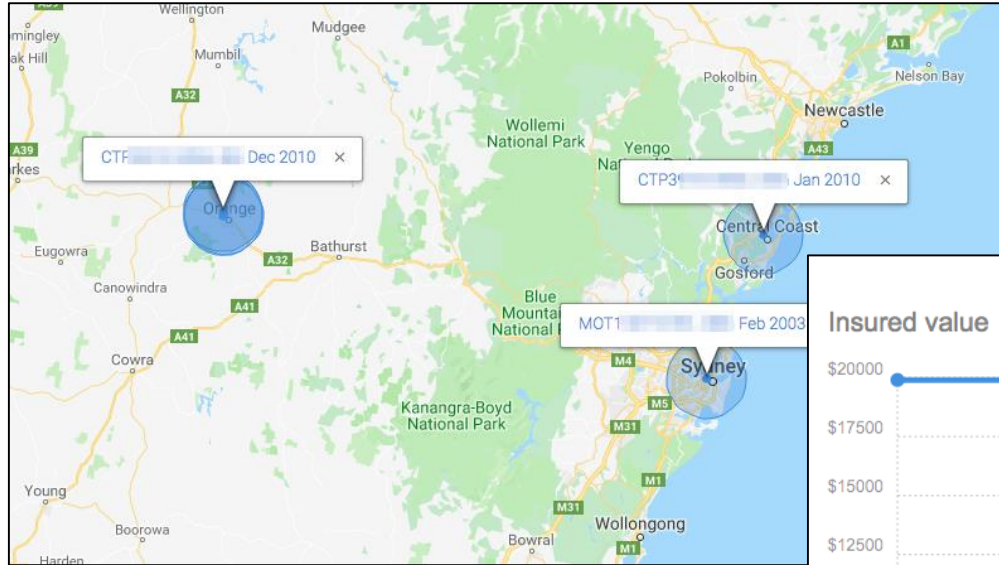
2016 HONDA CR-V

NSW  
Nov 2017

18-11-2016: [redacted] 3525  
BILLING CLIENT, DRIVER, INSURED, MAIL ADDRESS



# Refocused Data



# *How*

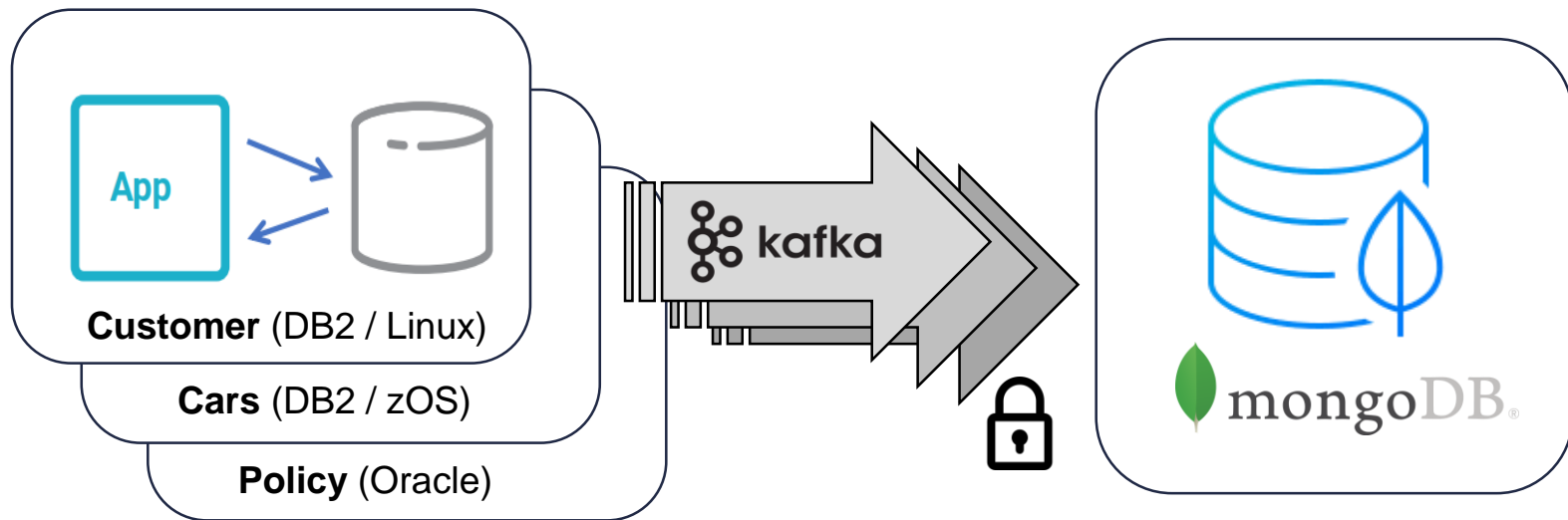
Did we build this?

iag



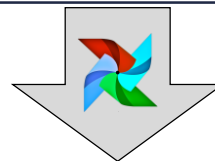
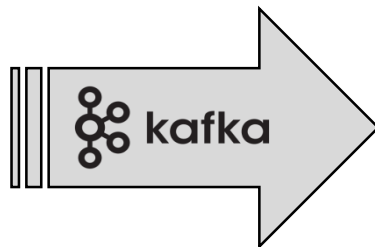
# System of record acquisition

## Architecture of capture



# Insight Generation

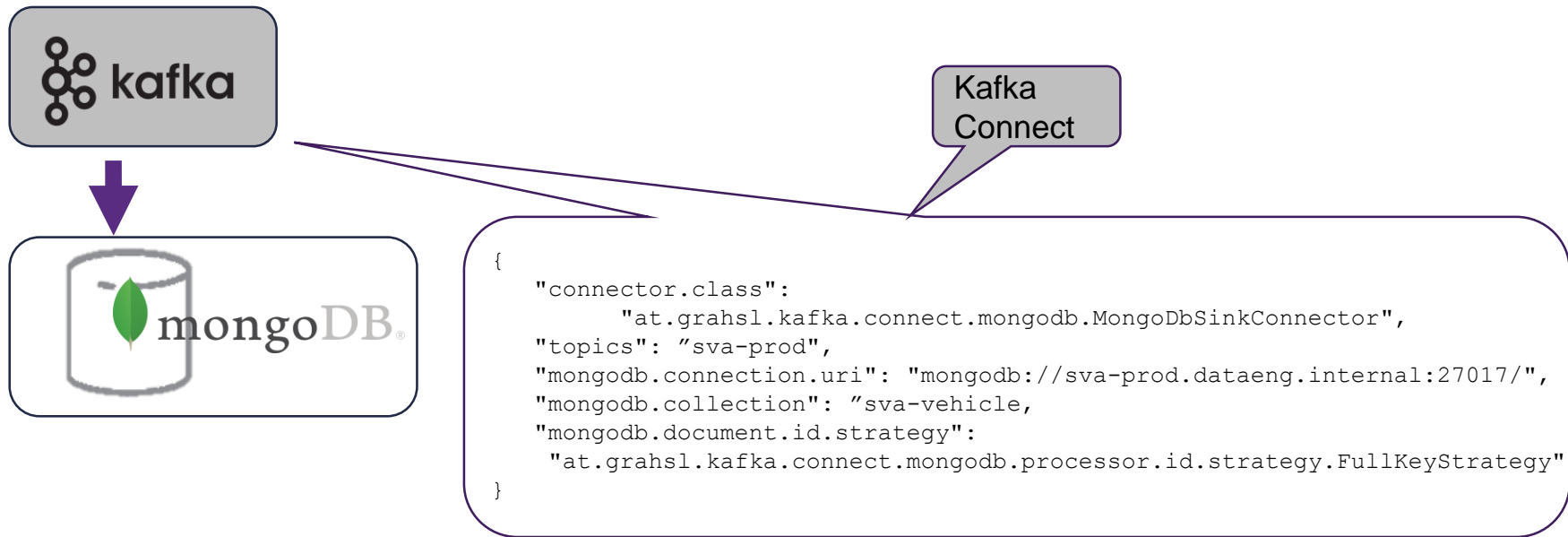
## Architecture of insights





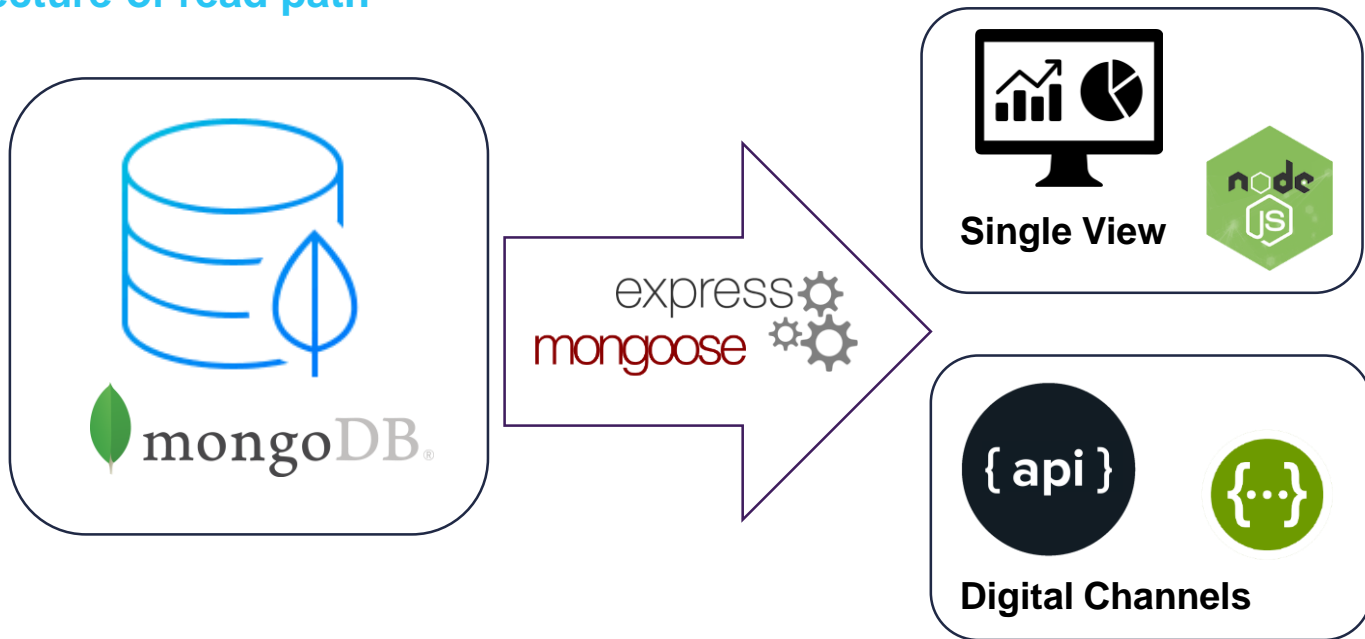
# Kafka Connect Sink

## Architecture of write path



# Serving Tier

Architecture of read path



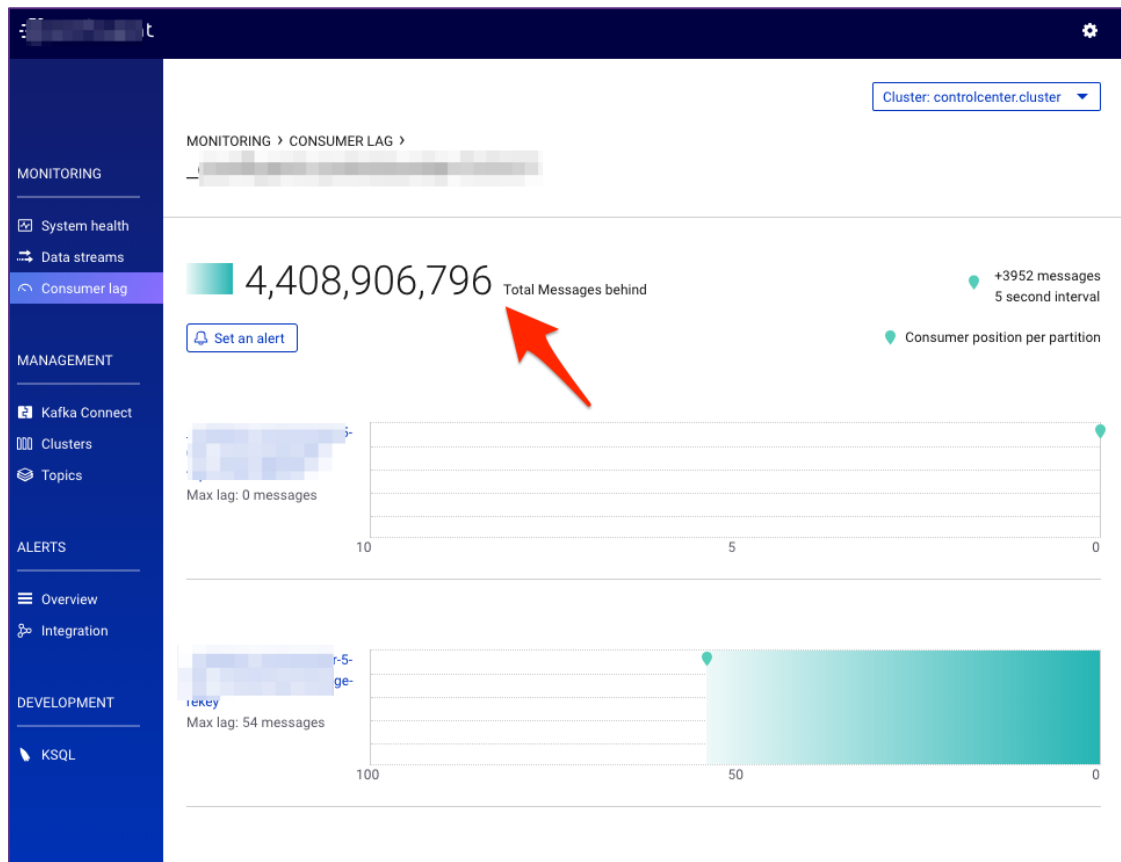
# What did we discover?

Slow to fast ... to *really* fast!



# Challenge

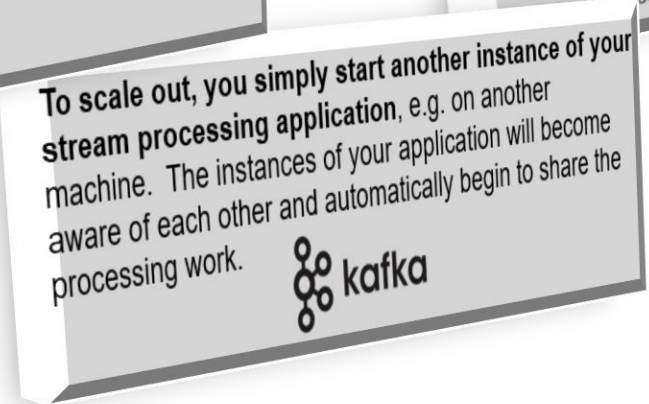
Lots of data





# Horizontal scaling?

## Theory



1. <https://docs.mongodb.com/manual/sharding/>
2. <https://kubernetes.io/docs/tutorials/kubernetes-basics/scale/scale-intro/>
3. <https://www.confluent.io/blog/elastic-scaling-in-kafka-streams/>

# Reality: Healthy Tech Competition



9:00 AM

Underwhelming ... it's time to use more docker apps.  
600/sec with 35% cpu on stream server



9:17 AM

Scaled



9:24 AM

Ingestion stats on mongo: 2M records. Pretty impressive  
considering the number of queries and index hits ...



10:01 AM

I can bump to 32



9:47 AM

Looking good on the 16 CPUs - Mongo is keeping  
up with Kafka on the ingest now



10:03 AM

Nice - starting to hammer. 40% across all CPU's



10:30 AM

Hah - mongo's winning ! Topic drained



10:46 AM

New record - we're at 5,070 records per second  
18 million / hour

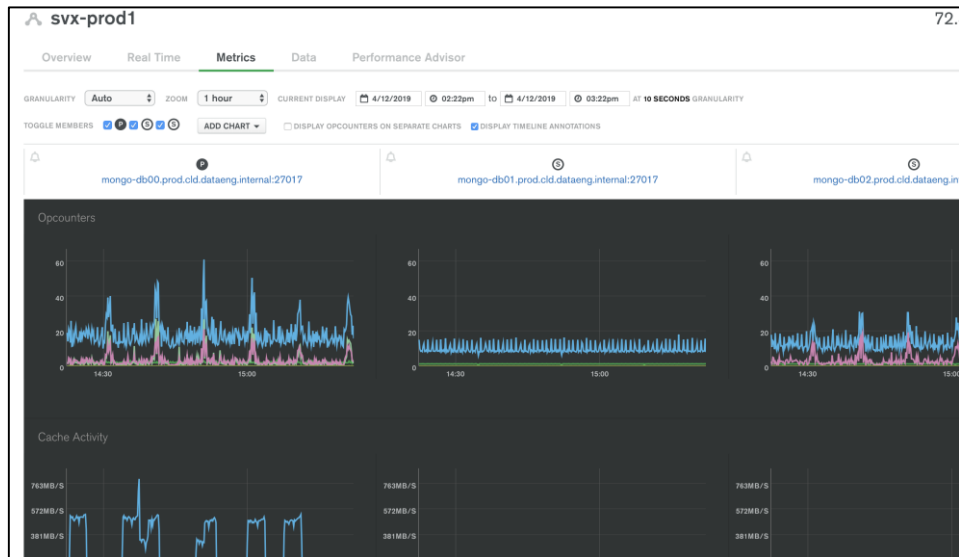
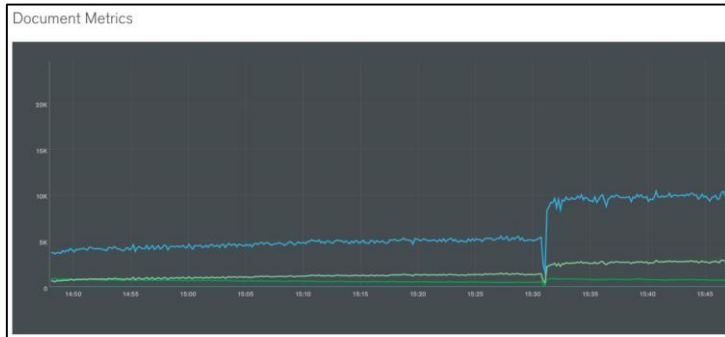
# Horizontal scaling ... meet efficient code

In numbers

30 records / sec



33,500 records / sec



# Who

Participates in feature delivery?

iag





# Feature Sprint

The screenshot shows the JIRA interface for a project named 'Teams in Space'. The left sidebar contains navigation links for Backlog, Active sprints, Releases, Reports, Issues, Components, and Project shortcuts. The main area displays 'All sprints' with a 'Switch sprint' dropdown. Below this, there are two columns: '12 To Do' and '4 In Progress'. The 'To Do' column contains two issues: TIS-37 (When requesting user details the service should return prior trip) and TIS-10 (Bad JSON data coming back from hotel API). The 'In Progress' column contains two issues: TIS-68 (Homepage footer uses an inline style - should use a class) and TIS-17 (Engage Saturn's Rings Resort as a preferred provider). Each issue has a 'SeeSpaceEZ Plus' button and a status indicator.

The screenshot shows the 'svx.policy\_insights' API endpoint. The 'Documents' tab is selected, and a 'FILTER' button is visible. Below the filter, there are buttons for 'INSERT DOCUMENT', 'VIEW', 'LIST', and 'TABLE'. A document is displayed with the following fields:

```
{
  "_id": "Object",
  "svx_policy_key": "HUON-604",
  "cap_product": "POLICY AND PRODUCT",
  "cap_model": "RETENTION OFFER",
  "cap_feature": "PERCENTILE",
  "kafka_id": 1651,
  "svx_policy_key": "HUON-CTP-04",
  "cap_product": "POLICY AND PRODUCT",
  "cap_model": "RETENTION OFFER",
  "cap_feature": "PERCENTILE",
  "feature_value_numeric": 30,
  "feature_value_character": null
}
```

The screenshot shows two API endpoints: 'addresses' and 'customers'. The 'addresses' endpoint has a 'POST' method with a search path. The 'customers' endpoint has several 'GET' methods with search paths and descriptions:

- GET /customers Search for customers
- GET /customers/count Count the number of customers
- GET /customers/{\_id} Find the customer with the specified id
- GET /customers/{\_id}/activities Find activities for the customer
- GET /customers/{\_id}/invitations Find journey invitations for the customer
- GET /customers/{\_id}/policies Find policies for the customer
- GET /customers/{\_id}/relations Find other customers related to this customer
- GET /customers/{\_id}/vehicles Find vehicles for the customer

The screenshot shows the 'Products' API endpoint. The 'Current' tab is selected, and a list of products is displayed. Each product has a 'CURRENT' status and a 'CTP' value. A summary card on the right shows the 'CAP Features' for the selected product, including the expected profit (\$-76.35) and the chance of renewal (81.1%).

Product	CTP
CTP	CTP
CTP	CTP
CTP	CTP
CTP	CTP

# *What*

Does this mean to our customers  
& stakeholders?

iag



# Unlocking Geospatial Analytics

What we couldn't do with our RDBMS

Analytics team

- 15 years of geospatial policy history
- Goal: understand correlations between population growth and policy purchases
- This business problem had been “on the table” and unsolved for over 3 years
- The following analysis generated from Mongo was done fast ... very, very fast

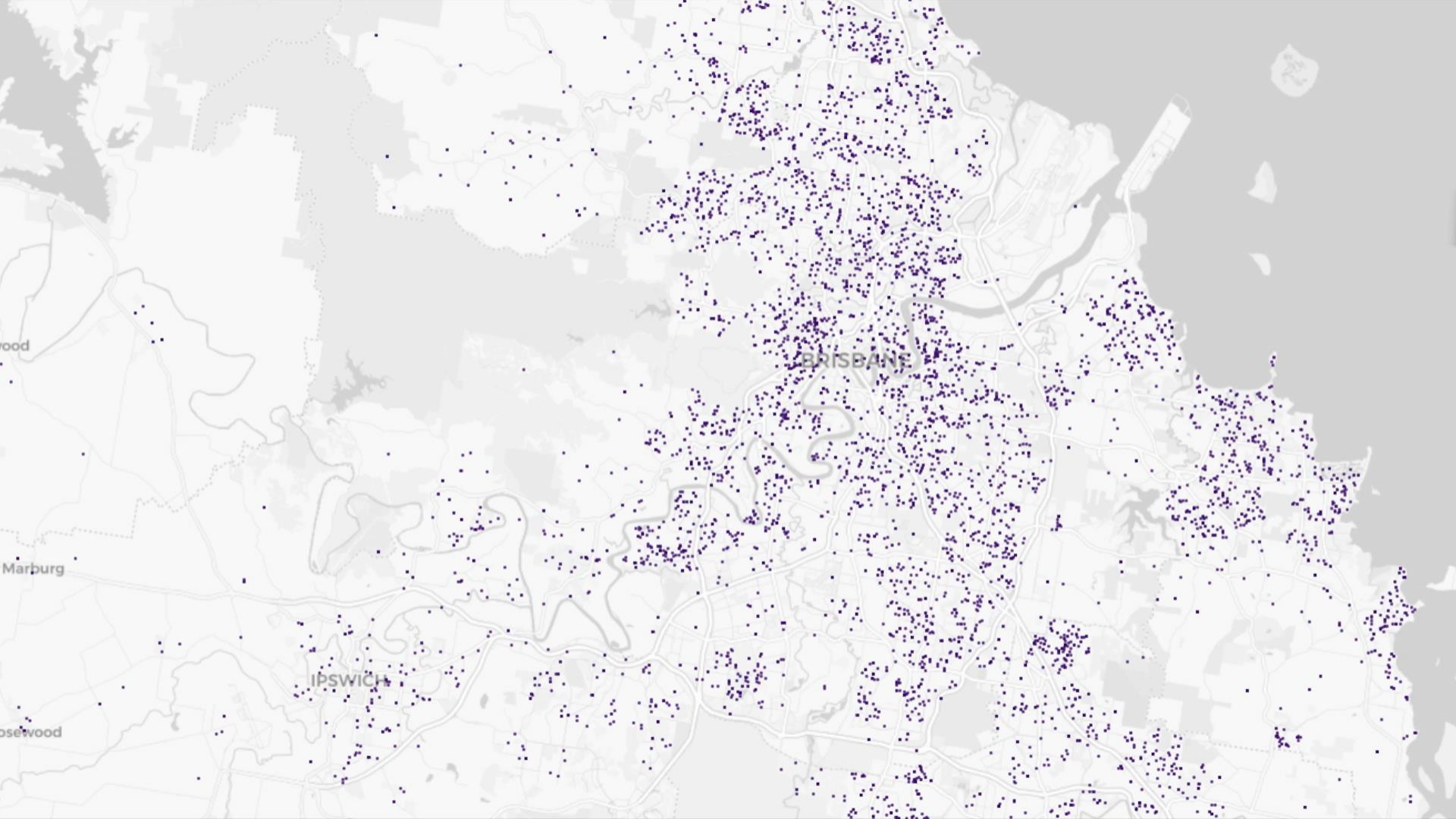


# Geospatial Over Time

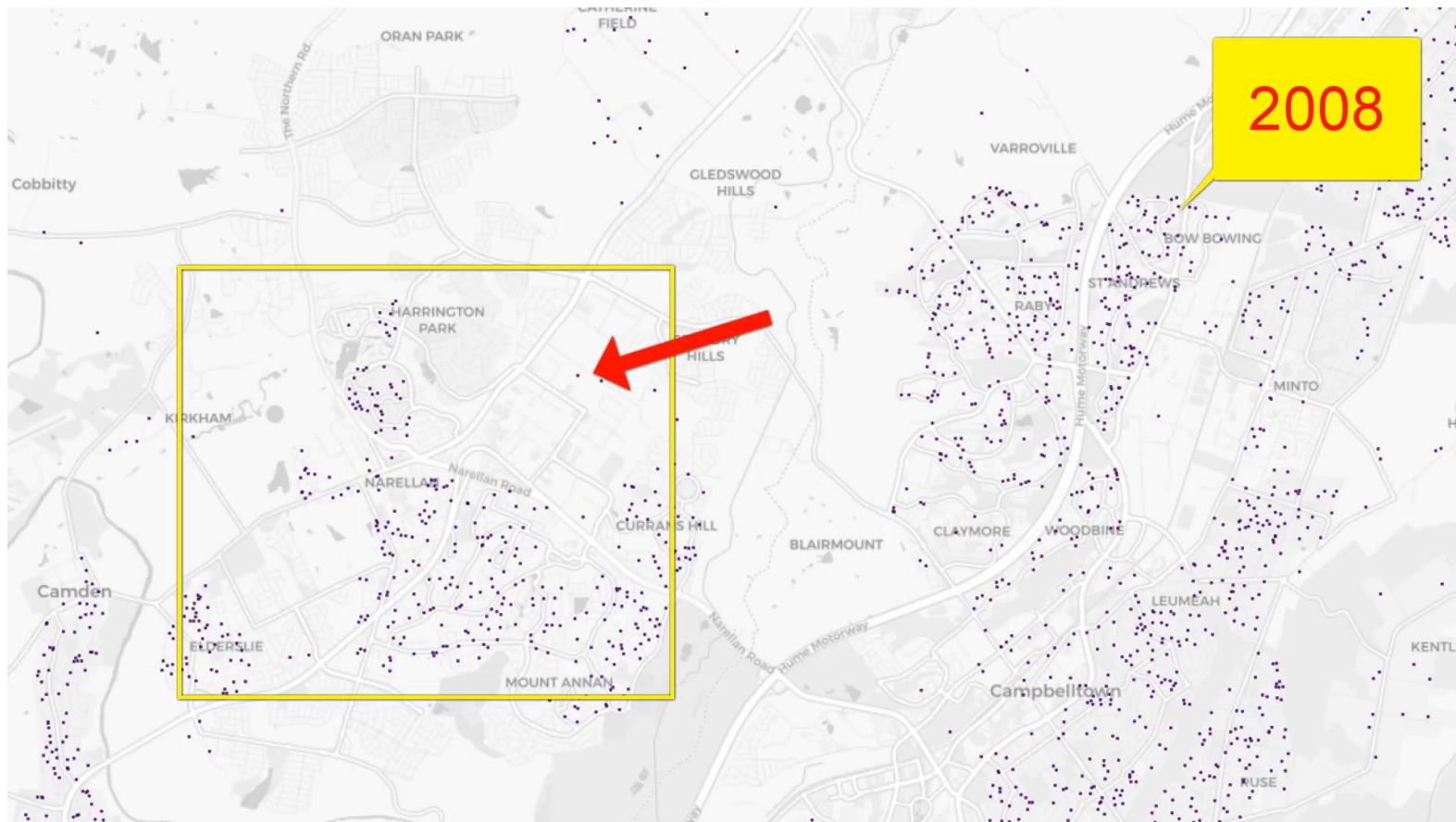
“Creating a duplicated policy map from traditional IAG data systems not only involved complex queries but was error prone due to data quality issues.

SVx and the mongo platform brings the data to a single place and allows easy extraction for multiple different use cases”





# Geospatial Over Time





# Conclusions

## In summary

- Why build a single customer view?
- How did we build this?
- Who participates in feature delivery?
- What does this mean to our customers & stakeholders?



# ***Thank-you!***

Any questions???

**iag**

