Single view of … *Everything*

Streaming 4.4 billion events with MongoDB & Apache Kafka

Simon Aubury

Principal Data
Engineer Architect

Where

Are these guys from?

Australia

NRMA INSURANCE
SGIO
SGIC
CGU
wfi
swann insurance

New Zealand

STATE
ami
N1
nac INSURANCE
LANTERN insurance
swann insurance

Asia

N1
AmAssurance
AKURNIA INSURANS
ASURANSI PAROLAMAS
SBIGeneral INSURANCE
AAA ASSURANCE

Why

Build a single customer view?

We've a lot of data

Context #1

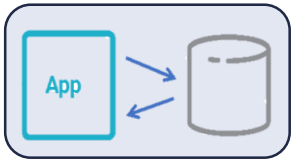
mvyear	mvmake	mvmodel	mvbody
1886	RUDGE	PENNY FARTHING	MBIKE
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1896	FORD	QUADRICYCLE	CONVT
1900	MERCEDES		CONVT

We have a lot of systems

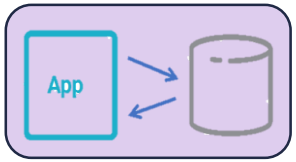
Context #2

Single View

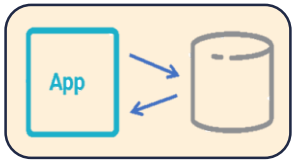
Digital Channels



Vehicles Data



Customer Data

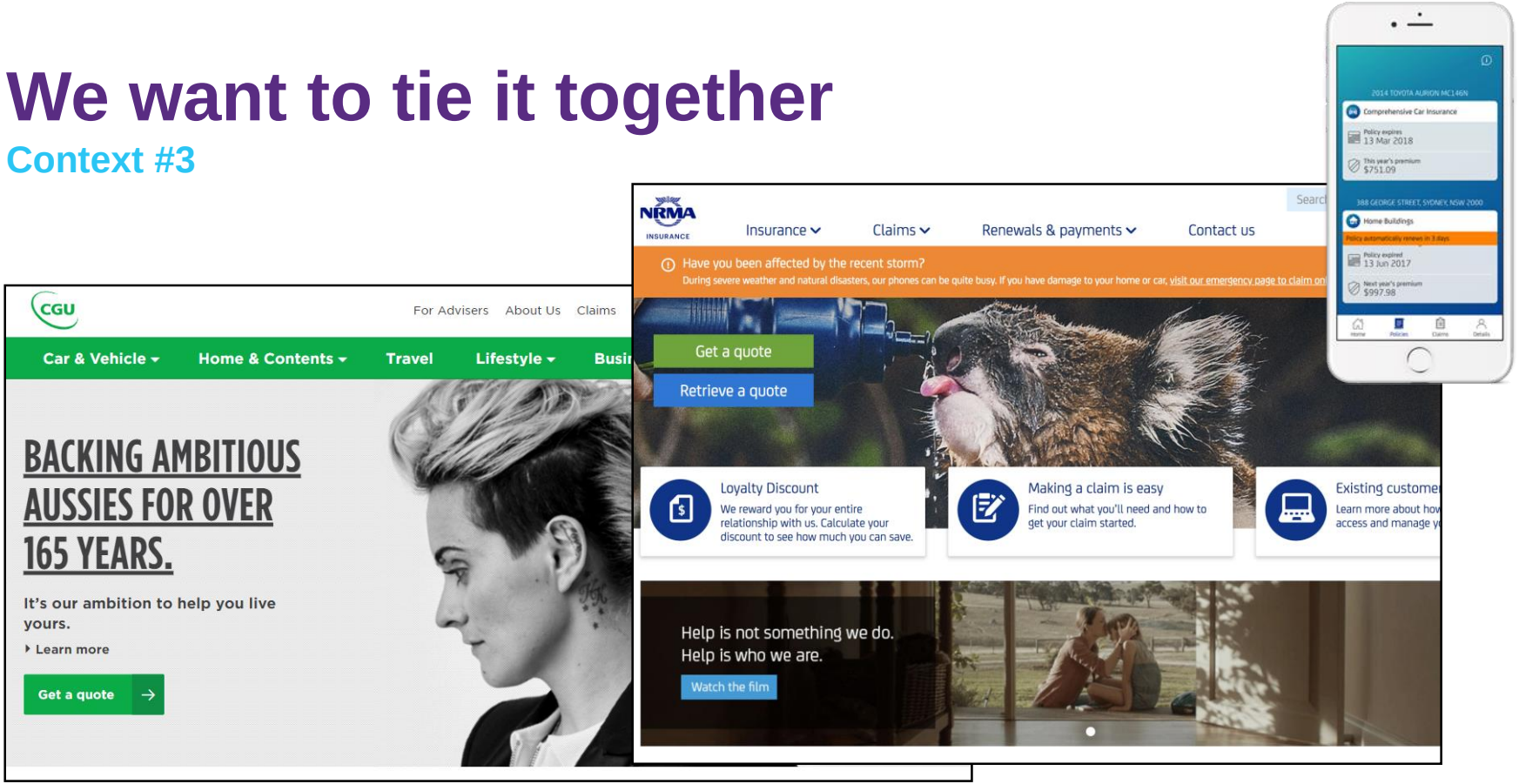


Policy Data

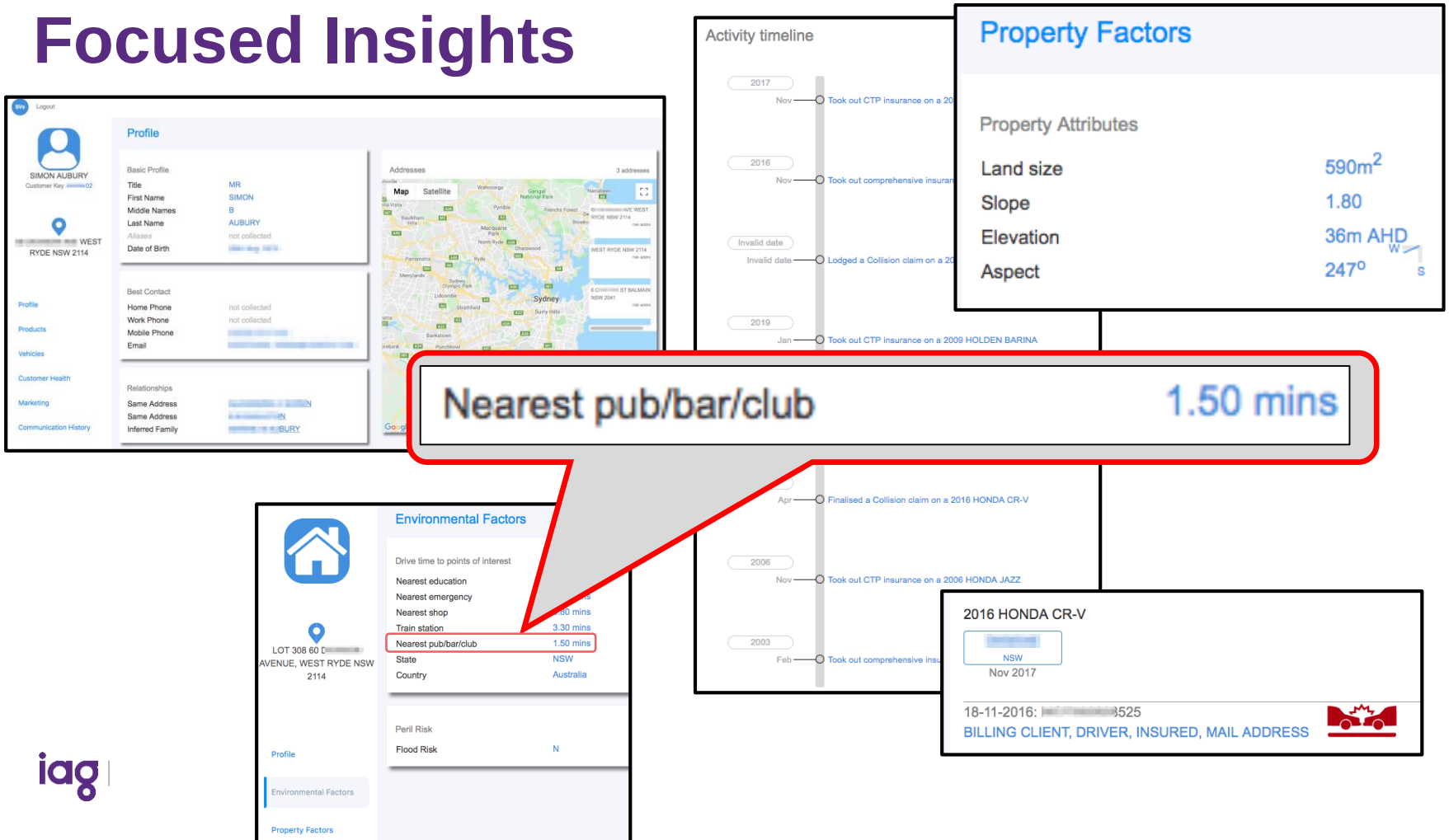
Modelling & Scoring

We want to tie it together

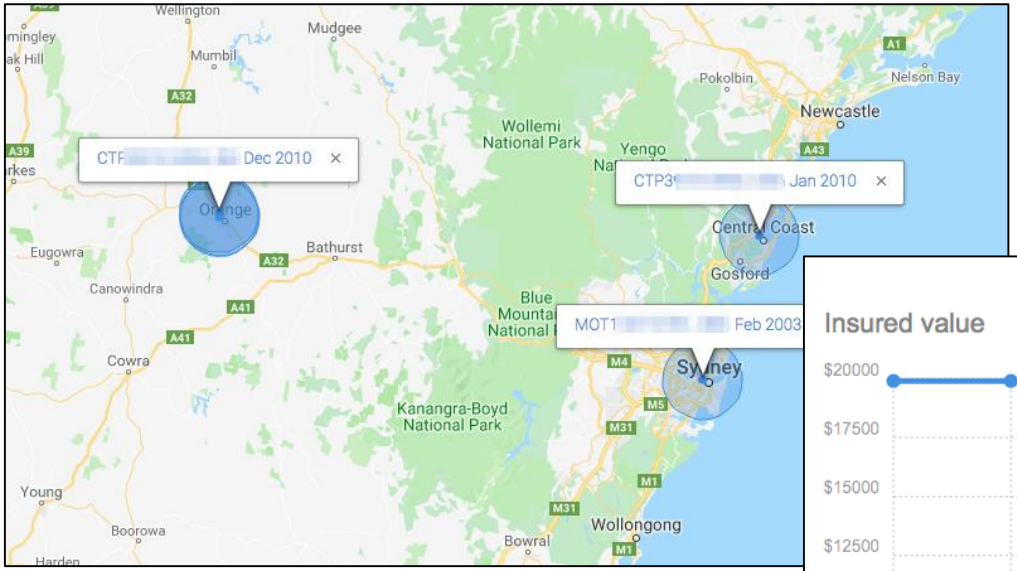
Context #3

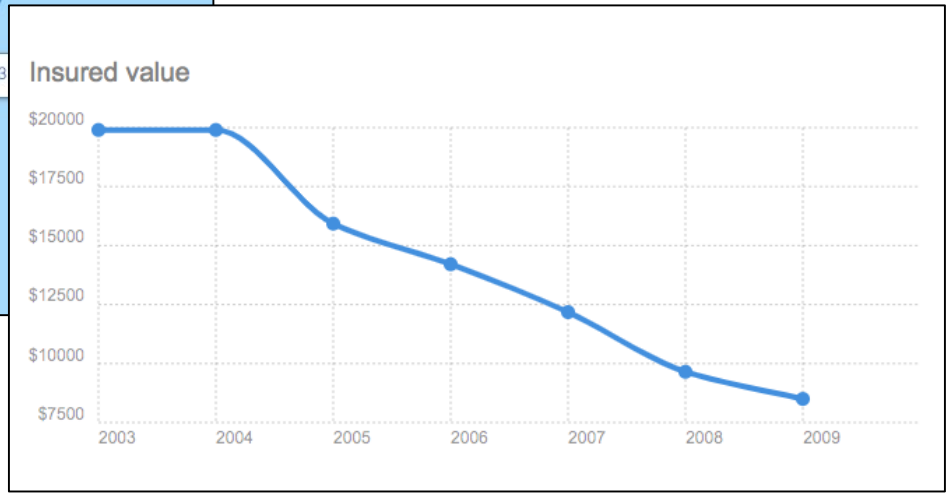


Focused Insights



Refocused Data



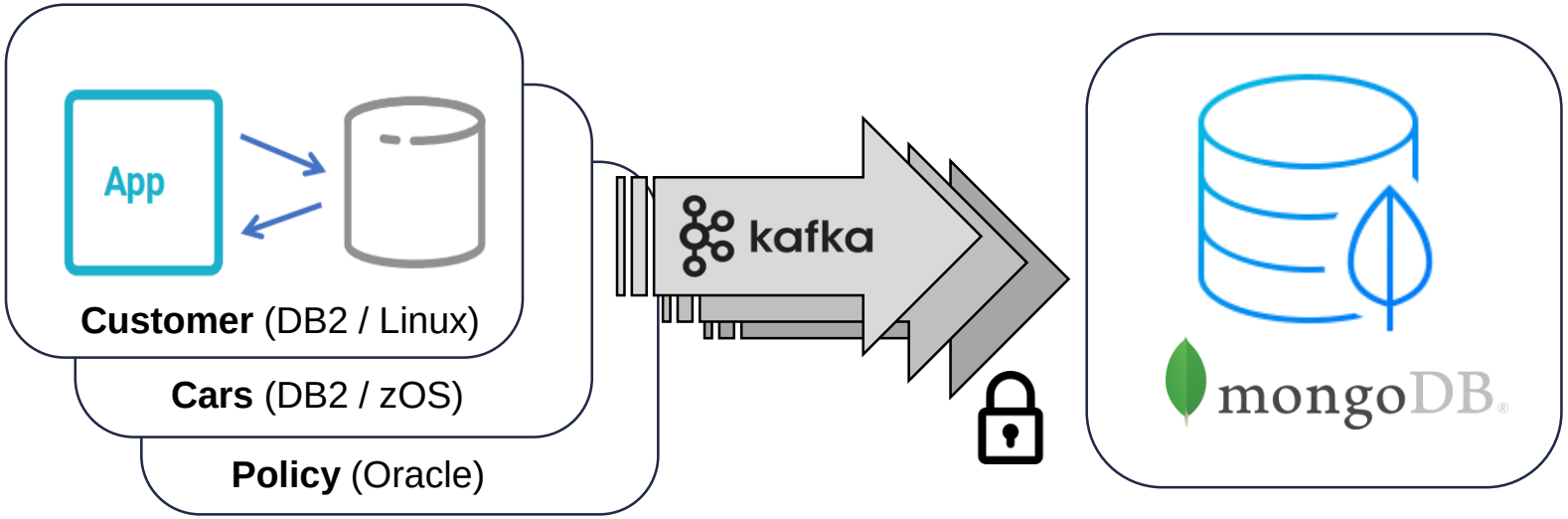


How

Did we build this?

System of record acquisition

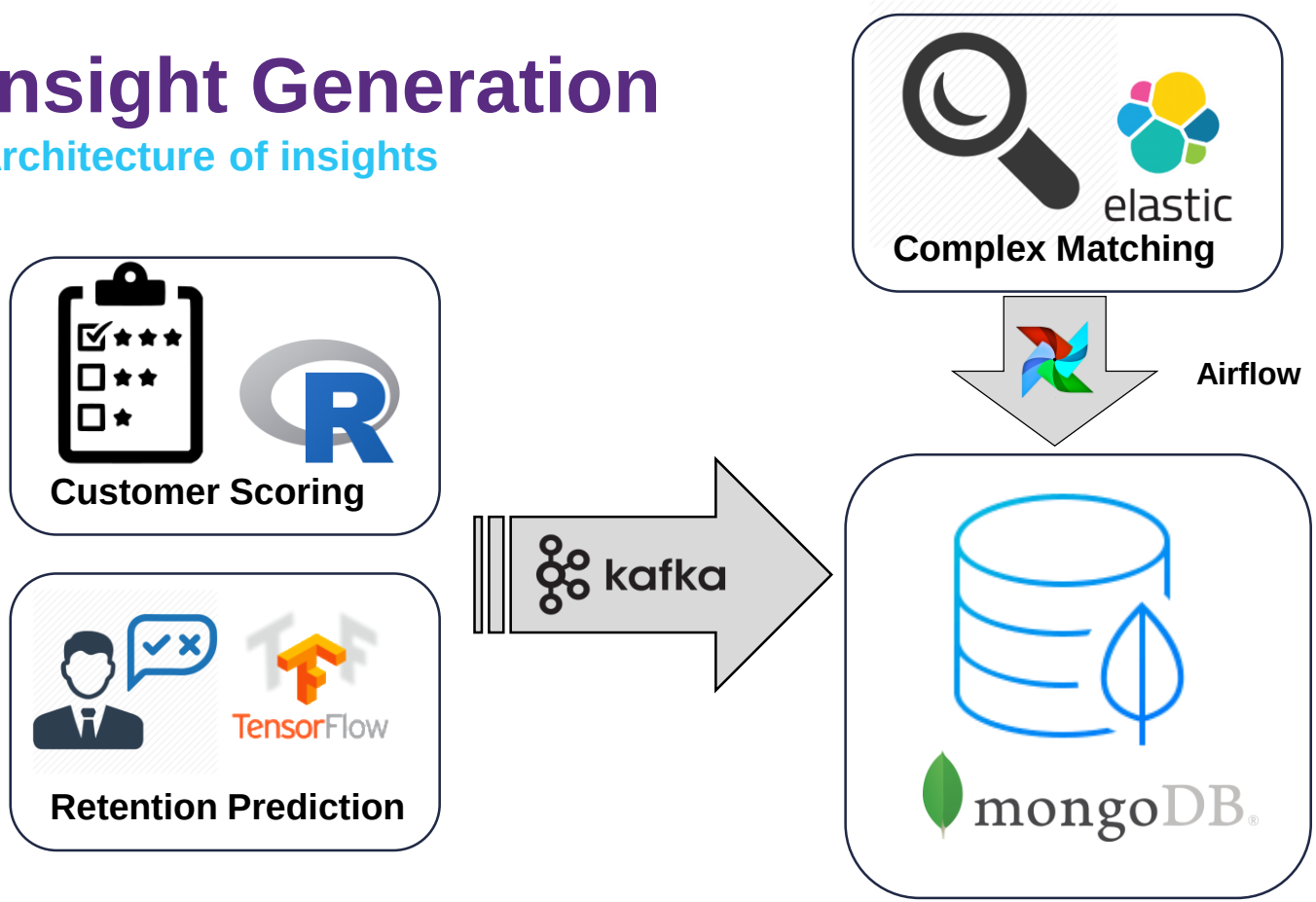
Architecture of capture



iag

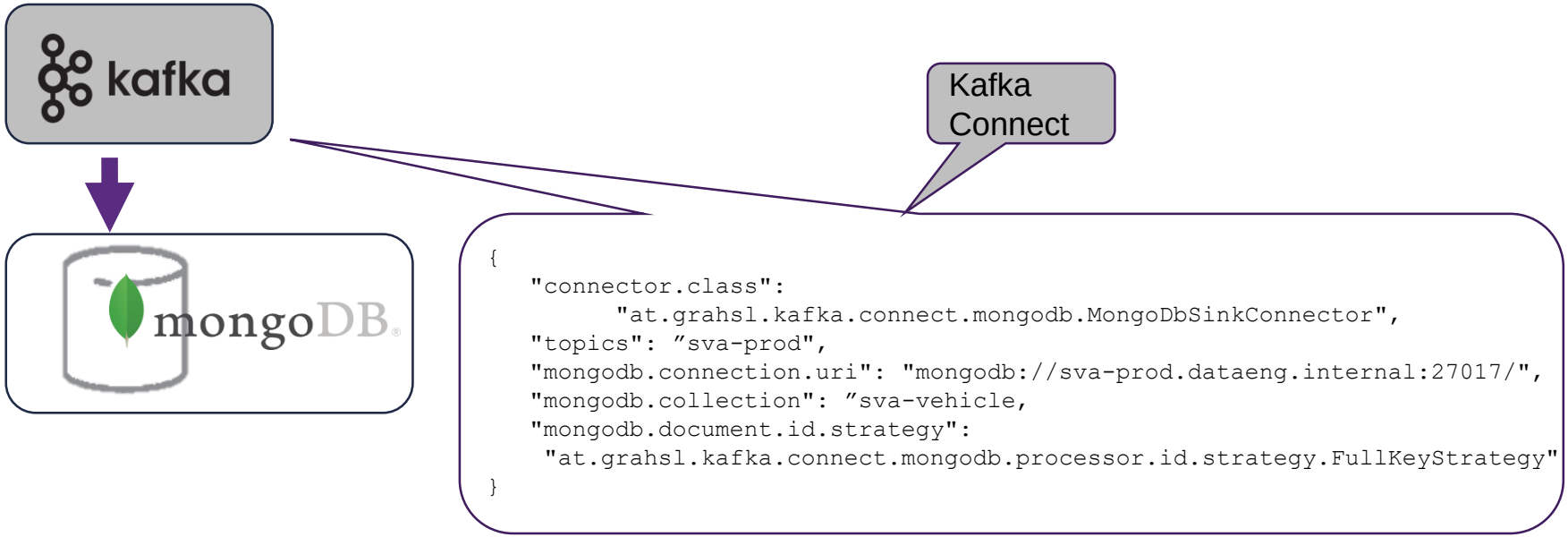
Insight Generation

Architecture of insights



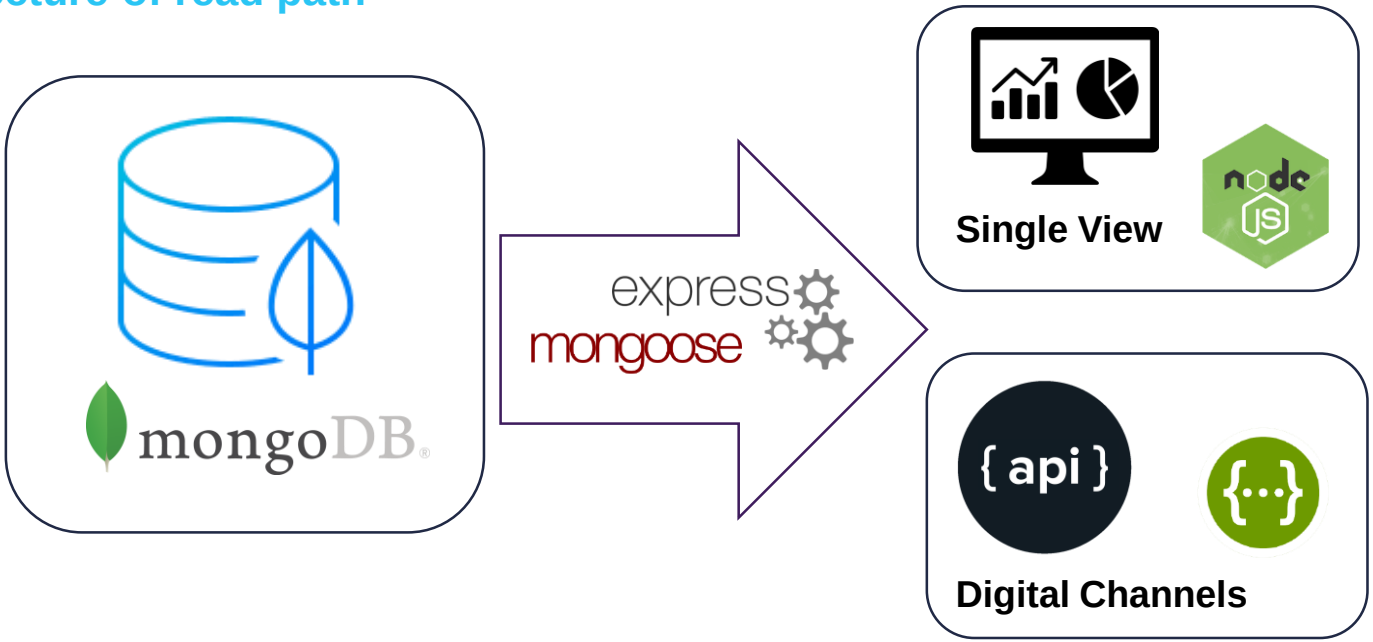
Kafka Connect Sink

Architecture of write path



Serving Tier

Architecture of read path

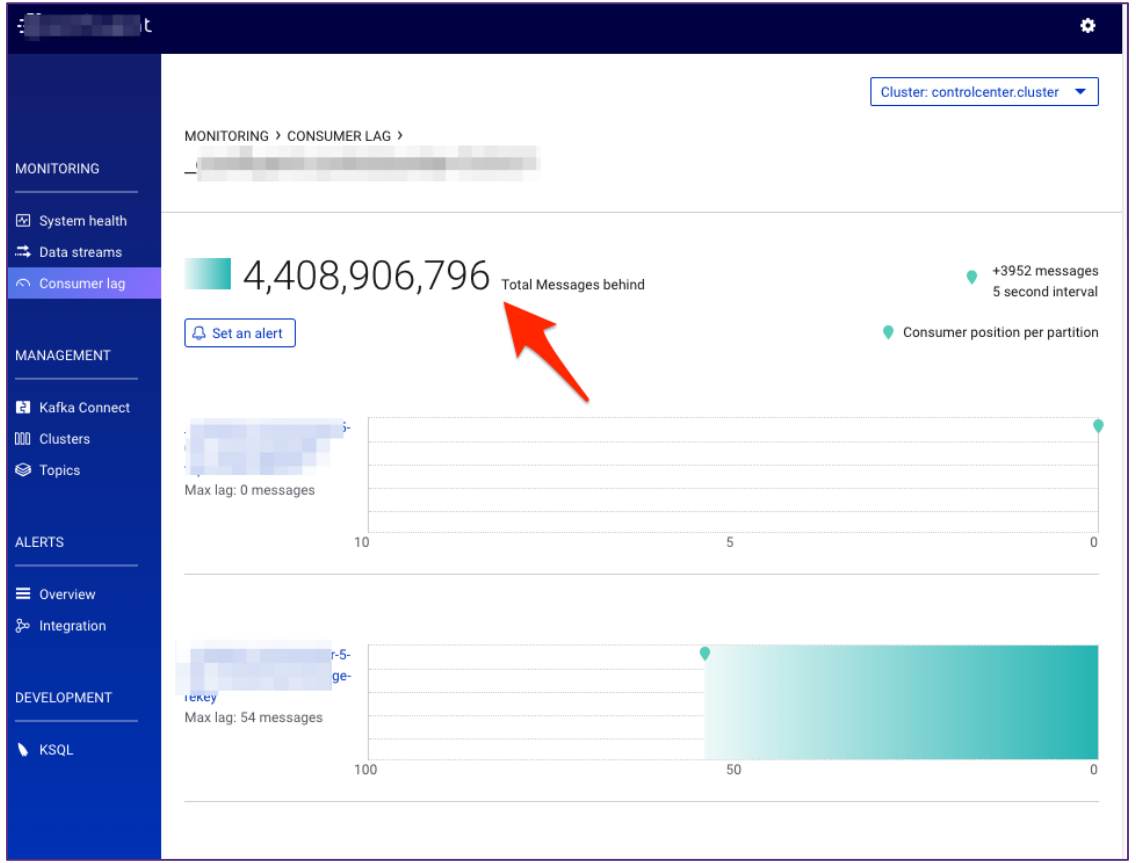


What did we discover?

Slow to fast … to *really* fast!

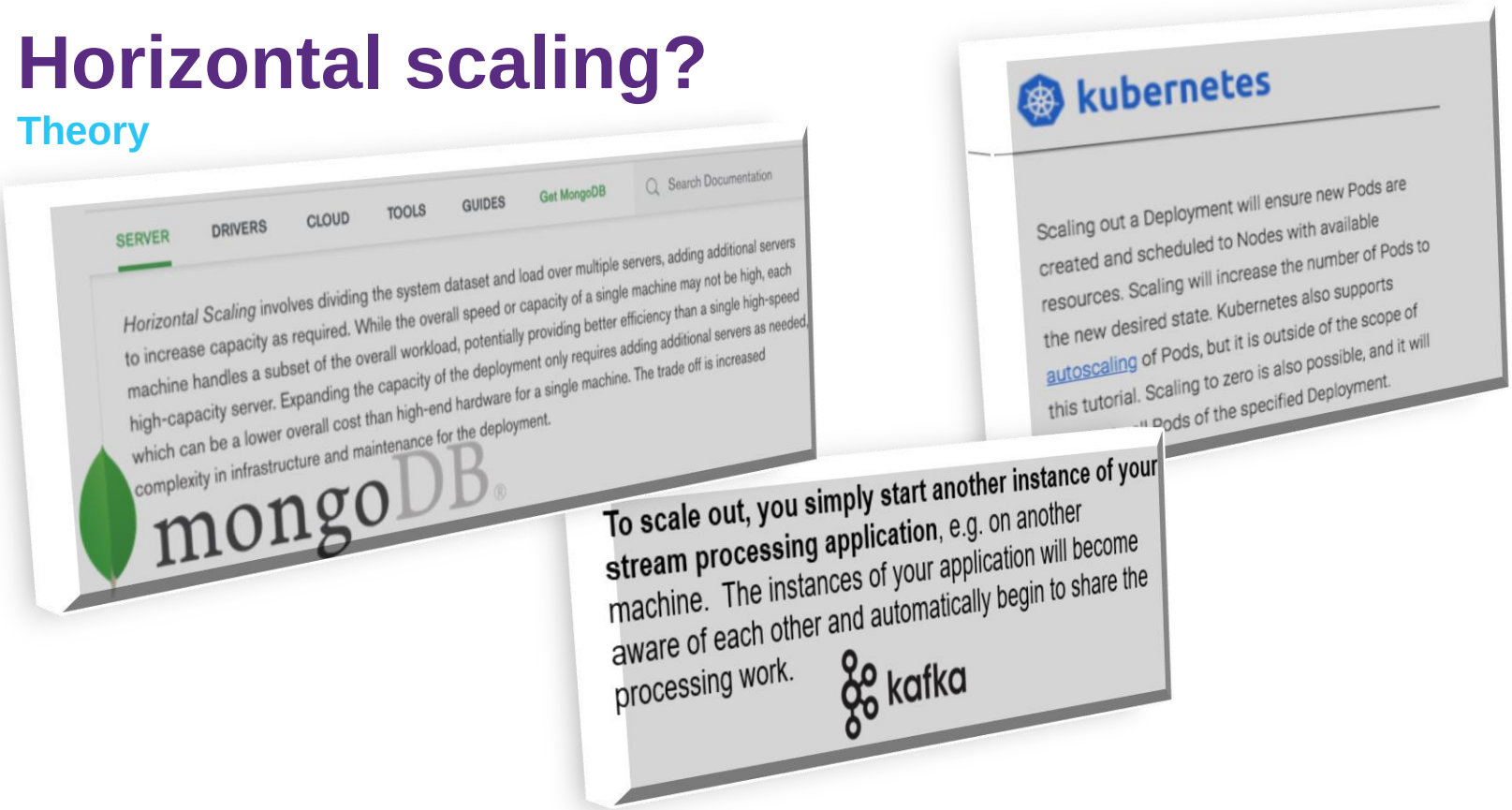
Challenge

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 ...

9:47 AM
Looking good on the 16 CPUs - Mongo is keeping up with Kafka on the ingest now

10:01 AM
I can bump to 32

10:03 AM
Nice - starting to hammer. 40% across all CPU's

10:30 AM
Hah - mongo's winning ! Topic drained

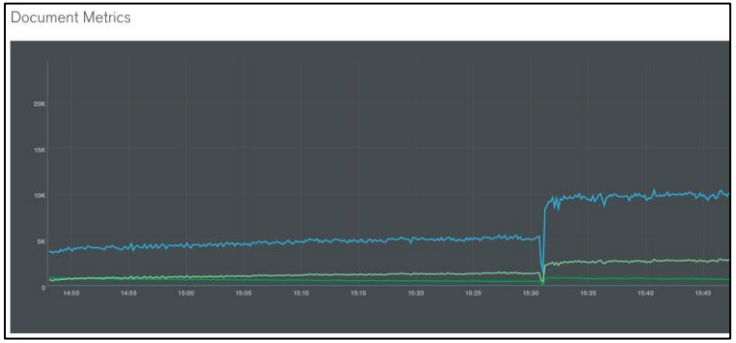
10:46 AM
New record - we're at 5,070 records per second
18 million / hour

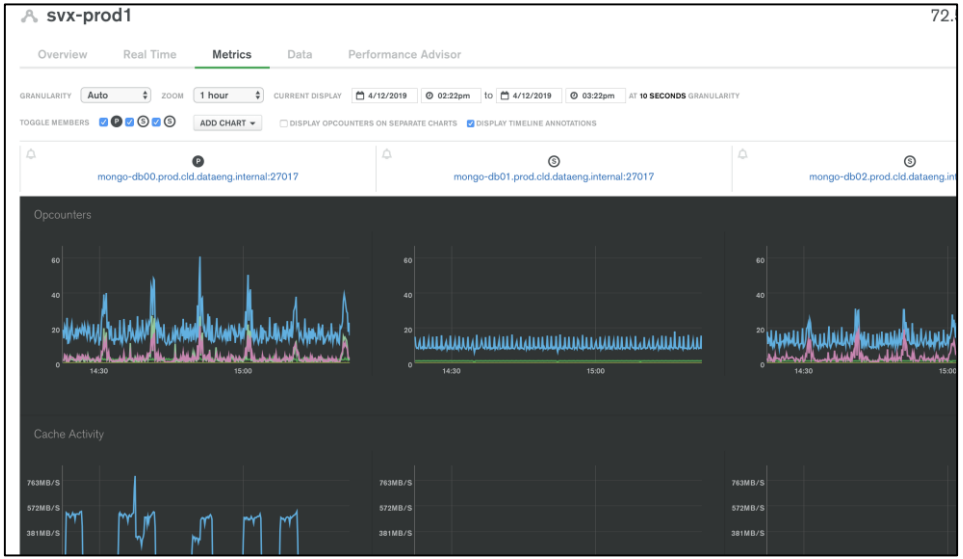
Horizontal scaling … meet efficient code

In numbers

30 records / sec

33,500 records / sec

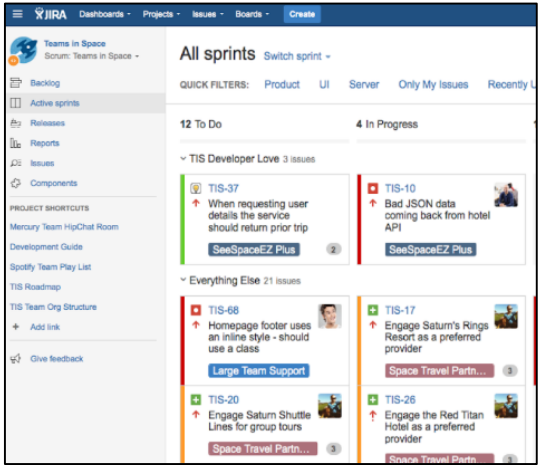


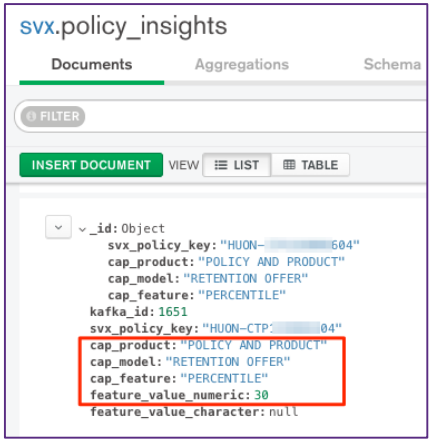


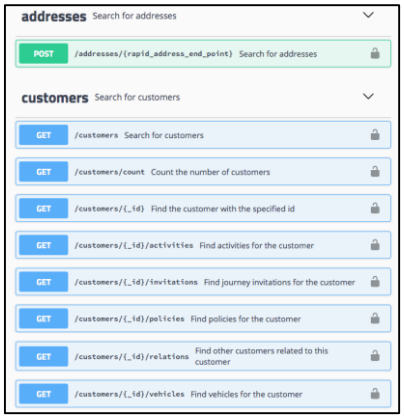
Who

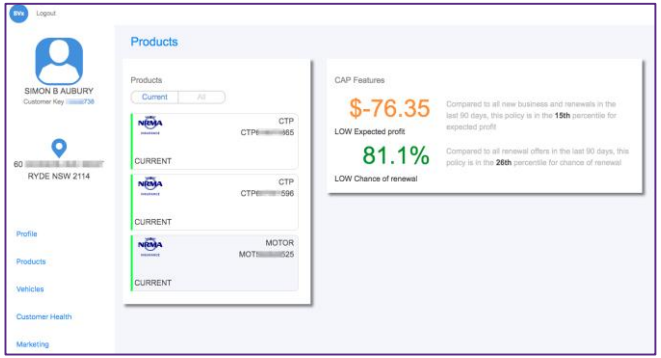
Participates in feature delivery?

Feature Sprint









What

Does this mean to our customers & stakeholders?

Unlocking Geospatial Analytics

What we ***couldn't*** do with our RDBMS

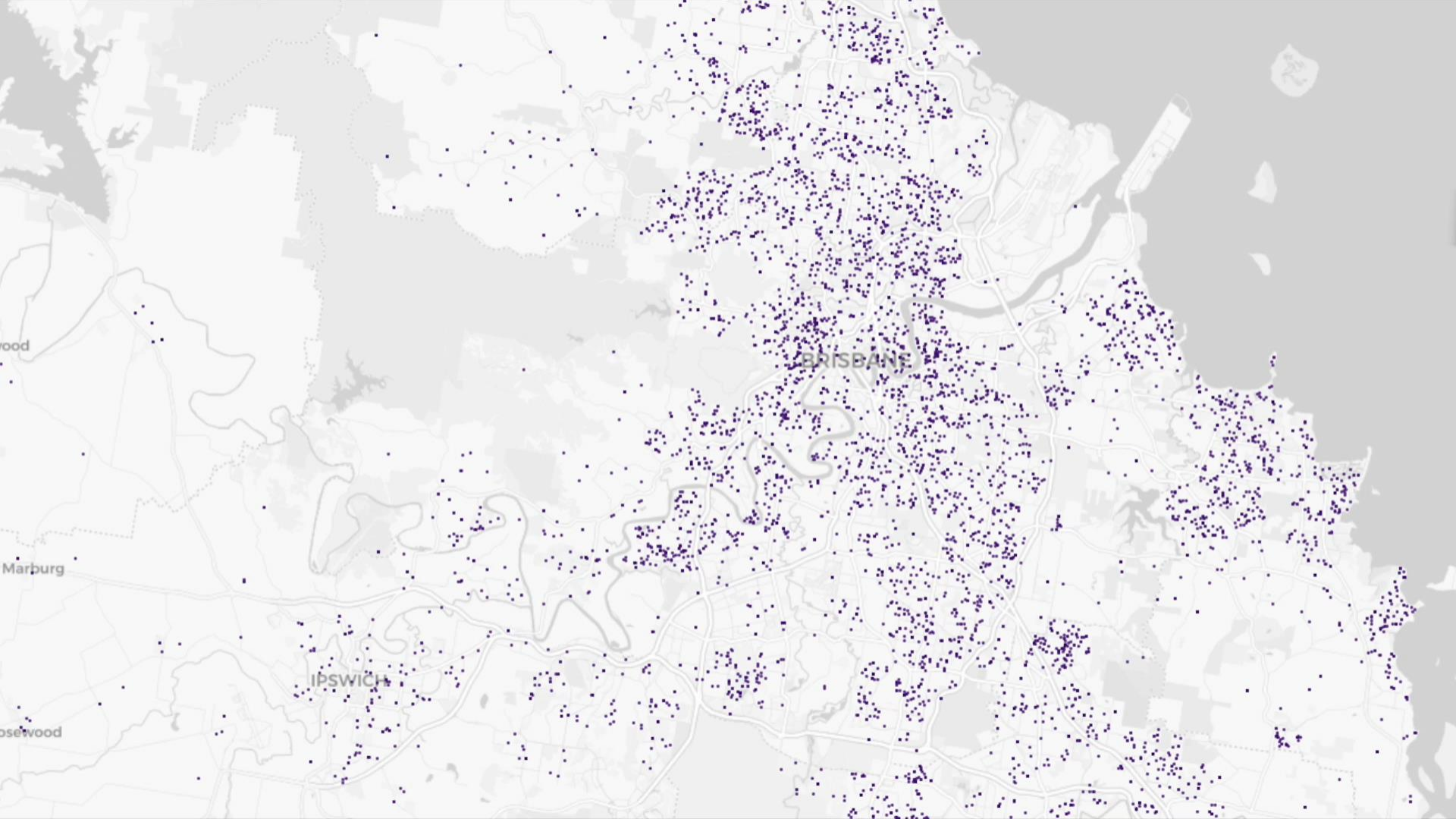
Analytics team

- 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"


BRISBANE
Marburg
IPSWICH

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???