
AWS
re:Invent

ADM304

Building the post-cookie identity graph for marketing

Carlos Rendon
Principal Technical Architect
Cox Automotive

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Introduction to Cox Automotive

An ideal consumer journey

The challenges of building that journey
- Tracking consumer identity across datasets
- The post-cookie world

How we built out our graph
- Choosing a graph database and Amazon Neptune
- Architecture
- Lessons learned

Results



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introduction to Cox Automotive

Transforming the way the world buys, sells, owns, and uses cars

>40,000 auto dealer clients across five continents
34,000-plus team members
Home to many brands including

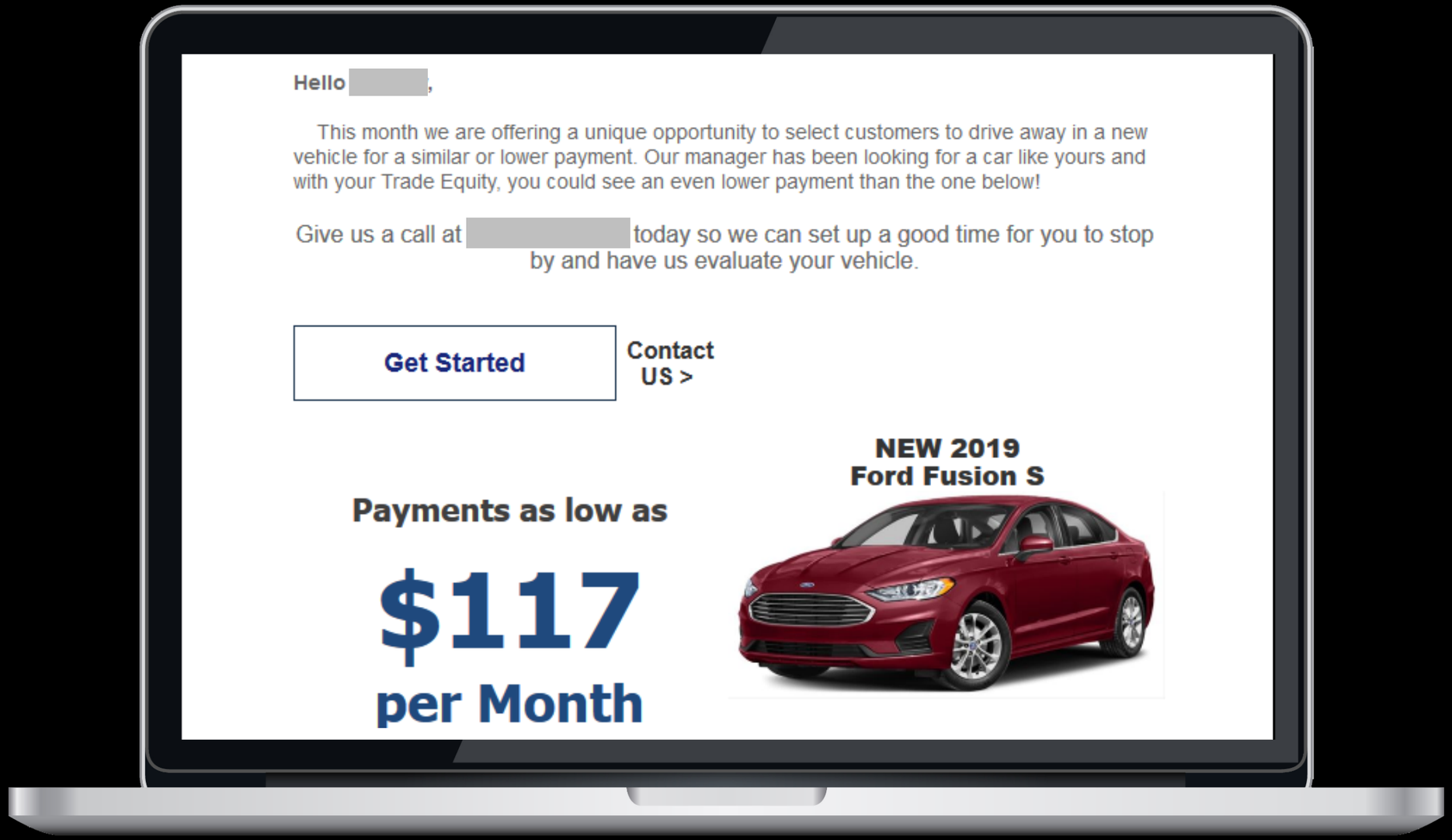
Autotrader	Kelley Blue Book	VinSolutions
Dealer.com	Manheim	vAuto
Dealertrack	NextGear Capital	Xtime

An ideal consumer journey

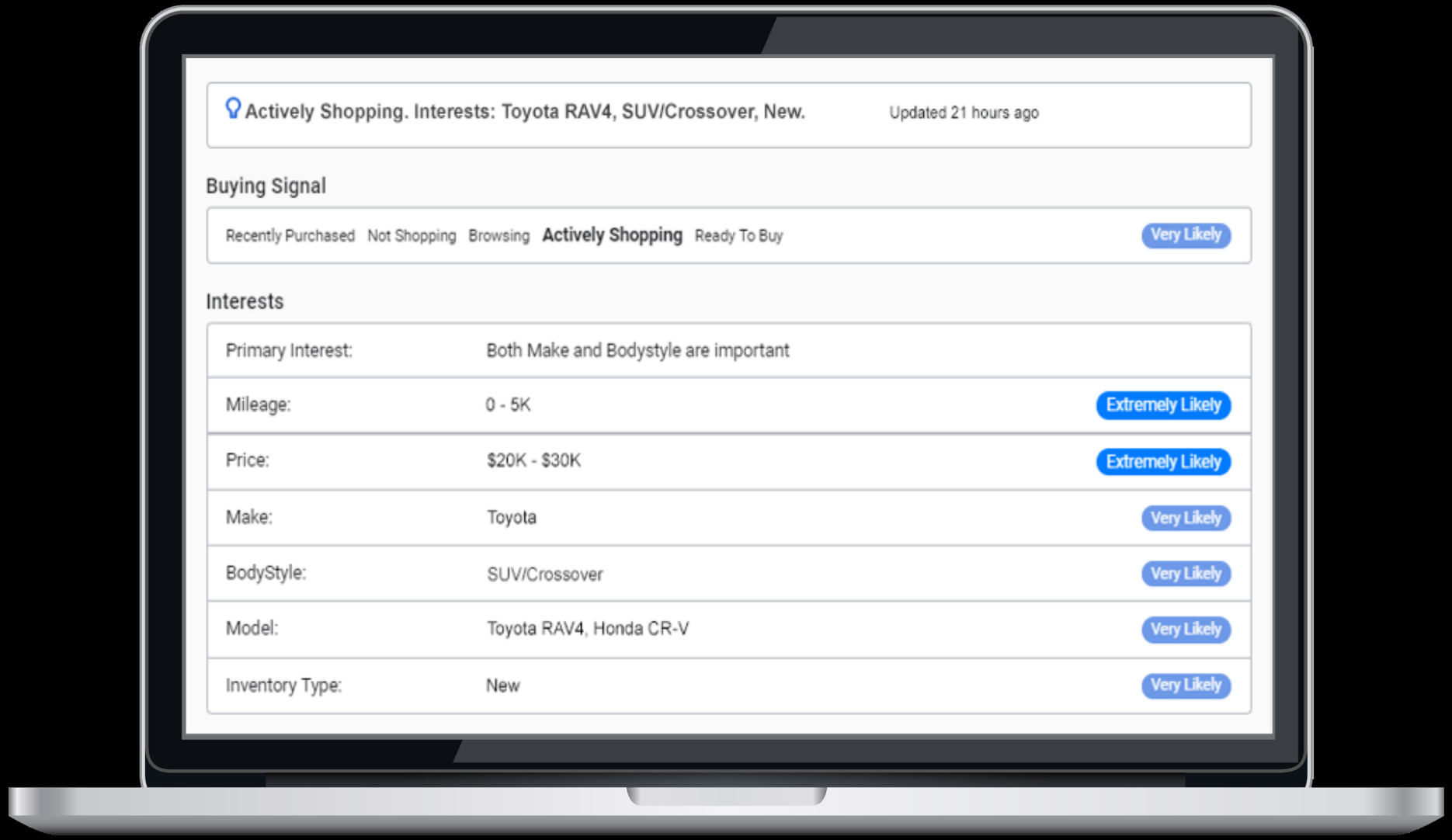


© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example: Targeted email

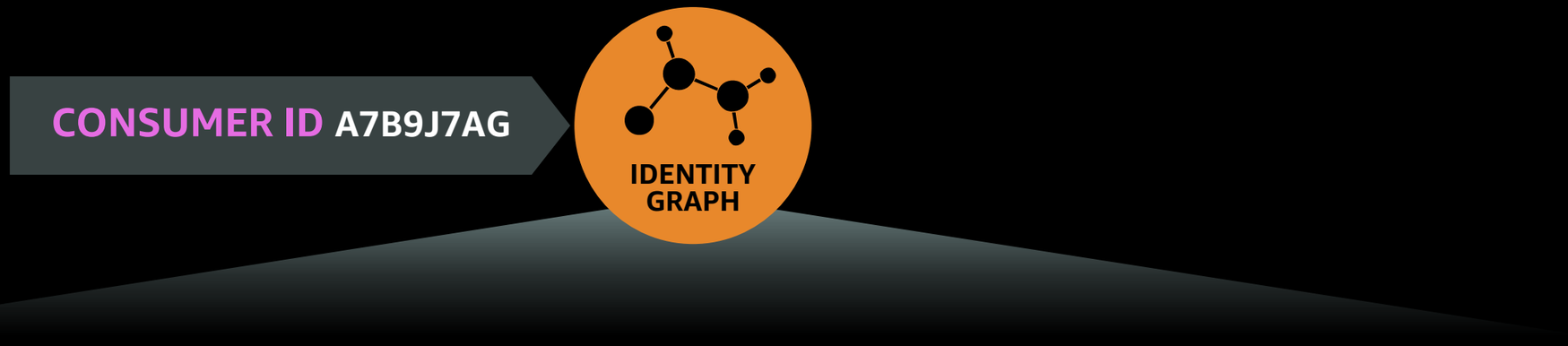


Example: Enriched data in CRM



What we want

A CONNECTED CONSUMER JOURNEY, COMBINED WITH MACHINE LEARNING, RESULTING IN RELIABLE PREDICTIVE INSIGHTS



LEASE/FINANCE PROPENSITY
Lease

CURRENT OWNERSHIP
Compact Sport Sedan

ONLINE RESEARCH
Full-Size SUV

LIFE EVENT
New Baby

HOUSEHOLD INCOME
$70K–$80K

COX AUTOMOTIVE SOURCES

Autotrader · Kelley Blue Book · DEALER.COM · VinSolutions · Dealertrack · xtime

EXTERNAL SOURCES

LiveRamp · Adobe · Acxiom

Use cases

CONSUMER SOLUTIONS

- Targeted advertising
- Website personalization
- Vehicle recommendations

Autotrader

Kelley Blue Book

DEALER SOLUTIONS

- Advertising dealerships and dealership websites
- Website personalization
- Email marketing
- Lead enrichment
- CRM insight mining

DEALER.COM

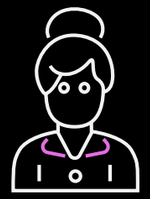
VinSolutions

The challenges of building that journey



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

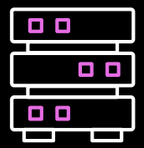
Challenges



Consumer identity is not shared across datasets



It’s getting harder to track consumers online

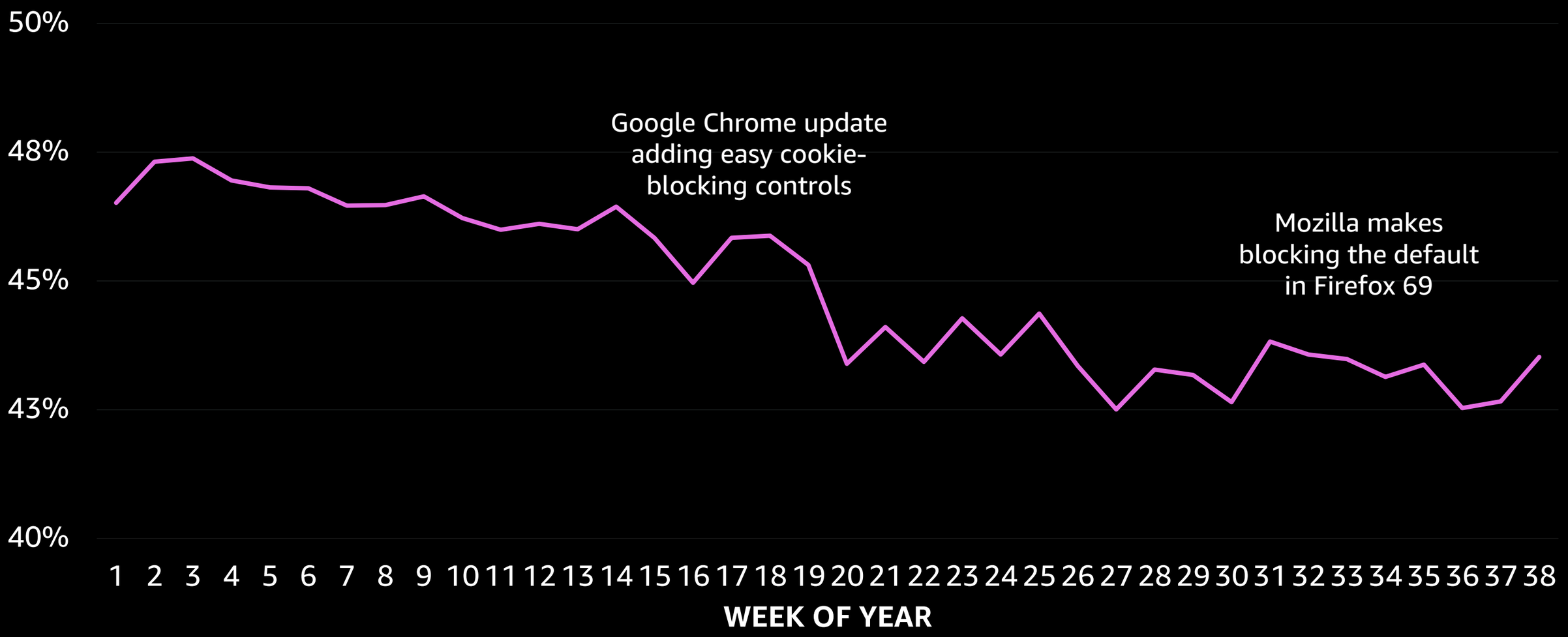


There’s a lot of data

Who is Carlos Rendon in my datasets?

Weblogs – a cookie?	5OsB5Hoh9guLNmOgNAbv4zkc uLNmOee5Hoh9bsv47eycM
An email address?	Carlos.Rendon@localhost
CRM – a unique key?	74621916397
Weblogs – an IP address?	175.64.236.232 2600:8902:700:2c8:2cba:d2f8:faf5:b825

Third-party cookie match rates in 2019



Most consumers are not trackable with third-party cookies!

Source: Cox Automotive Pixall tagging

Cox Automotive has a lot of data

Every year we see

39B
online interactions

21M
sales records

2.7B
online visits

158M
online leads

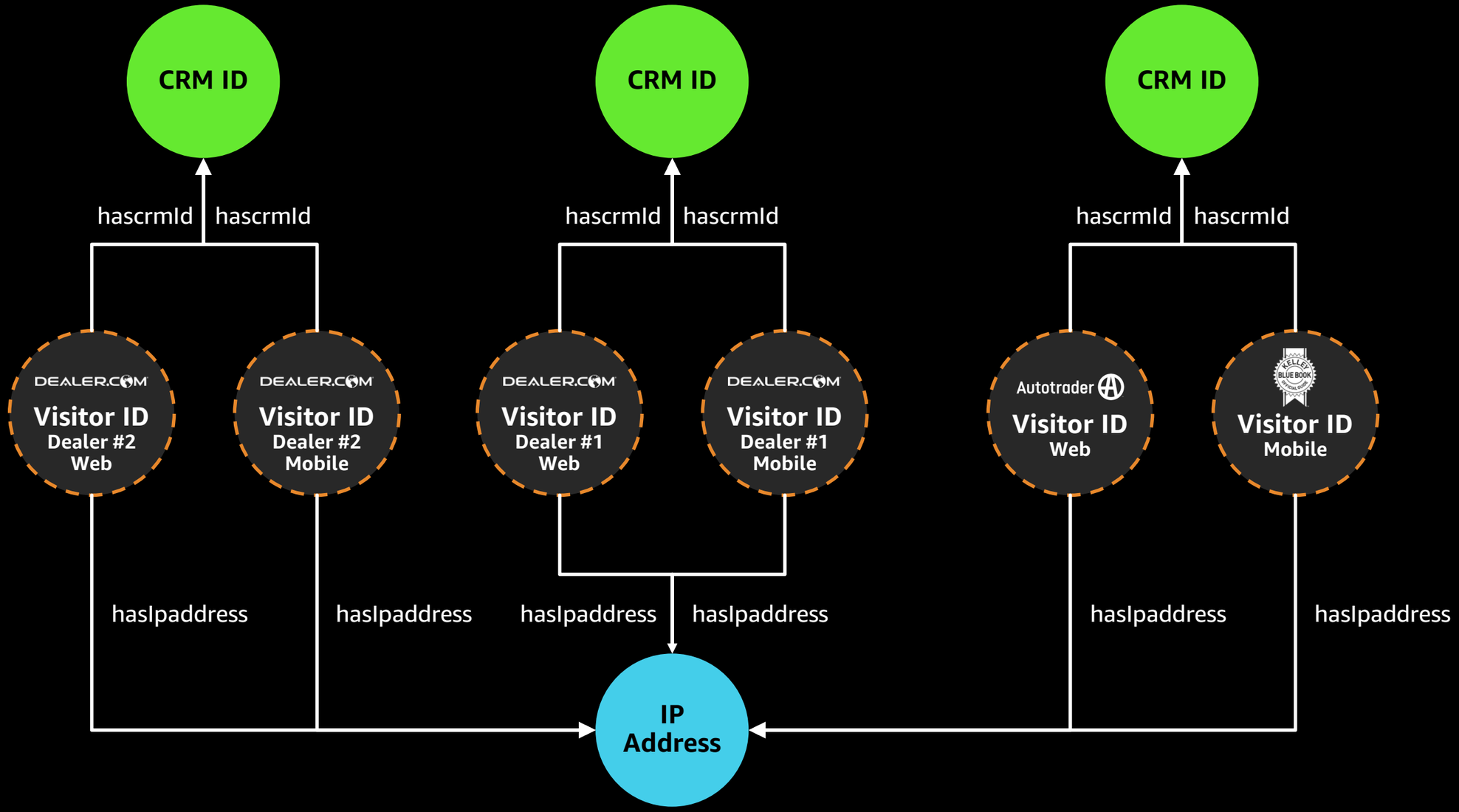
Building the identity graph



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is an identity graph?



Creating the relationships

Deterministic matching of online to offline

Leads submitted online with cookie attached flow to CRM

CRM emails to consumers with links to website to obtain cookie

Probabilistic grouping of cookies into “households”

Data science algorithm that allows us to tie cookies without third-party cookies

Using IP address as primary key

Not householding coffee shops, cell phone towers, offices

Why use a graph database?
Why Amazon Neptune?

Data model is easily expressed

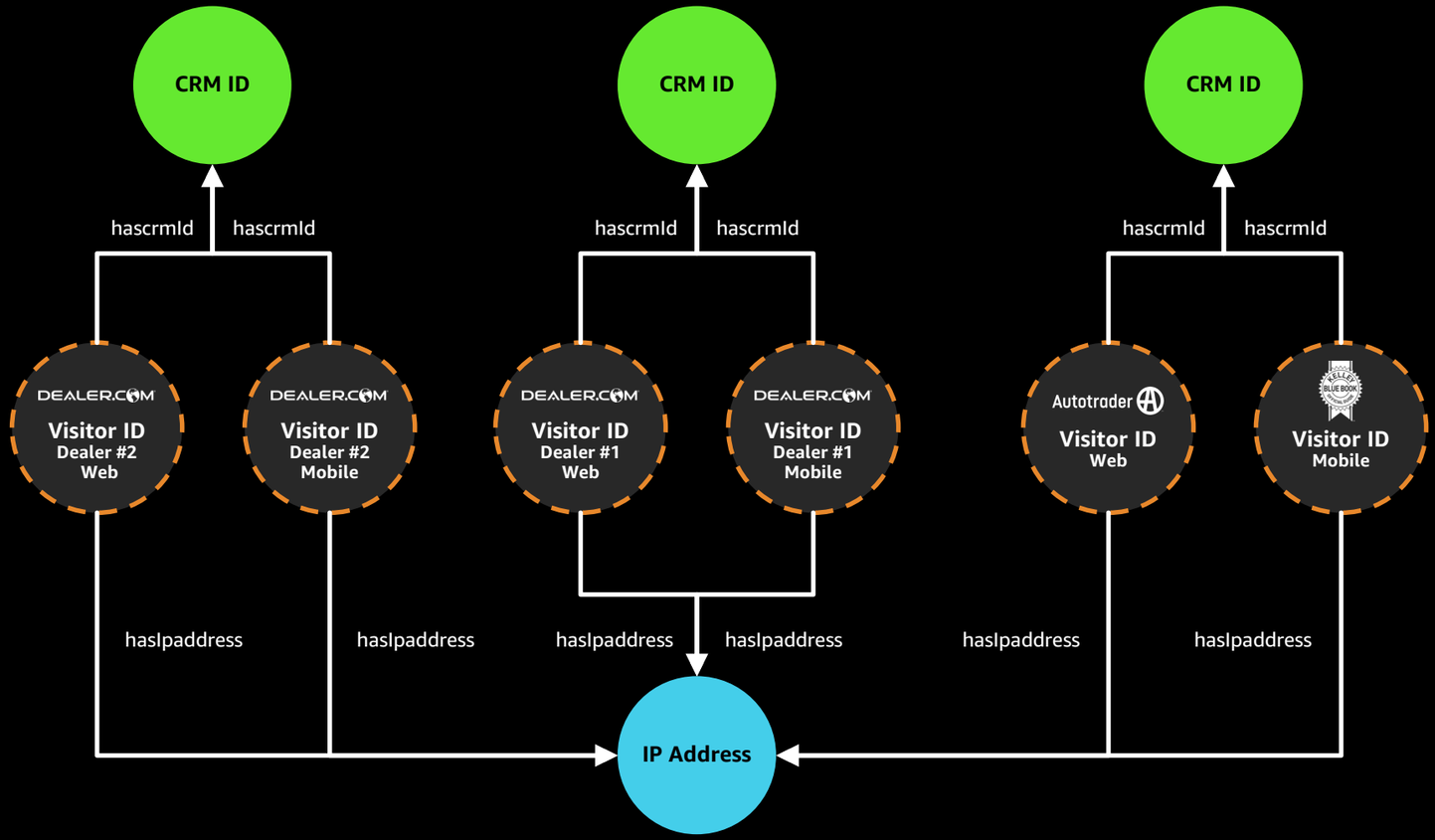
Good out-of-the-box query performance without tuning

Managed by Amazon and it's…

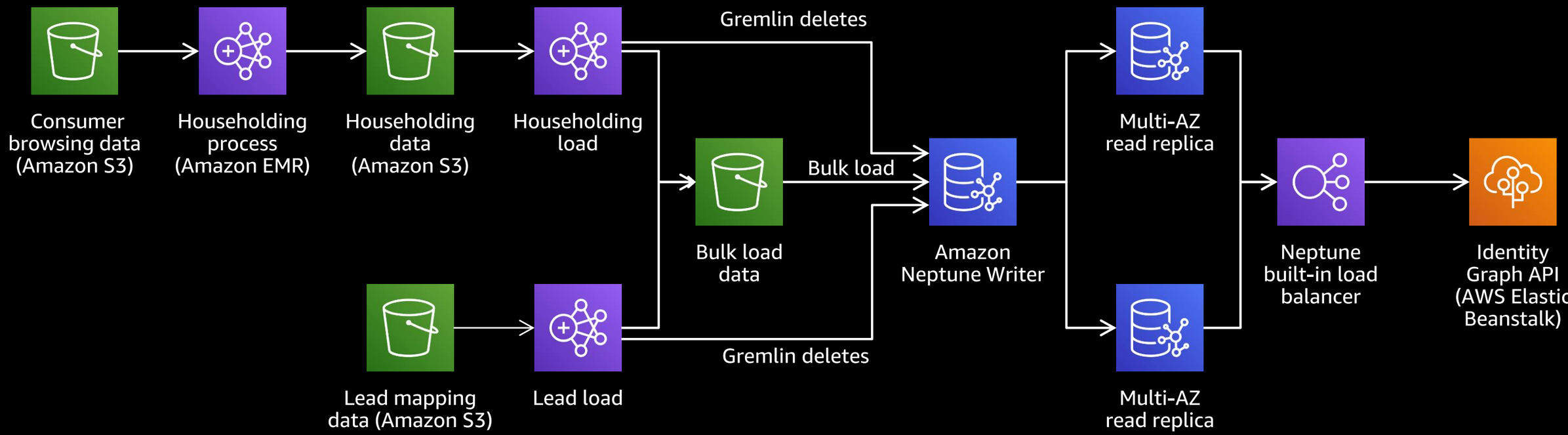
- Cheaper to run than an in-memory database
- Faster to query (while still being flexible) than key-value pairs
- Easier to design a schema and query than relational

Querying Neptune using Gremlin

```
g.V().has("name", visitorId)
        .out("hasIpAddress")
        .in("hasIpAddress")
        .out("hascrmId")
        .in("hascrmId")
        .values("name")
```



Architecture



Neptune statistics

db.r5.2xlarge instance type

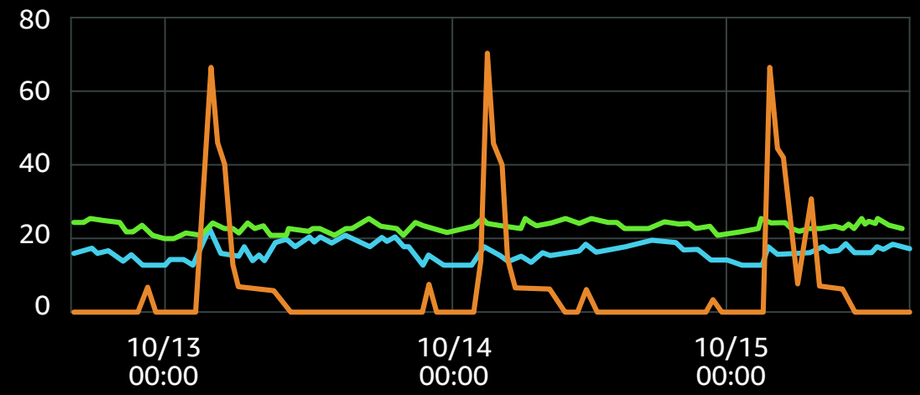
3 instances

~2,000 queries per second

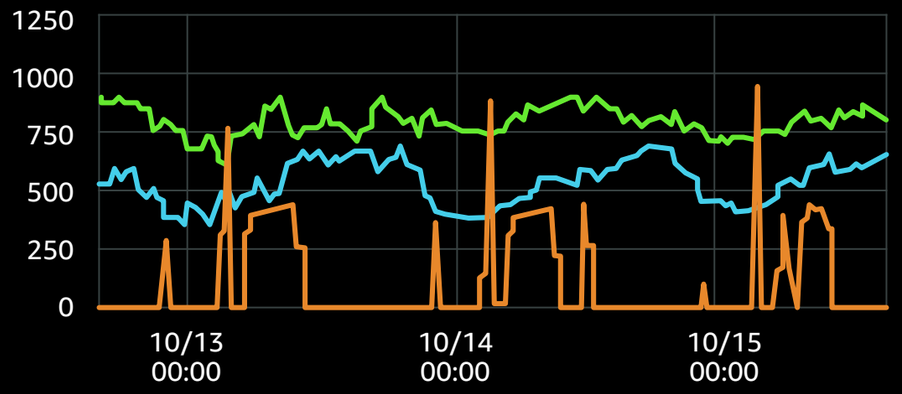
~500 million vertices

1.6 TB storage in Neptune

CPU Utilization (percent)



Gremlin requests (count/second)



Incremental data loads

Loading 500 million vertices takes a long time

Bulk-load API is fastest way to load Neptune,
need to provide specially formatted text files in Amazon S3

Loading only incremental changes keeps it fast

Bulk-load API does not perform deletes,
so we perform that with the Gremlin API

Neptune lessons learned

Train the team on graph databases

Use Amazon's published best practices https://amzn.to/383aNnR

Temporarily use a large instance size to load data fast

Perform incremental loads using
bulk-load API if you have a lot of data

Identity graph lessons learned

Make it real time

Privacy landscape changes all the time

Internet landscape changes (more than 40% of our traffic is IPv6)

Monitor your key metrics to look out for the above

You need consider permissible use of the data, both business and legal

Results



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Internal results

180% increase average online interactions per consumer

380% increase average cookies per consumer

Six use cases that benefit from identity graph today

Customer quotes

“Connect automotive intelligence has really helped us identify in-market shoppers within our CRM database, oftentimes prior to the customer raising their hand. This has allowed us to engage the customer earlier in the buying cycle, helping us prevent them being lost to another dealership.”

Kevin Patterson
Vice President, Nourse Chillicothe Automall

“Because we’re using this technology, we’re able to improve the interactions that we have with customers and really tailor conversations based on their interests.”

Andy Guelcher
General Manager, Mohawk Honda

We’re hiring

https://jobs.coxenterprises.com/



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.