

Scaling ML in Digital Advertising

Claudia Perlich

Chief Scientist

@claudia_perlich

dstillery

Programmatic Advertising

The screenshot shows the New York Times homepage with several programmatic advertisements highlighted in red boxes:

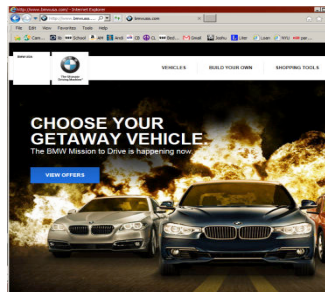
- Top Left:** "THE ALL-NEW CADILLAC XTS" advertisement.
- Top Right:** A small image of a Cadillac XTS car.
- Search Bar:** "ING DIRECT" advertisement.
- Left Sidebar:** A vertical list of categories including "WORLD", "U.S.", "POLITICS", "NEW YORK", "BUSINESS", "DEALBOOK", "TECHNOLOGY", "SPORTS", "SCIENCE", "HEALTH", "ARTS", "STYLE", and "OPINION".
- Main Content:**
 - Fiscal Crisis in States Will Last Beyond Slump, Report Warns** by MARY WILLIAMS WALSH and MICHAEL COOPER. Includes a "PLAY TIMESCAST" button.
 - Unity Government in Israel Disbanding Over Draft Dispute** by JODI RUDOREN and RICK GLADSTONE. Includes a "PLAY TIMESCAST" button.
 - Cautious on Growth, Bernanke Offers No Hint of New Action** by BINYAMIN APPELBAUM.
- Right Side:**
 - OPINION** section with articles like "Islamic School for Girls" and "Brooks: Capitalism Debate".
 - MARKETS** section with S&P 500, Dow, and Nasdaq data.
 - GET QUOTES** section with "My Portfolios" and "Go" button.
 - Advertisement:** "The New York Times TRY IT NOW 4 WEEKS FOR 99¢ CLICK HERE".
- Bottom Section:**
 - CAMPAIGN 2012** header.
 - Mormons' First Families Rally Behind Romney** by JIM RUTENBERG.
 - Inside the Romney No. 2 Hunt: First, Do No Harm** by ASHLEY PARKER and MICHAEL BARBARO.
- Bottom Right:** A large video advertisement for "INTRODUCING THE ALL-NEW CADILLAC XTS" with a "CLICK FOR SOUND" button and "LEARN MORE" link.

300 Million (US) consumer



Interacting with brand sites

Using Digital Devices



Measurement

Conversion

If we win an auction we serve ad of brand

100 ms to bid

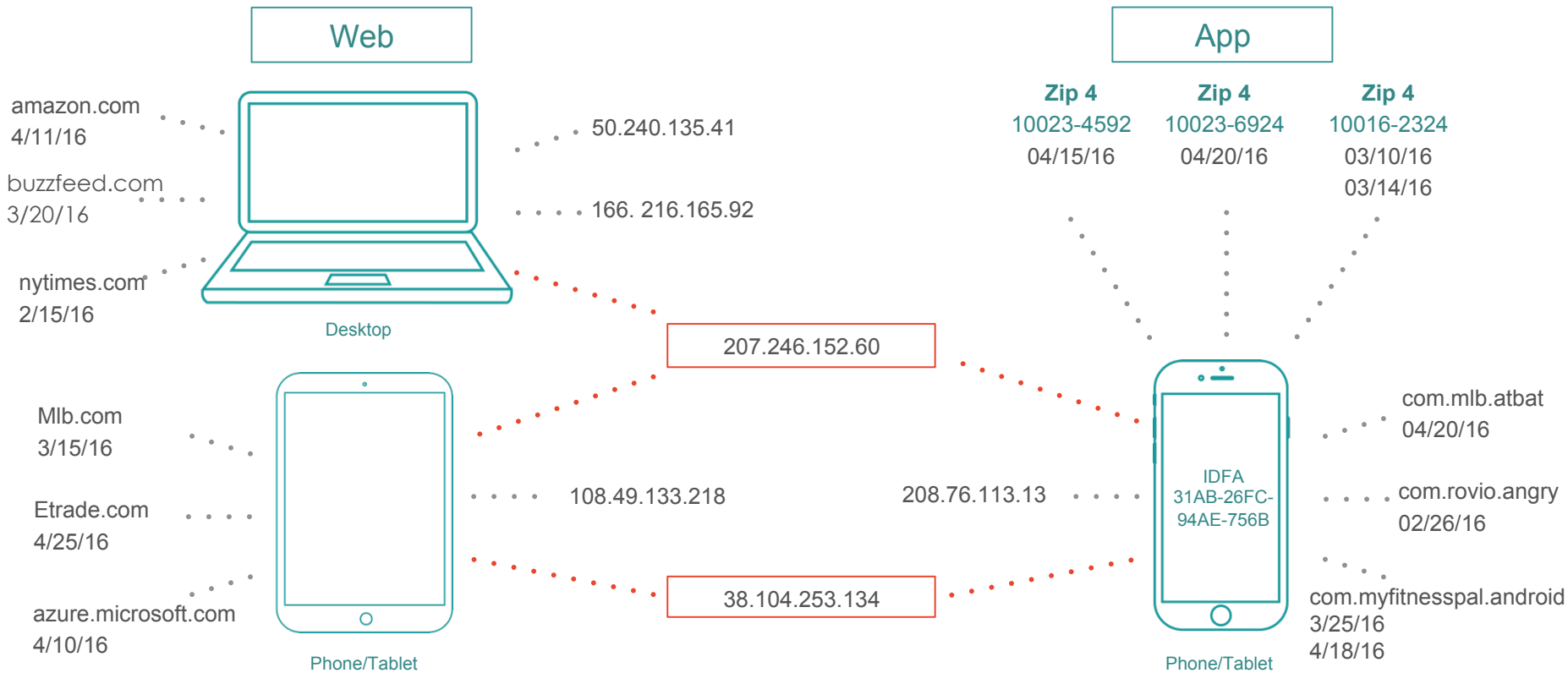
Work with Brand



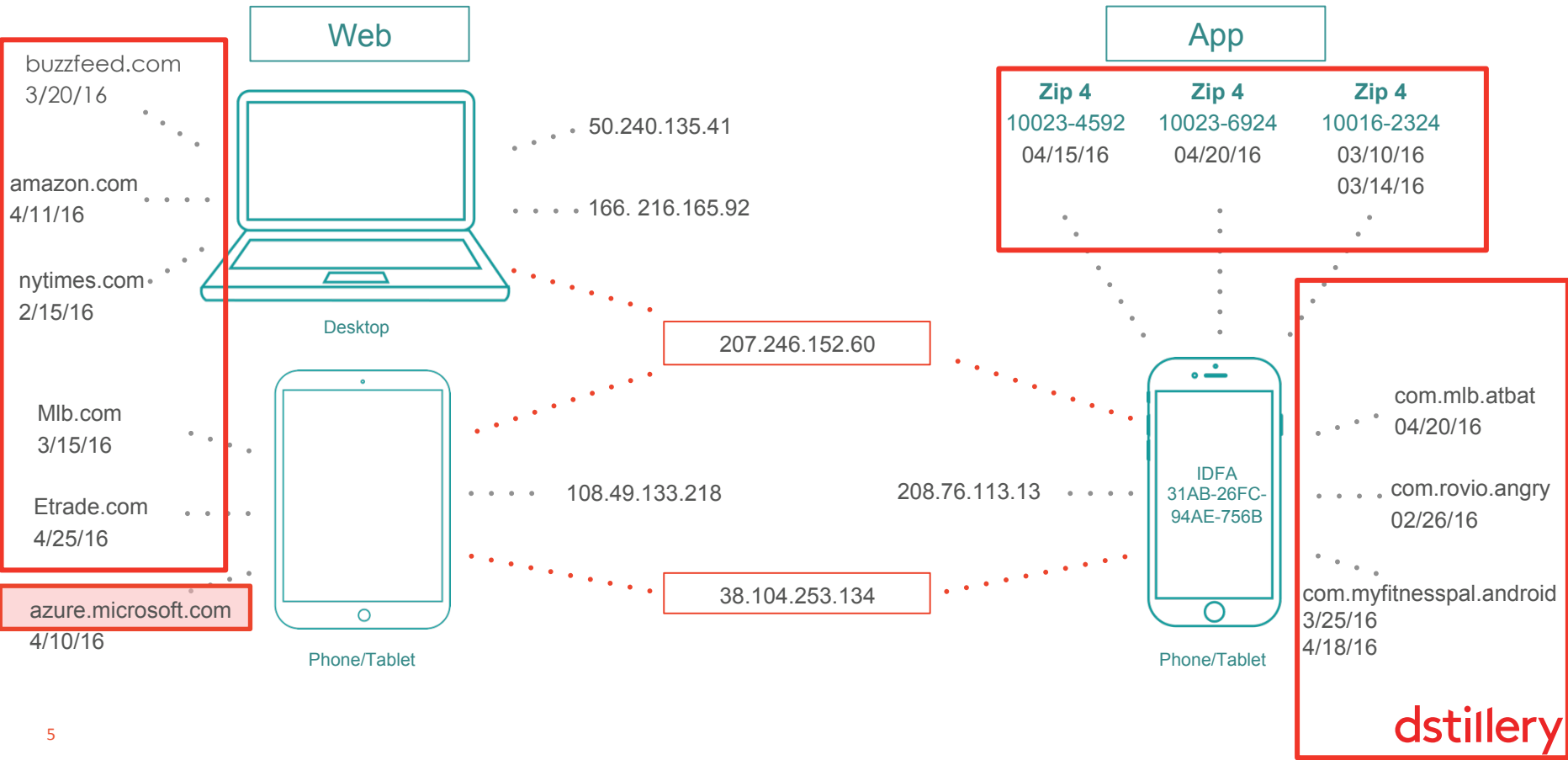
100 Billion bid requests per day

dstillery

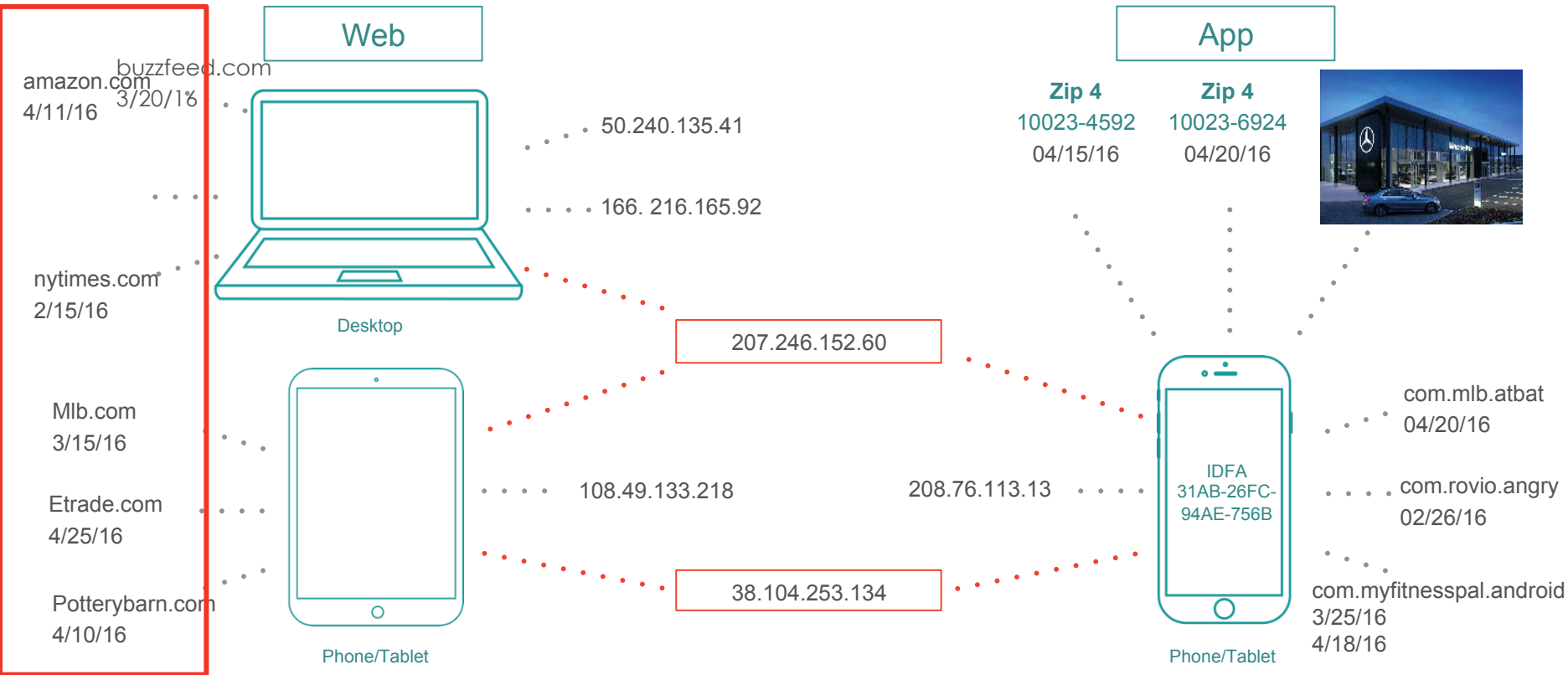
Dstillery Data: 100 Billion events per day across devices



Predicting events on pretty much everything



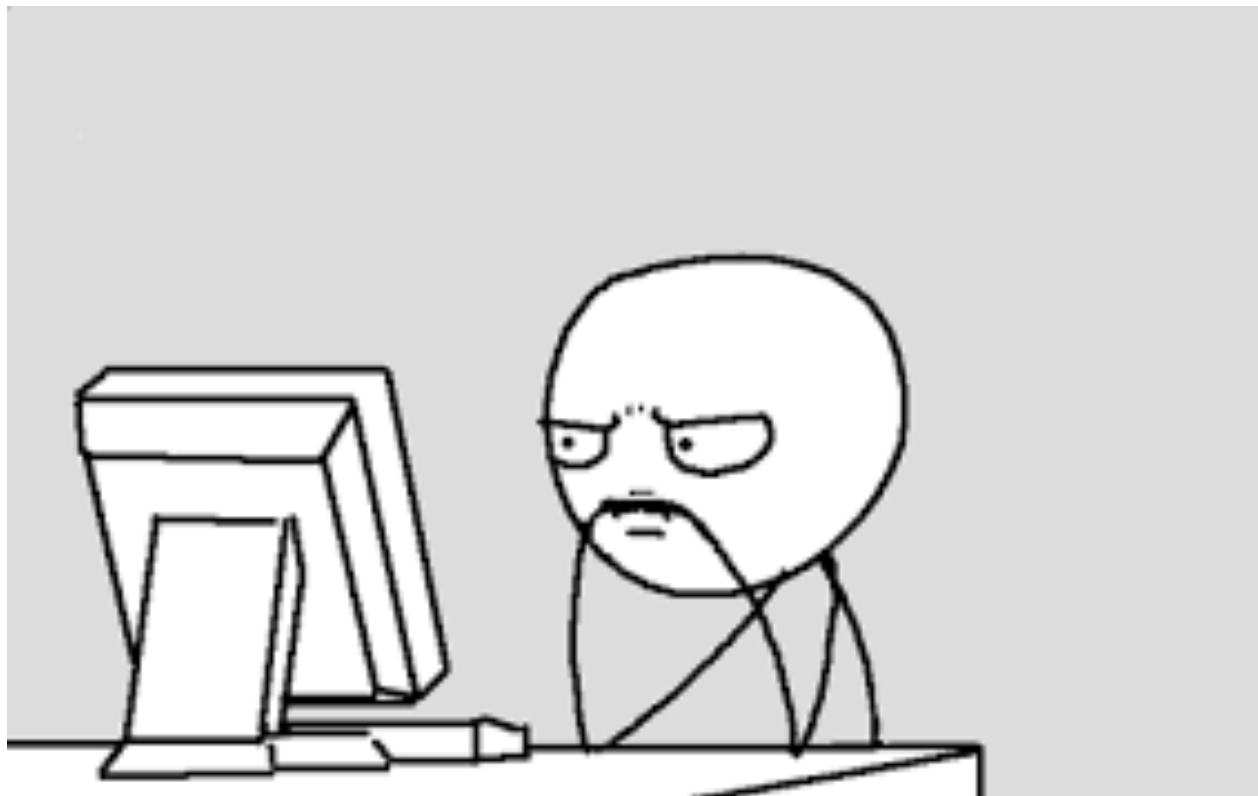
Predict who will go to a Mercedes Car dealership ..



Lets talk scaling? Most Limited resource?



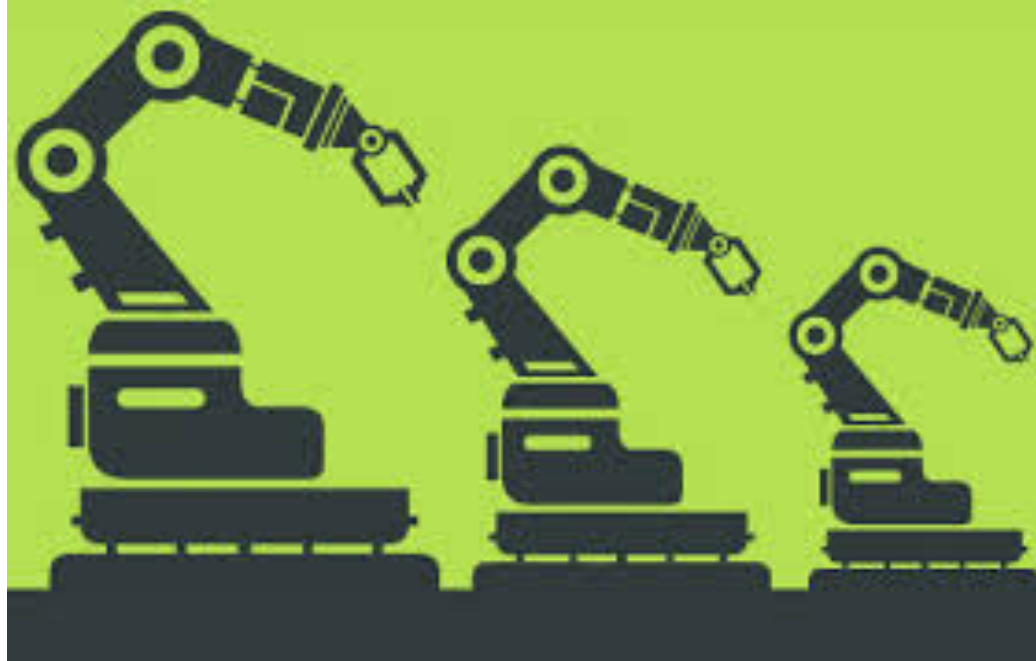
Personally: Waiting for your model estimation sucks!



Tip of the iceberg ...



World of Digital Advertising



Automation beats scaling

Scaling the Solution, not the ML ...

Build/manage thousands of models with limited DS resources ...

- All kinds of target/baserates:
various types of conversion, click, viewability, video completion, store visits, purchases
- Typically low predictability
- Create additional model for new target with a flip of a switch in hours
- Manage the feature space and automate deployment

Score ~200 Million examples daily against 10K models

- Less than 0.005 ms each

Economic efficiency

DS Dashboard ...

317

Campaigns on Optimus (out of 432 active campaigns; 249 auto set prices; 83 auto create targetgroups)

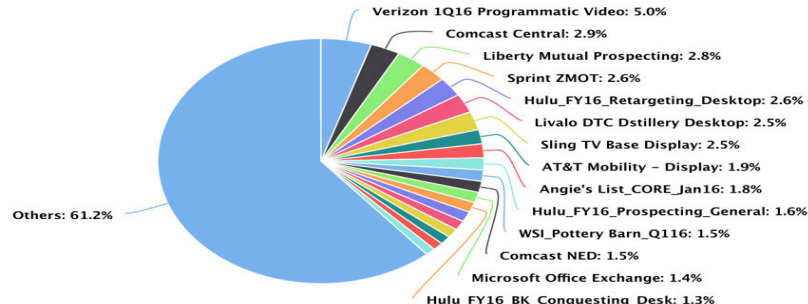
3116

SGD models were built in the past 24 hours

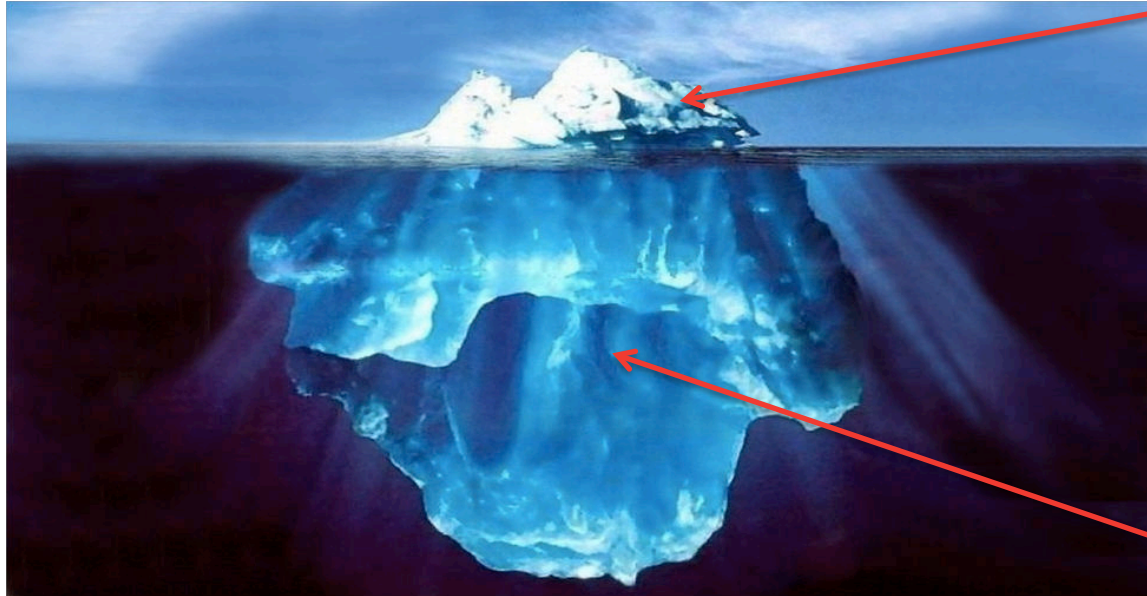
Projects:

- [CFP Sampler Specs](#)
- [Data Nexus](#)
- [DS Crosswalk](#)
- [Inventory Stats](#)
- [Optimus Campaigns](#)
- [Place Still Tag Index](#)
- [Seg Avails](#)
- [Sport](#)
- [Segrank Campaigns](#)
- [Exchange Specific Stats](#)
- [Optimus Decider Audit Trail](#)

Top campaigns by daily revenue



Automating Machine Learning

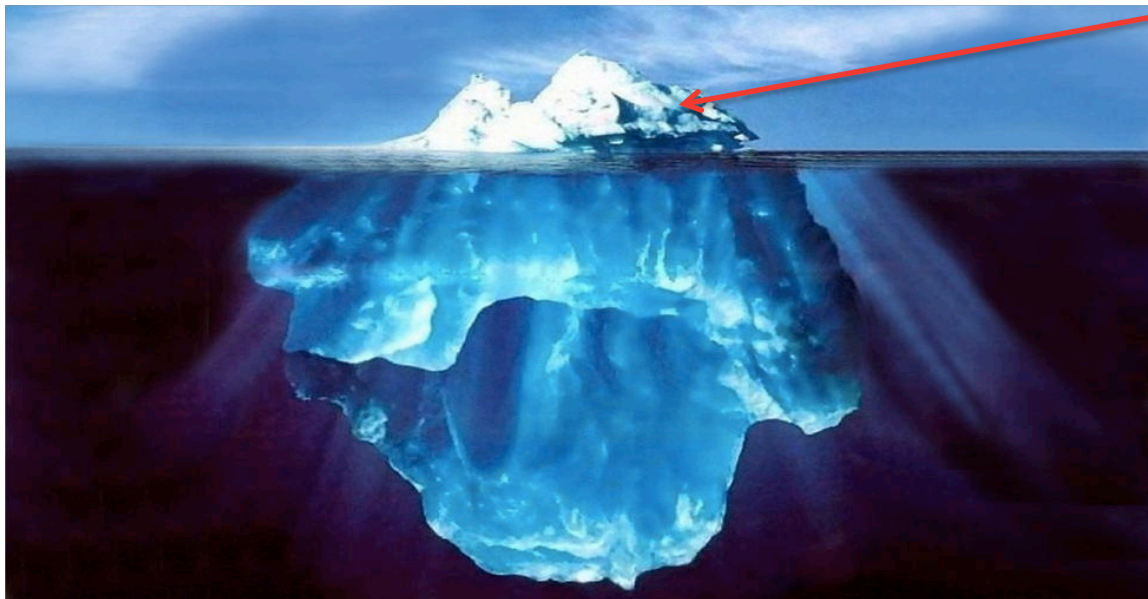


Strategic ML Choices ...


Infrastructure

Automating Machine Learning

Strategic ML Choices ...



ML Strategy: robustness & interpretation beats peak performance

- Feature engineering
 - Sparse binary or hostnames: easy to code all kinds of data
 - ~100 Million (possible) features
 - Stay 'close' to the raw data
 - Hashing to ensure consistent feature space
 - One MUST fit all ...
 - Algorithm
 - Logistic regression (and some Naïve Bayes)
 - Stochastic gradient descent
 - Penalties & Heuristics for Learning Rate Estimate
 - Constant refresh every ~3 days
 - Stacking & Transfer Learning & Sampling
- 



SGD Model (Model ID: 1728493)

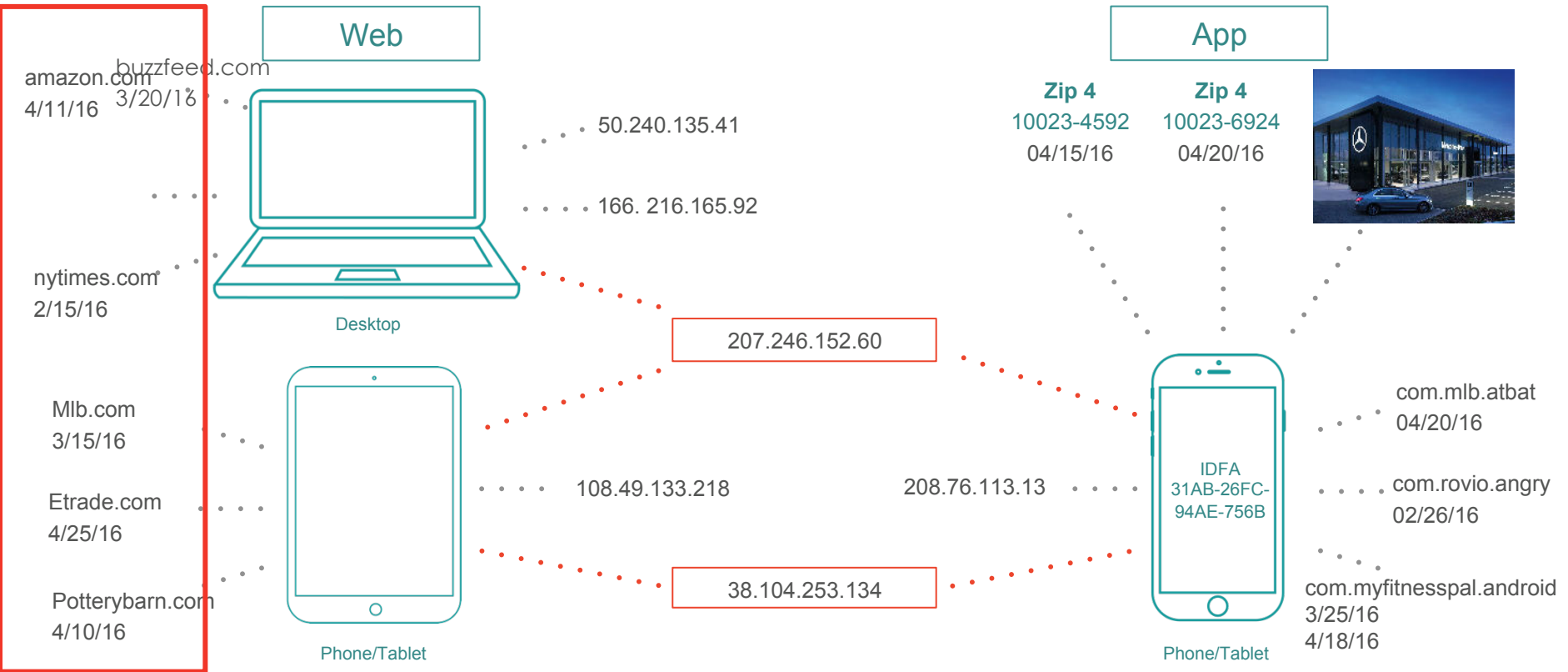
Marketer	Microsoft (Empowering Media SF) [6000]
Offer	Microsoft Azure Evergreen [7599]
Offer Level	40: Free Trial [19945]
Model Created At	2016-03-06 12:36:39
Model Event End Time	2016-03-06 17:40:00
# Positives	217
# Negatives	2,170
# Features	362,059
Parent Model	1725331
HDFS Path	hdfs://hdfs2/user/sgd/models/19945/OL19945_FM2_BS20160306123636.model

Top Features with Highest Scores

Show entriesSearch:

Feature	Coef
[46303095]	3.117
blogs.msdn.com [33892607]	3.010
www.why-recycle.com www.technet.com [46550201]	2.932
www.lz95.org [35757707]	2.779
www.dorisandjillycook.com www.dotnetfunda.com www.edenprairiecenter.com [35375642]	2.508
www.onebox.com www.avanquest.com [35325508]	2.467
www.office.com [46380052]	2.410
[46305049]	2.297
www.verbasoftware.com [46446087]	2.294
www.rosebyrne.org www.socialapocalypse.com www.milfmovies.biz www.milwaukee-wisconsin-wi.com www.msexchange.org [35355270]	2.288

Predict who will go in 7 days to a Mercedes dealership



Identify people who will go to Mercedes dealerships



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www.fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

'In the market' signal



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www.fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

Real Estate



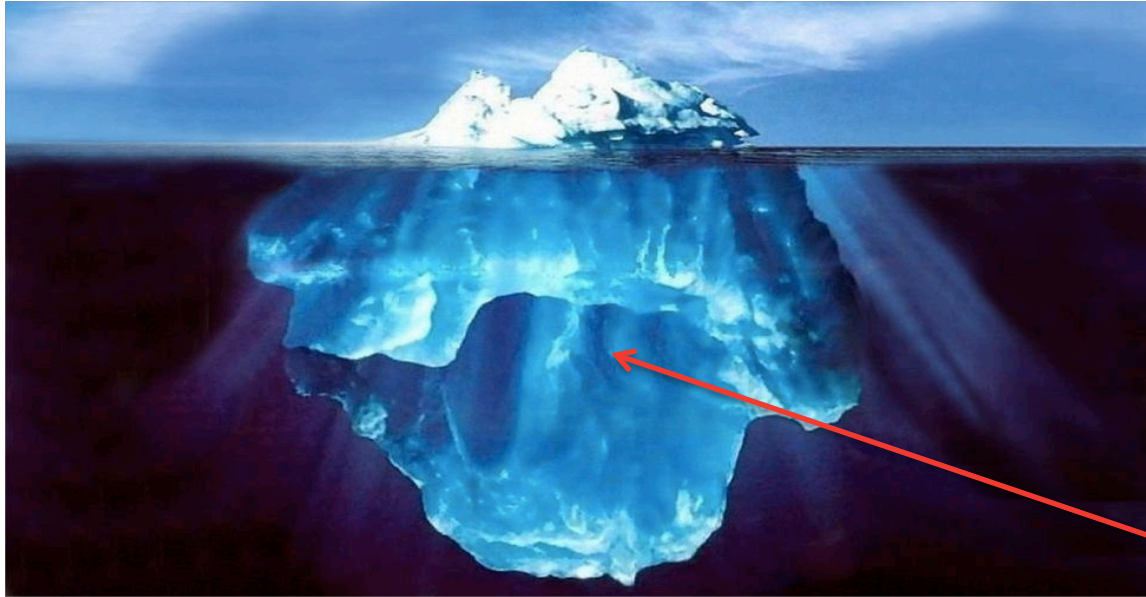
URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www.fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

Fitness ...



URL	Logistic Parameter
www.mercedesremotestart.com	2.97
www.girardtoyota.com	2.92
www.bluelineprop.com	2.91
www.toyotaoflaredo.com	2.70
www.pikespeakacura.com	2.44
www.motorcarsacura.com	2.39
www.hobsonrealtors.com	2.35
www.fayettevillencmls.com	2.35
www.coast2coastmixtapes.com	2.28
www.fwproperties.com	2.26
www.dealeron.com	2.25
www.smartfitnesscoach.com	2.23
www.pearsonlongman.com	2.21
www.pattersonauto.com	2.14
www.acuraofmodesto.com	2.11
www.f2it.co.uk	2.10
www.keyeshyundai.com	2.09
www.stevenscreekhyundai.com	2.06
www.columbusrealtors.com	2.05
www.whitmanathletics.org	2.05
www.napletonautosales.com	2.02
www.dealerelite.net	2.00
www.valuinsight.com	1.99
www.travel-shows.com	1.98
www.gorealestateagents.com	1.98

Automating Machine Learning



Infrastructure

Consumer Sensors Pipe
Data Into ..

Streaming Data Infrastructure Unlocking Access to 50 Trillion record Consumer
History..

Empowering
Scientists & Machines

Consumer Events:

- 100B Daily Events
- 20+ data integrations
- Clickstream
- App usage
- Ecommerce sales
- Cash register sales
- Precise Location

Context Data:

- User
- Device
- Location
- URL
- IP

Pre Processing Layer

- **Normalize**
- **Append**
- **Clean**
- **Primed for new Data Sources**

Extensible Rules Engine(proprietary software)

- **Identify** - events of interest
- **Process** - Define processes to invoke
- **Store** - route data to storage location

Scoring System

200 Million Devices Daily
Scored against 10,000+ models

Proprietary Augmentation Processes

- Crosswalk
- Homing
- Behavioral histories
- Location

Augmented Records

- Optimized for analysis
- Full consumer history
- Xx kb per record
- Customized for task

Machine Learning

- 100% Automated
- Millions of features
- 10,000 models daily

GeoSpatial Engine

- 10M points of interest

Data Scientists

- Research Playground
- Efficient prototyping of new products

Universal DataStore

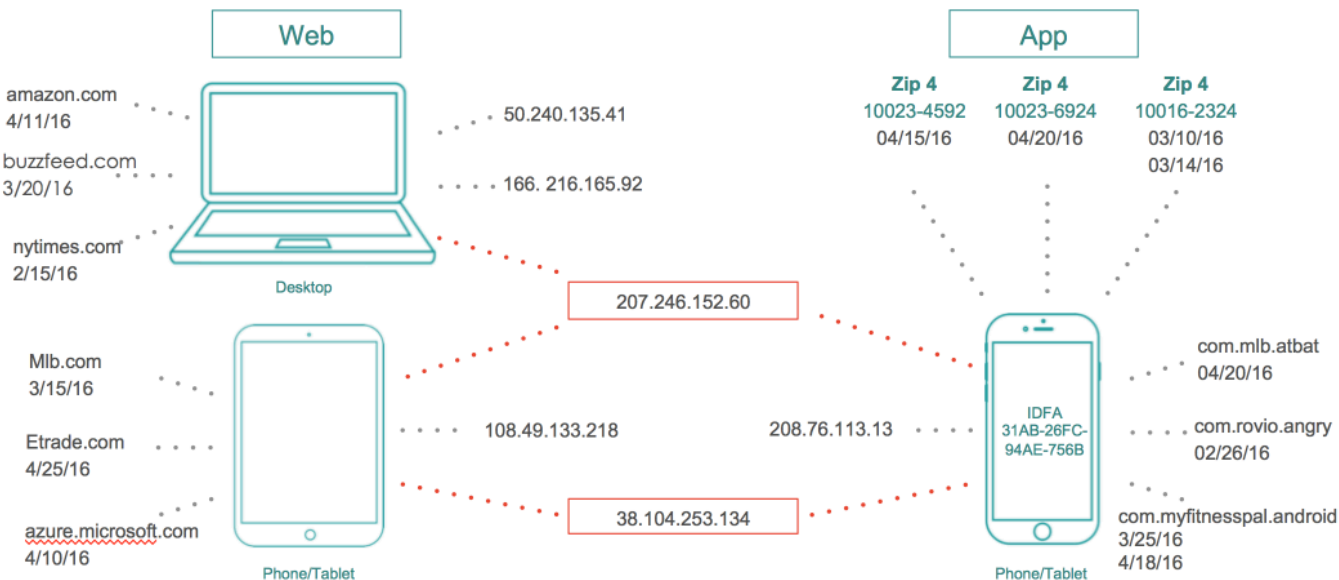
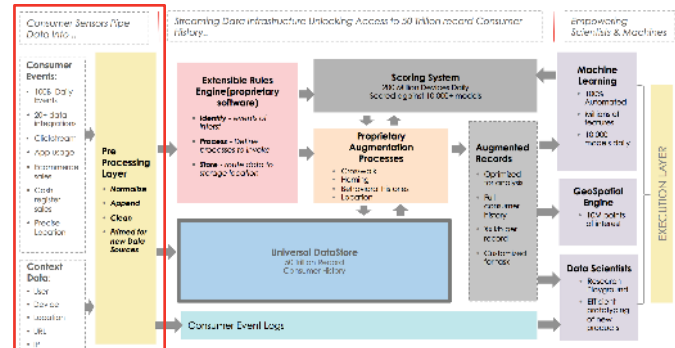
50 Trillion Record
Consumer History

Consumer Event Logs

EXECUTION LAYER

Consumer Events Data

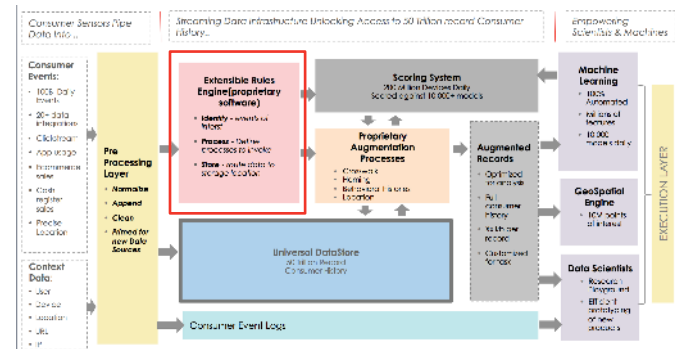
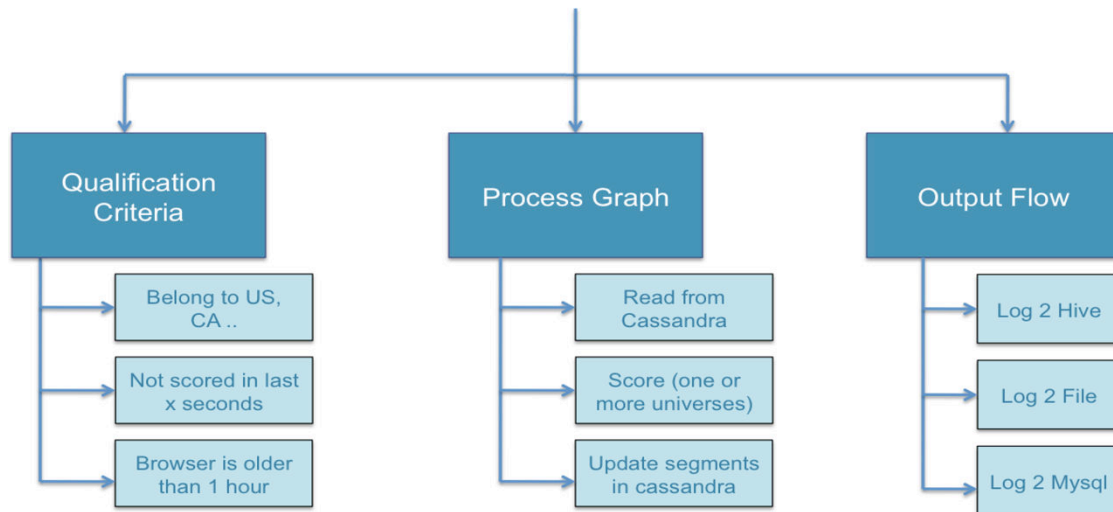
- Cleaning & consistent format
- Bot detection
- Appending contextual information



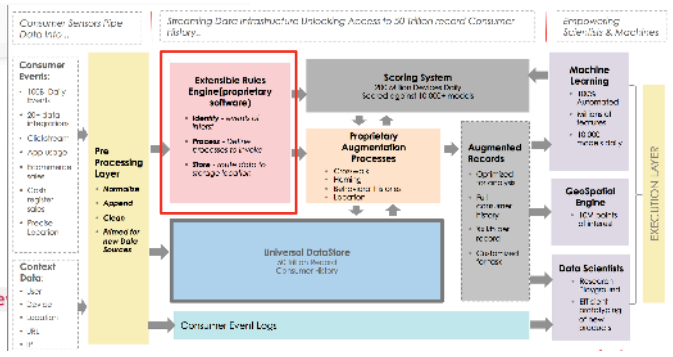
Sampler

- Extensible Rules Engine

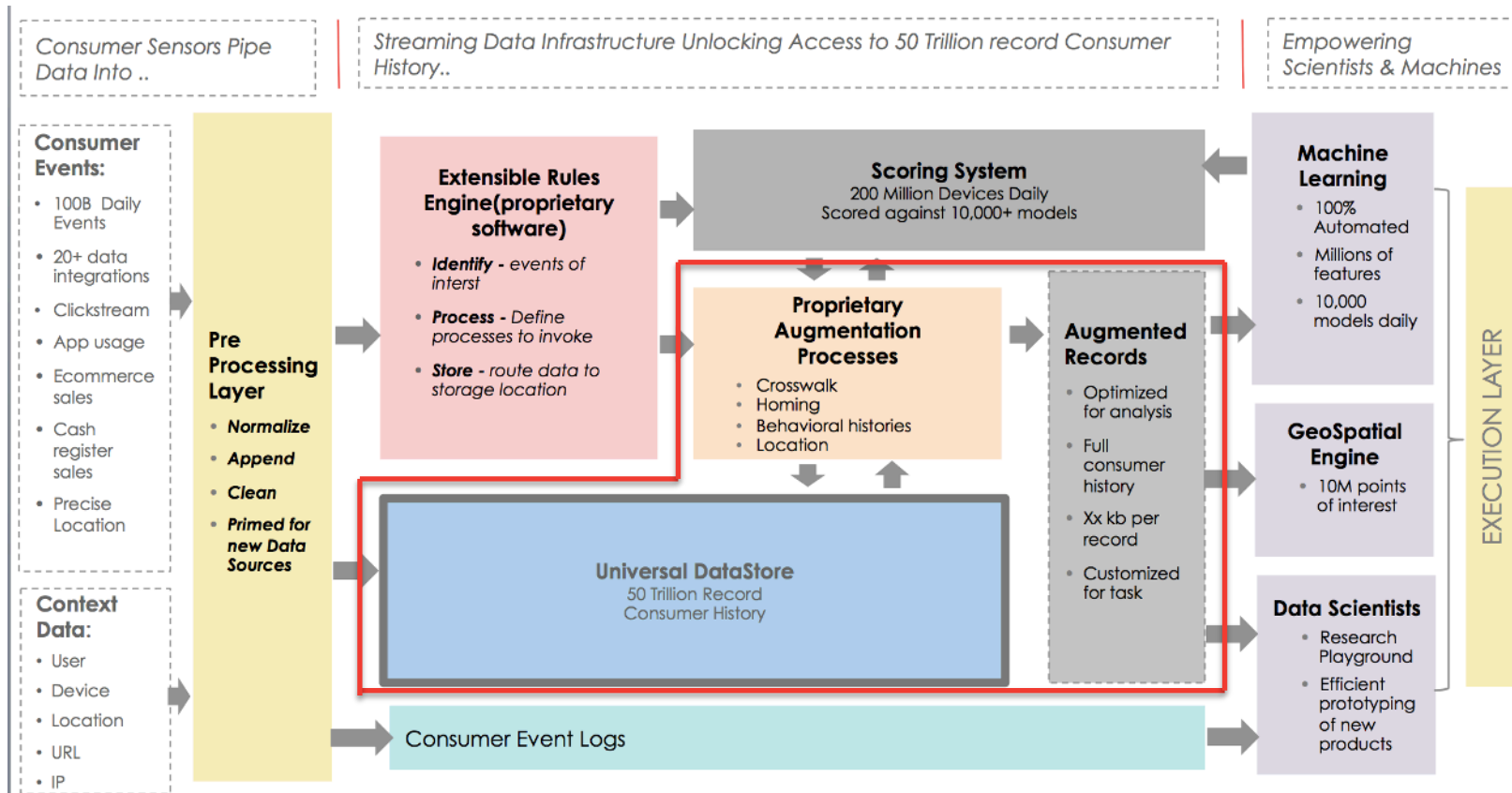
Samplers / Filters



id	112
desc	SGD RON Sampler
process	init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);feature-map(2);score(display,5:6:11:14:17:18,false,-1,false,true);cfp-to-kafka(cfpBaseline)
mode	ON
destination	
precondition	none
sampler_interval	86400
spec	{ "isQ": "return (event.eventType == 6 && event.audienceType == 72 && (event.getCountry() == \"US\" event.getCountry() == \"CA\" event.getCountry() == \"GB\" event.getCountry() == \"ZZ\" event.getCountry() == \"unknown\");", "onQ": "", "onR": "" }
type	groovyTemplate
version	3
id	113
desc	Fracture INS
process	init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);feature-map(2);cfp-to-kafka(cfpBaseline);
mode	ON
destination	
precondition	none
sampler_interval	0
spec	{ "isQ": "return (event.getBrowserId() % 1000 == 0) && event.getBrowserId() > 999999999 && event.environmentType != \"APP\";", "onQ": "", "onR": "" }
type	groovyTemplate
version	2
id	114
desc	Bid Sampler 0.125%
process	init-request;populate-cookies-from-req;refresh-ssc(adh,vstcnt,lbh,sglst,browser_prop);cfp-to-kafka(cfpBaseline)
mode	ON
destination	
precondition	none
sampler_interval	0
spec	{ "isQ": "return ((new Random()).nextInt(800) == 1 && event.environmentType != \"APP\" && event.eventType == 10 && (event.getCountry() == \"US\" event.getCountry() == \"CA\" event.getCountry() == \"GB\"));", "onQ": "", "onR": "" }
type	groovyTemplate
version	2



Event Data Enrichment



Augmented Event Data (Json)

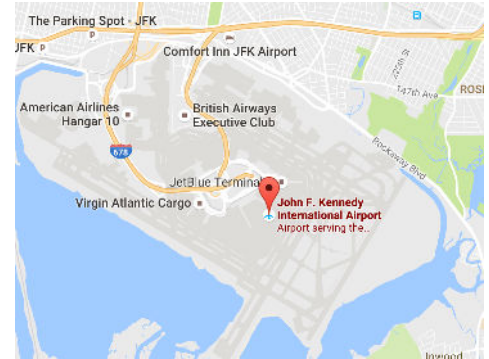
EventData	Data describing current event. Time, IP, etc.
GeoData	Info derived from precise physical location.
DeviceData	All data regarding the event device, including historical observations.
VisitData	List of historical visits by this device.
ScoresData	Thousands of propensity scores from the Dstillery machine learning system.
LinkedData	List of linked devices as well as list of linked visits.
SegmentData	Segments historically stored against device, as well as segments derived from all of the above augmented features.

Claudia Perlich

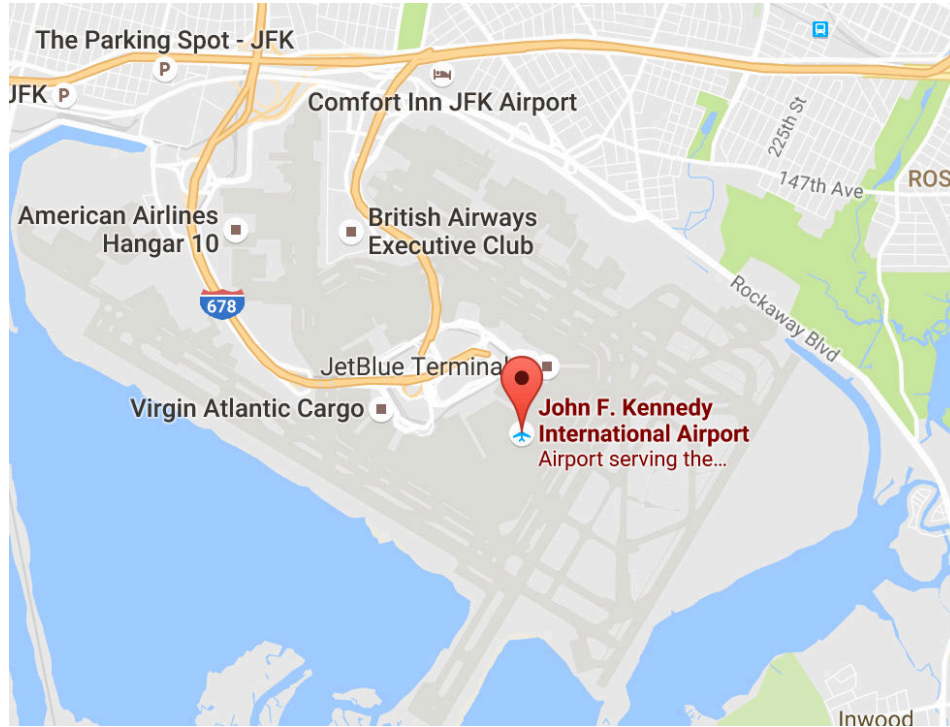
Thank You!

dstillery

Where do we find frequent traveler?



What do you think indicates people going to JFK?



URL

www.iglesiaelfaroinc.org
www.jumpseatnews.com
www.bluelineprop.com
www.ktxdtv.com
www.southjefffootball.org
www.unitedafa.org
www.parliamenthouse.com
www.yunghova.com
www.interlinetravel.com
www.aclin.org
www.swissport.com
www.gcsanc.com
www.swacu.org
www.airlinepilotcentral.com
www.homotrophy.com
www.beggfuneralhome.net
www.tvathletics.org
www.2shopper.com
www.nextmagazine.com
www.dailyjocks.com
www.pullzone.com
www.diamondoffshore.com
www.myerspolaris.com
www.ryandeyer.com
www.okllo.com
www.ifihadtochoose.com
www.ivoirmixdj.com

Logistic Parameter

2.38
2.25
2.21
2.14
2.1
2.09
2.07
2.06
2.03
2.03
2.03
2.01
2.01
1.97
1.97
1.94
1.92
1.91
1.91
1.87
1.87
1.86
1.86
1.86
1.84
1.83
1.83

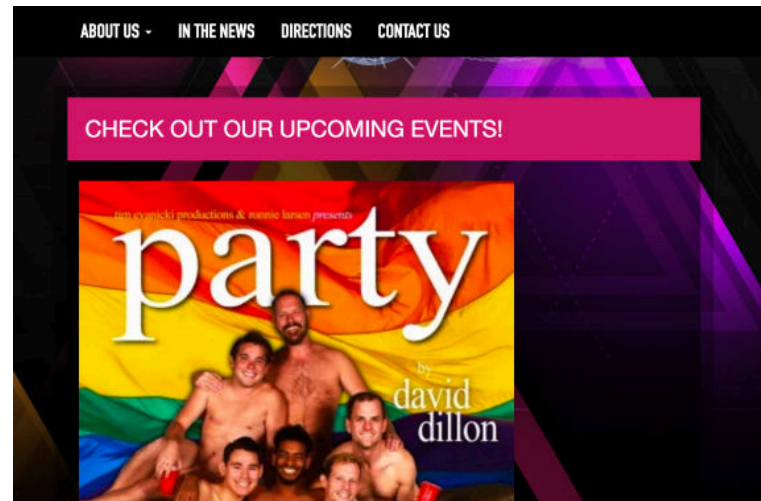
URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.okllo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83



URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.okllo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83

URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.okllo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83

URL	Logistic Parameter
www.iglesiaelfaroinc.org	2.38
www.jumpseatnews.com	2.25
www.bluelineprop.com	2.21
www.ktxdtv.com	2.14
www.southjefffootball.org	2.1
www.unitedafa.org	2.09
www.parliamenthouse.com	2.07
www.yunghova.com	2.06
www.interlinetravel.com	2.03
www.aclin.org	2.03
www.swissport.com	2.03
www.gcsanc.com	2.01
www.swacu.org	2.01
www.airlinepilotcentral.com	1.97
www.homotrophy.com	1.97
www.beggfuneralhome.net	1.94
www.tvathletics.org	1.92
www.2shopper.com	1.91
www.nextmagazine.com	1.91
www.dailyjocks.com	1.87
www.pullzone.com	1.87
www.diamondoffshore.com	1.86
www.myerspolaris.com	1.86
www.ryandeyer.com	1.86
www.okllo.com	1.84
www.ifihadtochoose.com	1.83
www.ivoirmixdj.com	1.83



RECENT ARTICLES

Page 1 of 1,038 [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [Last](#)

Sebastian K. for All American Guys
 SEPTEMBER 16TH, 2016 / HUNK, MODEL, PHOTOGRAPHY / NO COMMENTS

[Like](#)
[Share](#)
[Tweet](#)
[StumbleUpon](#)
[Pin It](#)

SUBSCRIBE

[Poets \(RSS\)](#)
[Comments \(RSS\)](#)

TROPHY : GET US

[Become a FAN](#)
[Twitter](#)

TROPHY::LISTS

VISIT LOVELY AD

INNduge Palm Springs

Swimsuits Allowed

Predict who is a traveler?



People who work at
JFK ...

