

A decorative graphic in the top-left corner consisting of a 5x8 grid of small, dark gray dots.Abstract geometric lines in the background, including a diagonal line from the top-right to the bottom-left and another from the bottom-left to the bottom-right, creating a sense of movement and structure.

# **Advanced Feature Flagging: It's All About The Data**



DAVE KAROW

Continuous Delivery Evangelist







## **FEATURE FLAG REVIEW**

A quick review of feature flags and rollout strategies

## **HOW YOU MEASURE MATTERS**

Don't believe everything you see

## **FLAGS + DATA = EXPERIMENTATION**

Attribution, calculation and analysis, automatic 24x7

01

02

03



# FEATURE FLAGS





**DEPLOY**

When you  
want to

**RELEASE**

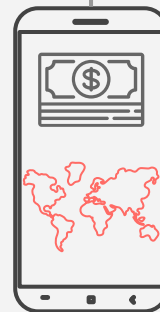
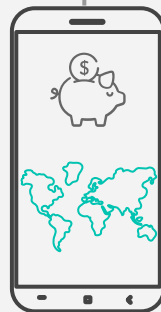
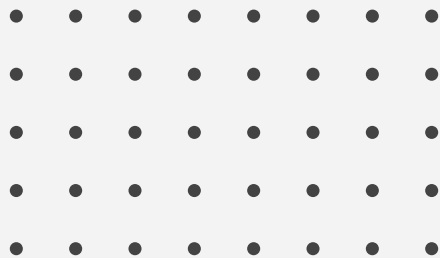
When you  
are ready





# `if{} else{} STATEMENT`

Separates code deploy  
from feature release





# TYPES OF FLAGS



**SWITCH**



**RAMPED**



**MULTI-VARIATE**





# ROLLOUT FRAMEWORK

## **SPEED**

How quickly can we reach a decision?



## **QUALITY**

How confident can we be of that decision?



## **RISK**

How can we minimize bad outcomes?



## **CREDIT**

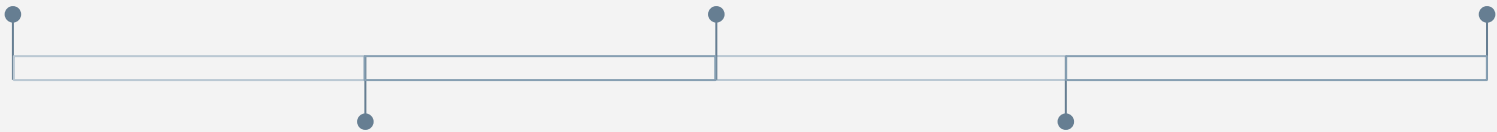
SQR: Balancing Speed, Quality and  
Risk in Online Experiments  
Ya Xu, et.al



# PHASES OF ROLLOUT

## DEPLOY

Code deployed,  
no exposure

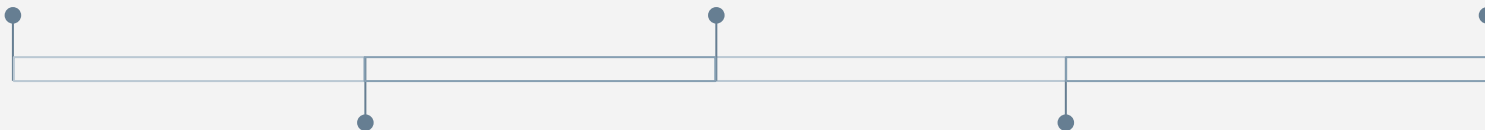




# PHASES OF ROLLOUT

## DEPLOY

Code deployed,  
no exposure



## ERROR MITIGATION

1-50% Ramp  
Identify bugs/crashes





# PHASES OF ROLLOUT

## DEPLOY

Code deployed,  
no exposure

## MEASURE

Maximum Power Ramp  
Understand impact

## ERROR MITIGATION

1-50% Ramp  
Identify bugs/crashes





# PHASES OF ROLLOUT

## DEPLOY

Code deployed,  
no exposure

## MEASURE

Maximum Power Ramp  
Understand impact

## RELEASE

Complete rollout

## ERROR MITIGATION

1-50% Ramp  
Identify bugs/crashes

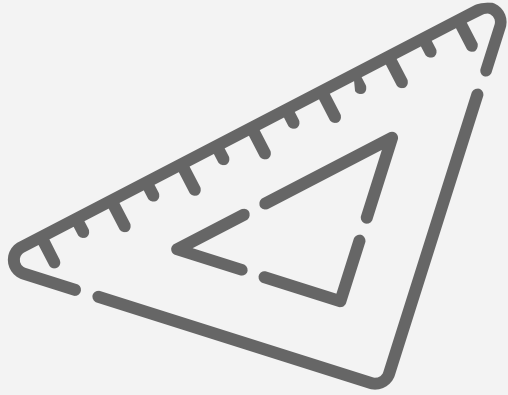
## SCALE MITIGATION

50-100% Ramp  
Identify scaling issues



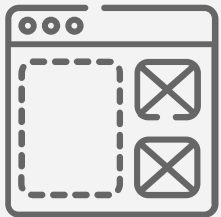


# HOW YOU MEASURE MATTERS





# DON'T BELIEVE EVERYTHING YOU SEE...



New Release



Metrics Change

**“Can’t we just  
change things  
and monitor  
what happens?”**



### Impressions Dropped only

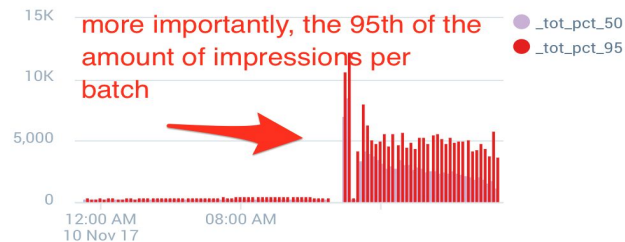
Last 3 Days



THERE IS NO DATA TO DISPLAY.  
SHOW IN SEARCH [🔍](#)

### Impressions Written

Last 24 Hours

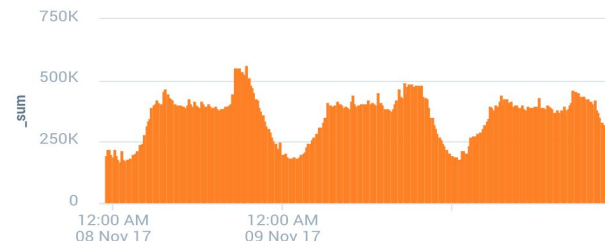


### Impressions writer write delays



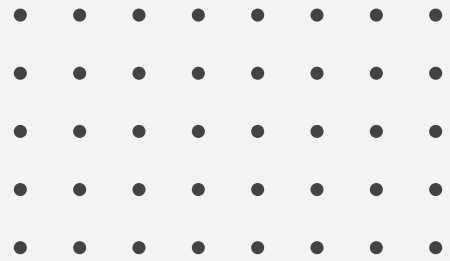
### Webhook impressions sent

Last 3 Days



# SYSTEMS DASHBOARD

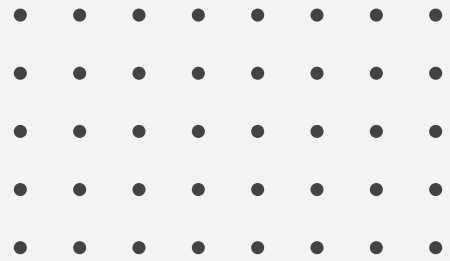




**FEATURE ENABLED TO 100%**







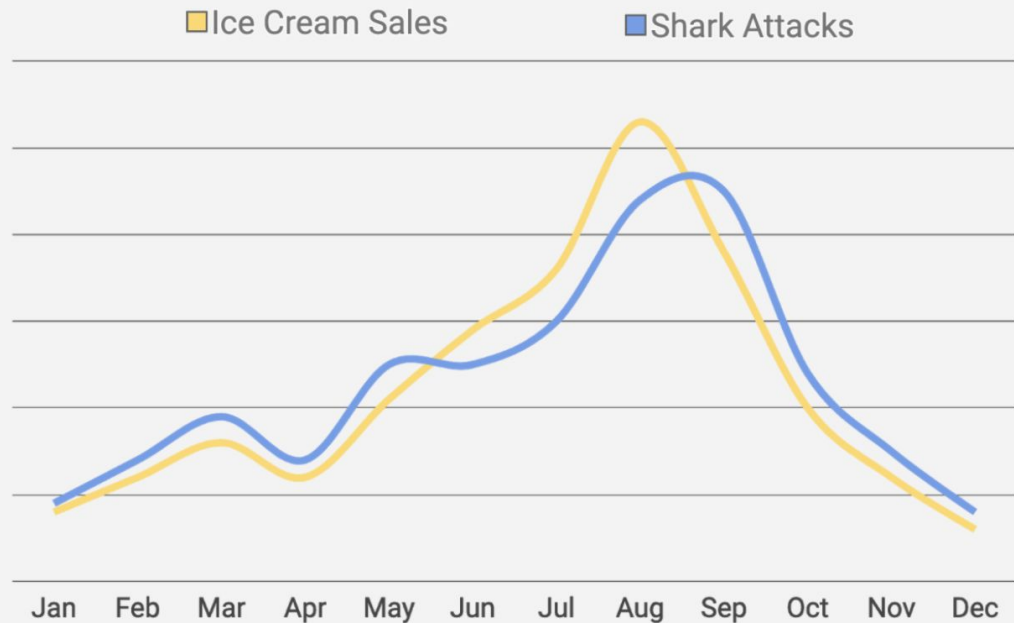
**FEATURE ENABLED TO 5%** →

← **FEATURE ENABLED TO 100%**





# CORRELATION IS NOT CAUSATION



**Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other**

**They are both caused by good weather, with lots of people at the beach, both eating ice cream and swimming in the sea**



Impressions Dropped only

Last 3 Days



THERE IS NO DATA TO DISPLAY.  
SHOW IN SEARCH [🔗](#)

Impressions Written

Last 24 Hours



15K  
10K  
5,000

more importantly, the 95th of the  
amount of impressions per  
batch

Impressions written

Last 3 Days

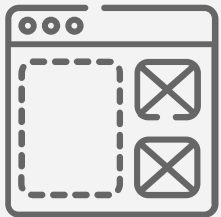


Not related to a feature change  
A customer was undergoing a DDoS

**SYSTEMS DASHBOARD**



# DON'T BELIEVE EVERYTHING YOU SEE...



New Release

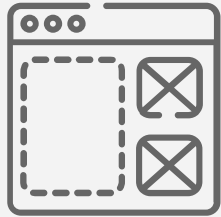


Metrics Change

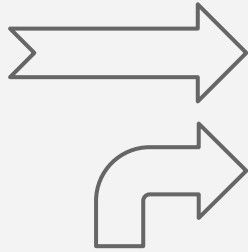
**“Can’t we just  
change things  
and monitor  
what happens?”**



# DON'T BELIEVE EVERYTHING YOU SEE...



New Release



Metrics Change



Everything else in  
the world

- **Product changes**
- **Marketing campaigns**
- **Global Pandemics**
- **Nice Weather**

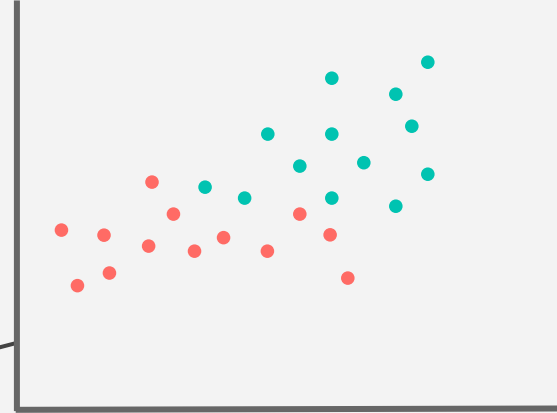
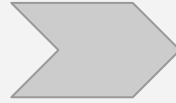
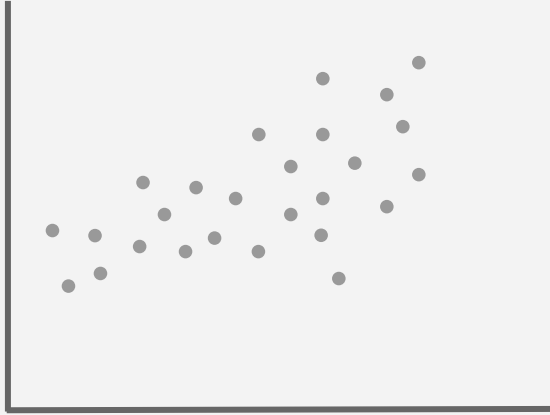


# MEASURING CAUSALITY



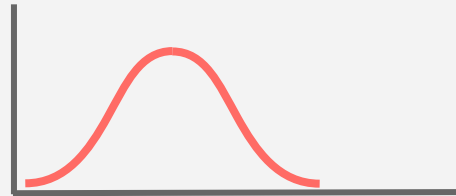
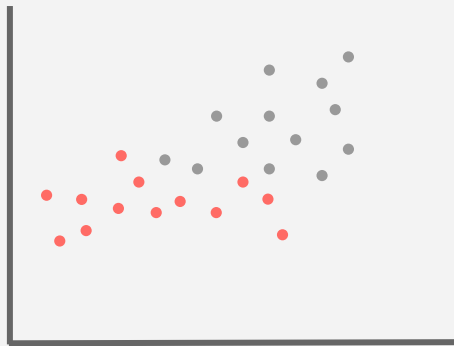
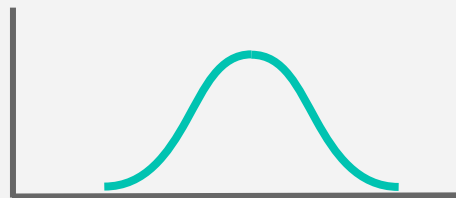


# ATTRIBUTION



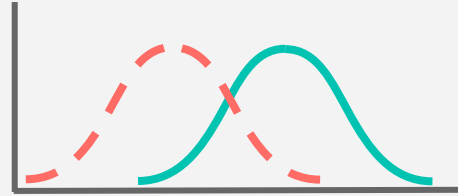
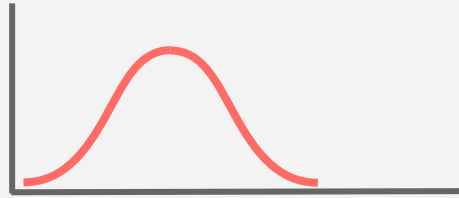
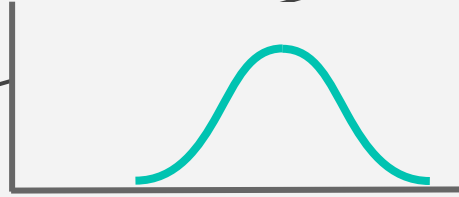


# CALCULATION





# ANALYSIS





## **TAGGING METRICS**

Most dashboarding tools allow tagging of data for segmentation



## **AD-HOC ANALYSIS**

Storing feature attribution in BI database for querying



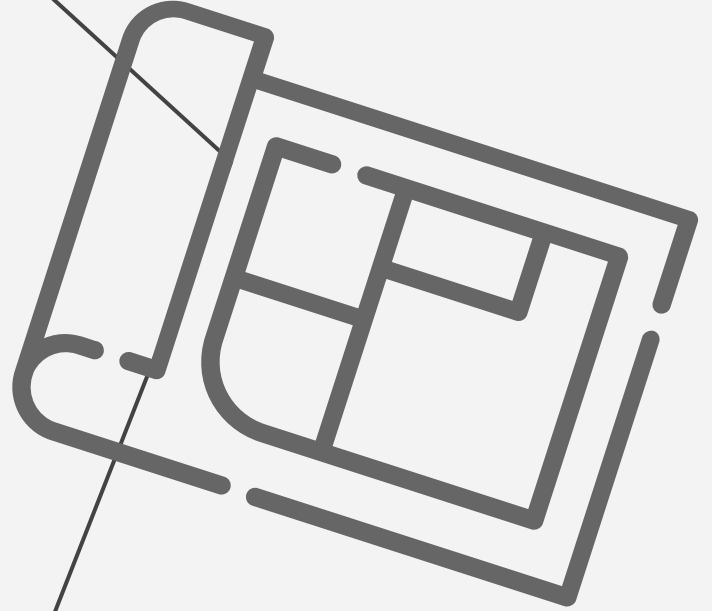
## **EXPERIMENTATION PLATFORM**

Homegrown or SaaS offerings provide both data collection and statistical analysis





**FLAGS + DATA =  
EXPERIMENTATION**





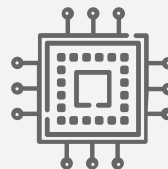
- 
- 
- 
- 
- 
- 
- 

TARGETING SYSTEM

TELEMETRY SYSTEM

STATISTICAL ENGINE

MANAGEMENT CONSOLE

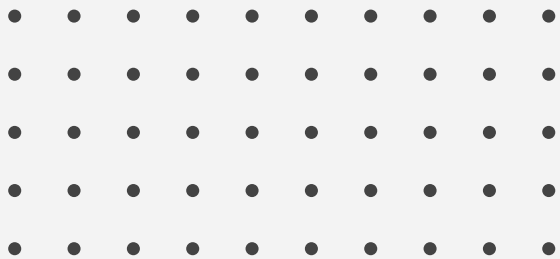


# ANATOMY OF AN EXPERIMENTATION PLATFORM

## CREDIT

Understanding Experimentation  
Platforms [Aijaz, Stuart, Jewkes]

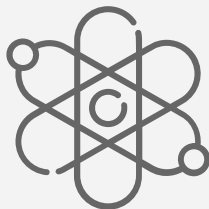




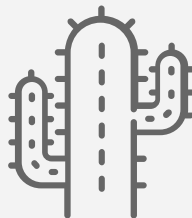
## TARGETING SYSTEM



FAST



RANDOM



STICKY

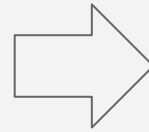
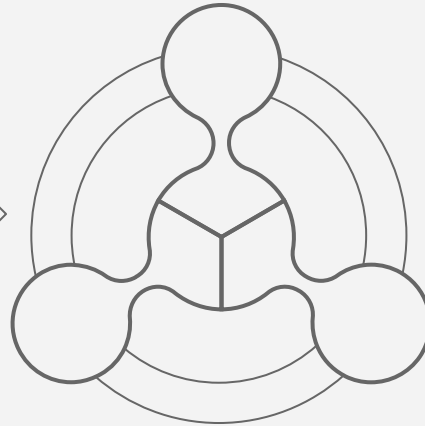
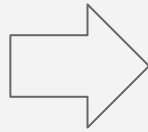


RELIABLE

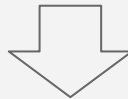


# TARGETING SYSTEM

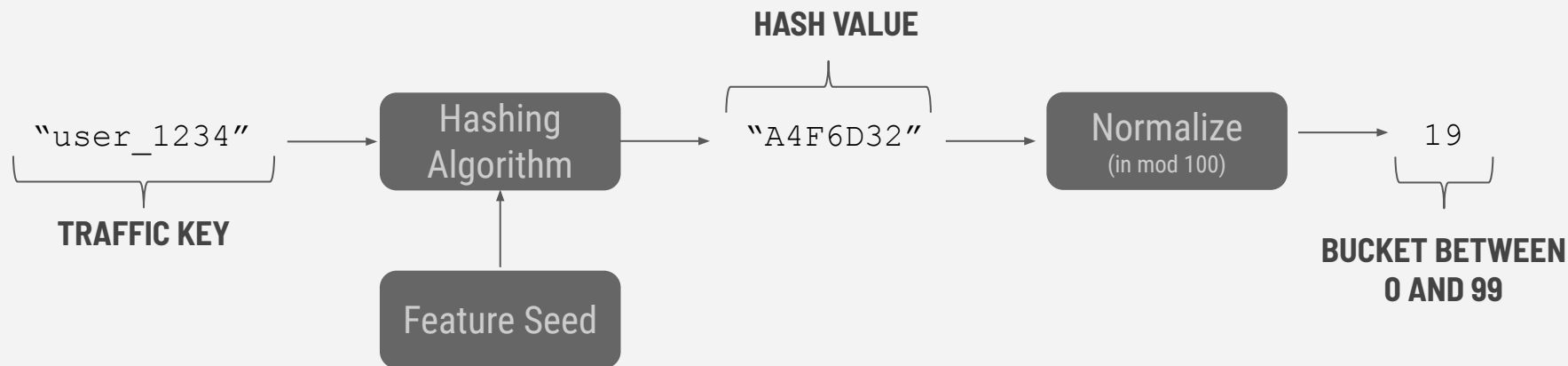
FEATURE  
TRAFFIC KEY



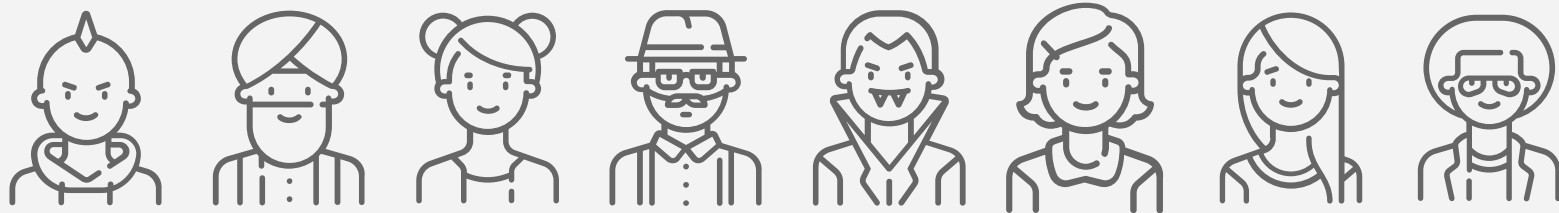
ASSIGNMENT



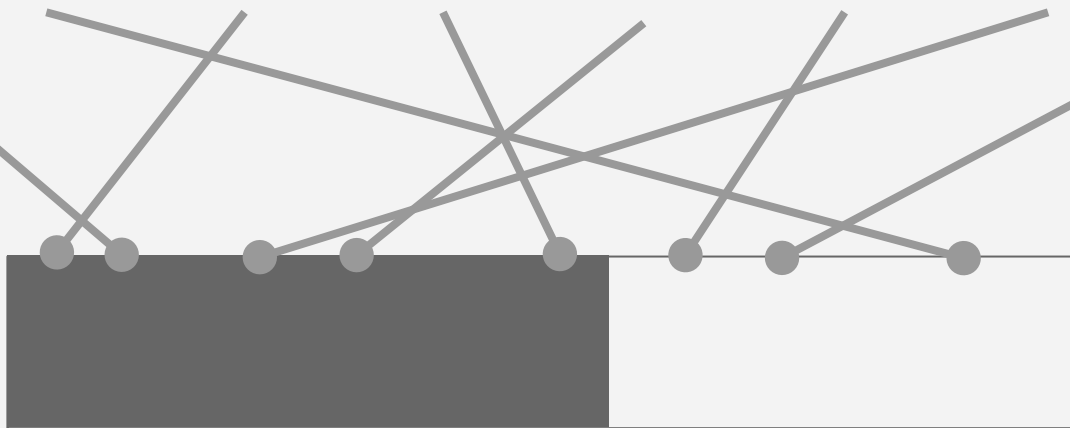








12 91 6 54 34 63 25 70





# TELEMETRY SYSTEM



track (String key, String eventType, double value)

## WRAPPER

Centralize tracking across tools

Ensure reliable delivery

Identify and exclude  
malicious traffic

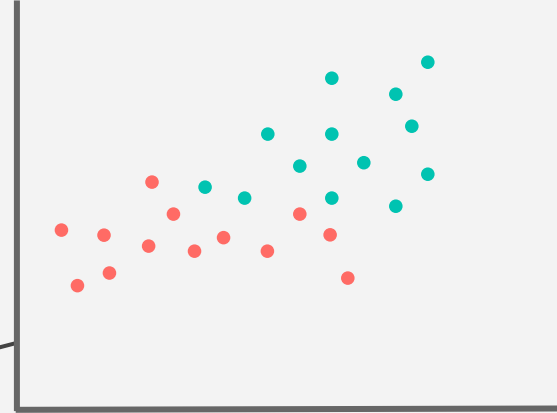
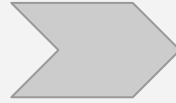
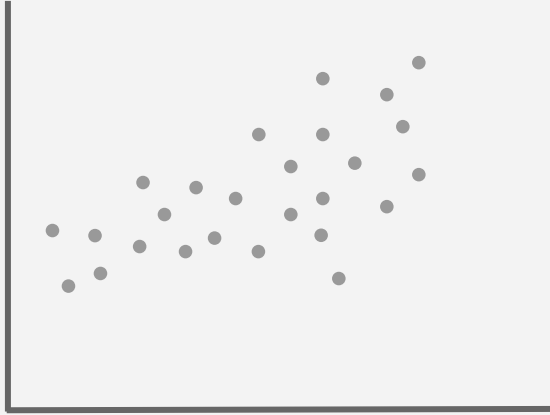
## SERVICE

Receive batches of events

Store in warehouse for  
processing

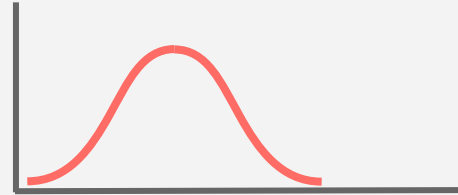
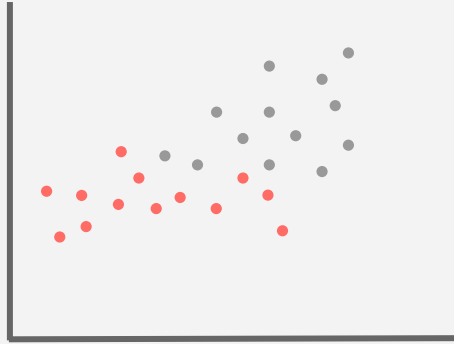
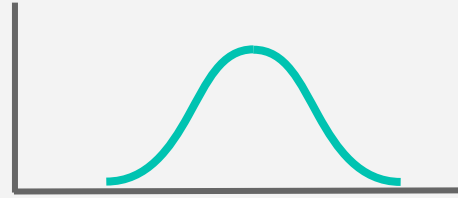
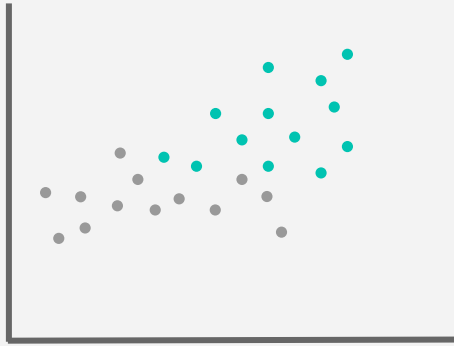


# STATISTICAL ENGINE



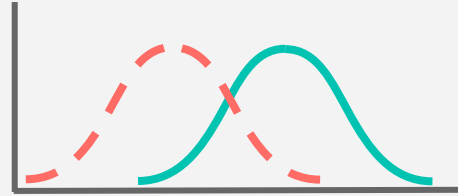
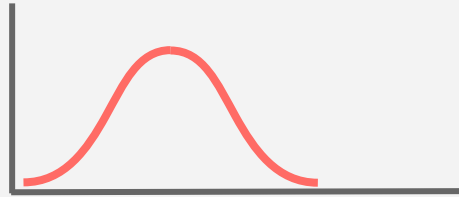
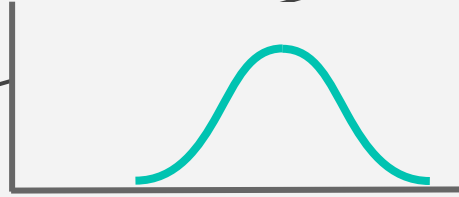


# STATISTICAL ENGINE



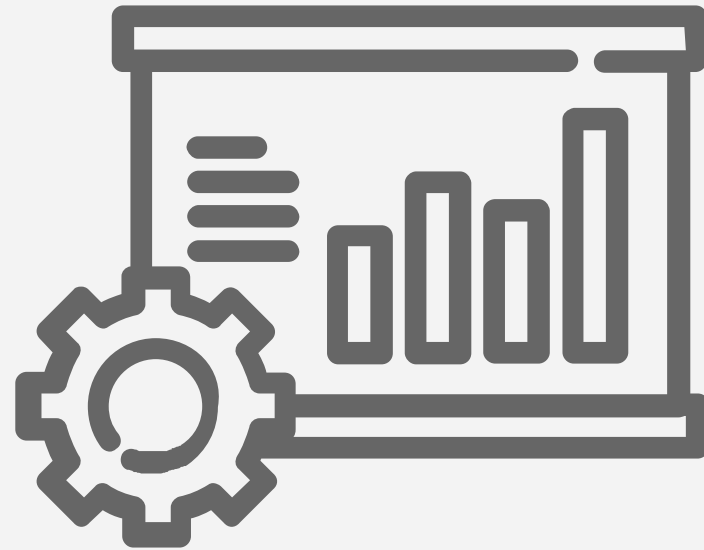


# STATISTICAL ENGINE

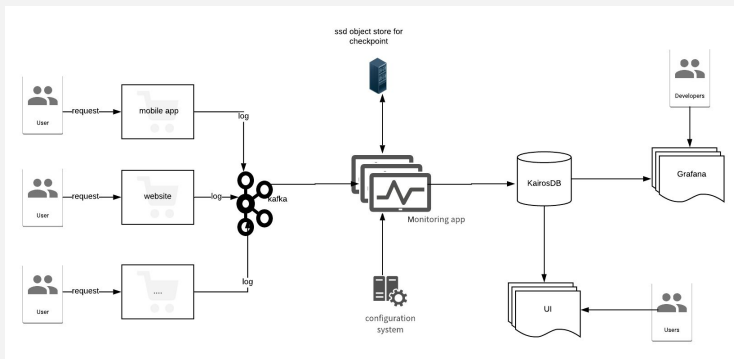




# MANAGEMENT CONSOLE

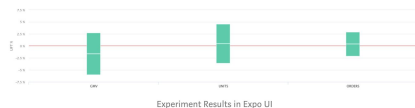






### It's All About the Data

When someone runs an A/B test, they expect results, and whether they are positive or negative, they need to be correct. One challenge we faced in building our own platform was building trust in the results. With the support of a strong product analytics team, and a revamped data pipeline, we were able to achieve that.



# EXAMPLE IMPLEMENTATION

Walmart EXPO

## Test to learn

- **Purpose:** Understand customer behavior and validate or invalidate a hypothesis
- Feature likely only launches if it is better than the production experience
- Ex. Guest Cart

## Test to launch

- **Purpose:** Mitigating risk by phasing out the rollout to customers, and ensuring no bugs are introduced
- Feature is necessary to launch for the business. Will launch if it is doing no harm
- Ex. ADA

EXPO

Demo Experiment

PREVIEWABLE

CREATED ON

SCHEDULED END DATE

EXPERIMENT TOTAL

CONTROL

TREATMENT 1

ASSIGNED LAYER

06/22/2018 05:33 PM PDT

None

0%

0%

0%

vtc - Layer 18

Watch

Start

Back to Draft

Cancel

Experiment Details

Factors & Variations

Revision History

Qualifying Conditions

Any

Base Templates

Hash Unit & Traffic Layer

vtc - Layer 18

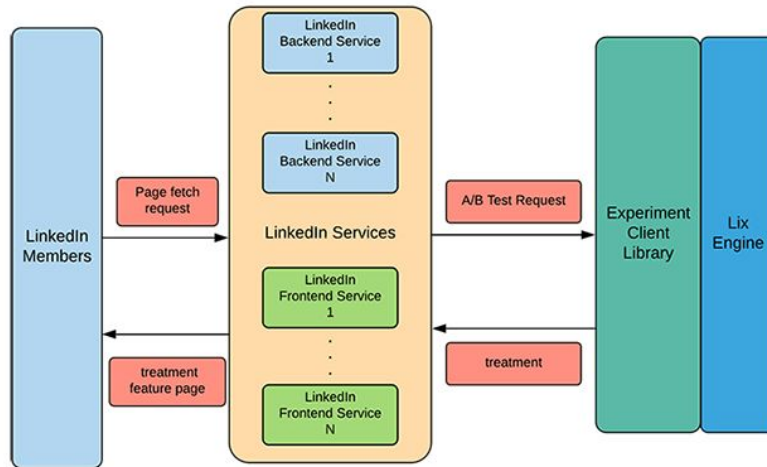
Mutual Exclusion Labels

LAYER\_18

Variations

Control	Variation Name	Variation Spec	Treatment	% of Layer Traffic	Bucket Assignment	Actions
Control	Demo Experiment Control	EYa7w	Modules: 0 Configs: 0 Origin: None Header/Cookie: 0	0%		
Treatment 1	Demo Experiment Variation	pX7PP	Modules: 0 Configs: 0 Origin: None Header/Cookie: 0	0%		





Students	Job Seekers	Everyone Else
Treatment	Treatment	Untreated
Control	Control	

# EXAMPLE IMPLEMENTATION

Linked In LiX

**LiX Failed on Site Speed**

Run Time: 2017-02-28 17:43:35 PT

**xmc.cache.V2.disable** [1622829](#)  
Experiment Id

LIX Key

**2017-02-28 14:43:20** **2017-02-28 17:43:20**  
Start (PT) End (PT)

**1**  
Rules Failed

**Failed Rules**

Page Key / Dimensions	Root Cause / Metric	Delta
oz-winner Geo: in Segment: 1 Treatment: treatment	Server Issue 50 pct. <a href="#">Harrier Debug Page</a>	197 ms (5.45%)

\* Run Time is the email sending time. Start and End is the time range of collecting analysis data.

[See Analysis Result](#)





Circuit Breaker Automated System

Technical/Setup issues

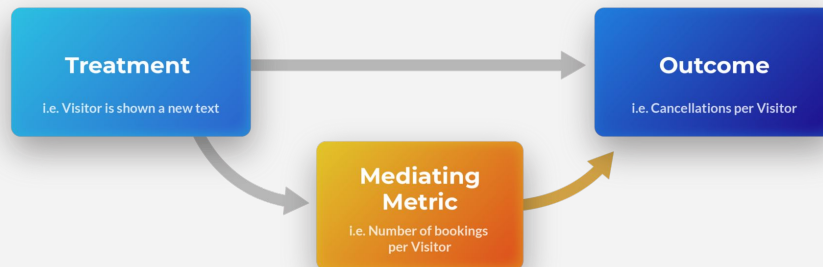
Too many errors and warnings.

variant	errors	warnings
base	0.0%	0.0%
v1	47.3%	81.8%

The experiment has been stopped to make sure we are not displaying a broken page to our customers. [Read more about automatic stopping.](#)

3 days ago

Back View helpful



request_warning_count - Number of warnings triggered							code 🔍 link 🔗
717707 (1h9m)	Exp. visitors	Goal server requests	Data validity	Sum	Exp. average	Goal average	Bar Graph
base (50%)	31 167 (49.66%)	22 (13.41%)	[BASE]	24 warnings	$770.0452 \times 10^{-6}$ warnings	1.0909 warnings	
variant 1 (50%)	31 594 (50.34%)	142 (86.59%)	More data points in this variant. This is bad, and the values are no longer comparable.	143 warnings	$4.5262 \times 10^{-3}$ warnings	1.007 warnings	

# EXAMPLE IMPLEMENTATION

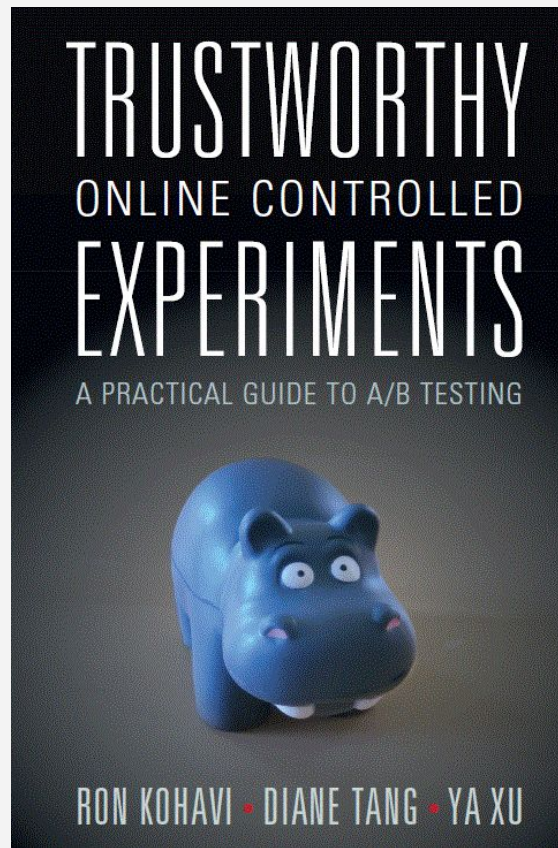
Booking.com



## LESSONS LEARNED

@ Microsoft + Google + LinkedIn

<https://experimentguide.com/>







## KILL THE RELEASE NIGHT

Decoupling deployment from release  
make release rituals go away.



## AUTOMATE DELIVERY WITH DATA

Independent feature rollouts are  
orchestrated by data, not people.



## EVERY FEATURE IS AN EXPERIMENT

Targeted feature rollout allows for rapid  
A/B tests and customer insights.



# THANKS!



Let's move to discussion!

[linkedin.com/in/davekarow](https://linkedin.com/in/davekarow)  
[dave.karow@split.io](mailto:dave.karow@split.io)  
[@davekarow](https://twitter.com/davekarow)

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**