



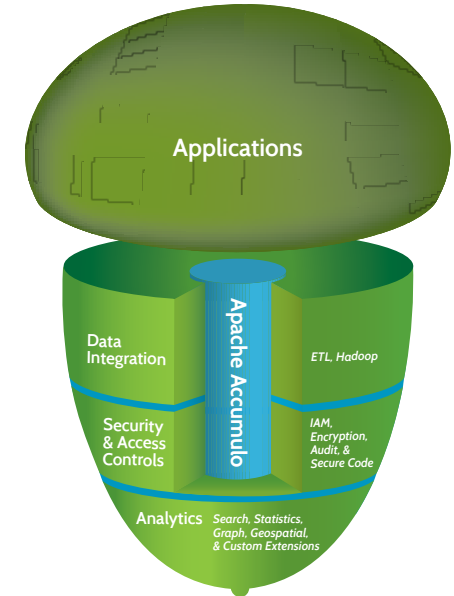
**sqrrl**

Secure. Scale. Adapt

**Bloomberg: NYC Data Business Event**



- Big Data & Security Experts
- Proven Analytic Expertise
- Apache Accumulo Creators, Committers, and Contributors
- sqrri analytics



# Enthusiasm for Big Data

Secure. Scale. Adapt.

## TAMING BIG DATA

BIG DATA INCLUDES DATA SETS WHOSE SIZE AND TYPE MAKE THEM IMPRACTICAL TO PROCESS AND ANALYZE WITH TRADITIONAL DATABASE TECHNOLOGIES

**BIG DATA MARKET FORECAST**  
\$ US BILLIONS

| Year        | 2009  | 2010  | 2011   | 2012   | 2013   | 2014   | 2015 | 2016 | 2017 |
|-------------|-------|-------|--------|--------|--------|--------|------|------|------|
| Market Size | \$1.1 | \$1.5 | \$18.9 | \$32.1 | \$48.8 | \$83.4 |      |      |      |

**GLOBAL MENTIONS OF "BIG DATA"**  
1211.34% INCREASE OVER BASELINE AVERAGE

"IT'S NO LONGER HARD TO FIND THE ANSWER TO A GIVEN QUESTION; THE HARD PART IS FINDING THE RIGHT QUESTION AND AS QUESTIONS EVOLVE, WE GAIN BETTER INSIGHT INTO OUR ECOSYSTEM AND OUR BUSINESS." - KEVIN WEIL

**CURRENT USES ACROSS THE BOARD**

- SENTIMENT ANALYSIS
- FRAUD DETECTION
- RISK MODELING
- RECOMMENDATION ENGINE
- NETWORK MONITORING
- CUSTOMER EXPERIENCE ANALYTICS
- MARKETING CAMPAIGN ANALYSIS
- CUSTOMER CHURN ANALYSIS
- RESEARCH AND DEVELOPMENT
- SOCIAL GRAPH ANALYSIS

**Case Studies:**

- facebook: stores, accesses and analyzes 30+ PETABYTES of user-generated data
- Linked in: processes and mines PETABYTES of user data to power "People You May Know"
- amazon: crunches click-stream and historical user data to recommend products
- Atamai: analyzes 75 MILLION events per day to better target advertisements
- JPMORGAN CHASE & CO.: analyzes web logs, transaction data, and social media to detect fraudulent activity
- Treato: tags Big Data to help researchers and physicians better determine patient treatments

**DECODING THE HUMAN GENOME USED TO TAKE TEN YEARS. IT CAN NOW BE DONE IN 7 DAYS.**

**THE OBAMA ADMINISTRATION IS INVESTING \$200 MILLION IN BIG DATA RESEARCH PROJECTS.**

**hadoop**: open source framework for storing, processing and analyzing massive amounts of distributed, multi-structured data

**MPP Analytic Database**

Big Data is the new definitive source of competitive advantage across all industries

[wikibon.org/bigdata](http://wikibon.org/bigdata)

## EXPLODING DATA

THE POTENTIAL OF BIG DATA

**WHAT IS BIG DATA?**

- 4 BILLION GB OF DATA ARE CREATED BY THE AVERAGE PERSON DAILY
- 235 TERABYTES OF DATA ARE STORED ON THE AVERAGE PERSON'S HARD DRIVE
- 40% OF THE PRODUCTION COST OF AN ALIEN SCIENCE FICTION FILM IS SPENT ON VISUAL EFFECTS
- 5% OF THE PRODUCTION COST OF A FILM IS SPENT ON VISUAL EFFECTS

**AMOUNT OF STORED DATA BY SECTOR**

| Sector                   | Amount |
|--------------------------|--------|
| GOVERNMENT               | 41B    |
| RETAIL                   | 38B    |
| FINANCIAL SERVICES       | 37B    |
| TELECOM                  | 35B    |
| ENERGY                   | 35B    |
| MANUFACTURING            | 34B    |
| TRANSPORTATION           | 32B    |
| HEALTH CARE              | 31B    |
| UTILITIES                | 31B    |
| EDUCATION                | 29B    |
| INSURANCE                | 29B    |
| ADVERTISING              | 27B    |
| MARKETING                | 27B    |
| CONSTRUCTION             | 26B    |
| PHARMACEUTICALS          | 24B    |
| SCIENTIFIC RESEARCH      | 23B    |
| RESEARCH AND DEVELOPMENT | 23B    |
| PROFESSIONAL SERVICES    | 21B    |
| TELEVISION               | 21B    |
| SPORTS AND ENTERTAINMENT | 21B    |
| MEDIA                    | 21B    |
| TELEPHONE                | 21B    |
| RESEARCH AND DEVELOPMENT | 21B    |
| RESEARCH AND DEVELOPMENT | 21B    |
| RESEARCH AND DEVELOPMENT | 21B    |
| RESEARCH AND DEVELOPMENT | 21B    |
| RESEARCH AND DEVELOPMENT | 21B    |

**WHAT'S IN THE DATA?**

- HEALTH: \$300 BILLION
- RETAIL: \$210 BILLION
- FINANCIAL SERVICES: \$170 BILLION
- TELECOM: \$147 BILLION
- ENERGY: \$147 BILLION
- MANUFACTURING: \$147 BILLION
- TRANSPORTATION: \$147 BILLION
- HEALTH CARE: \$147 BILLION
- UTILITIES: \$147 BILLION
- EDUCATION: \$147 BILLION
- INSURANCE: \$147 BILLION
- ADVERTISING: \$147 BILLION
- MARKETING: \$147 BILLION
- CONSTRUCTION: \$147 BILLION
- PHARMACEUTICALS: \$147 BILLION
- SCIENTIFIC RESEARCH: \$147 BILLION
- RESEARCH AND DEVELOPMENT: \$147 BILLION
- PROFESSIONAL SERVICES: \$147 BILLION
- TELEVISION: \$147 BILLION
- SPORTS AND ENTERTAINMENT: \$147 BILLION
- MEDIA: \$147 BILLION
- TELEPHONE: \$147 BILLION
- RESEARCH AND DEVELOPMENT: \$147 BILLION
- RESEARCH AND DEVELOPMENT: \$147 BILLION
- RESEARCH AND DEVELOPMENT: \$147 BILLION
- RESEARCH AND DEVELOPMENT: \$147 BILLION

**CAN WE CAPTURE THE FULL POTENTIAL OF BIG DATA?**

- DATA POLICIES
- DATA ACCESS
- TECHNOLOGY
- TALENT

## EVERY MINUTE of the DAY

- THE MOBILE WEB RECEIVES 217 NEW USERS.
- YOUTUBE USERS UPLOAD 48 HOURS OF NEW VIDEO.
- EMAIL USERS SEND 204,166,667 MESSAGES.
- GOOGLE RECEIVES OVER 2,000,000 SEARCH QUERIES.
- FACEBOOK USERS SHARE 684,478 PIECES OF CONTENT.
- CONSUMERS SPEND \$272,070 ON WEB SHOPPING.
- TWITTER USERS SEND OVER 100,000 TWEETS.
- APPLE RECEIVES ABOUT 47,000 APP DOWNLOADS.
- BRANDS & ORGANIZATIONS ON FACEBOOK RECEIVE 34,722 "LIKES."
- TUMBLR BLOG OWNERS PUBLISH 27,778 NEW POSTS.
- INSTAGRAM USERS SHARE 3,600 NEW PHOTOS.
- Flickr USERS ADD 3,125 NEW PHOTOS.
- NEW WEBSITES ARE CREATED. 571
- FOURSQUARE USERS PERFORM 2,083 CHECK-INS.
- WORDPRESS USERS PUBLISH 347 NEW BLOG POSTS.
- 3.8 PETABYTES
- 60%
- 50%
- 17%
- 13%

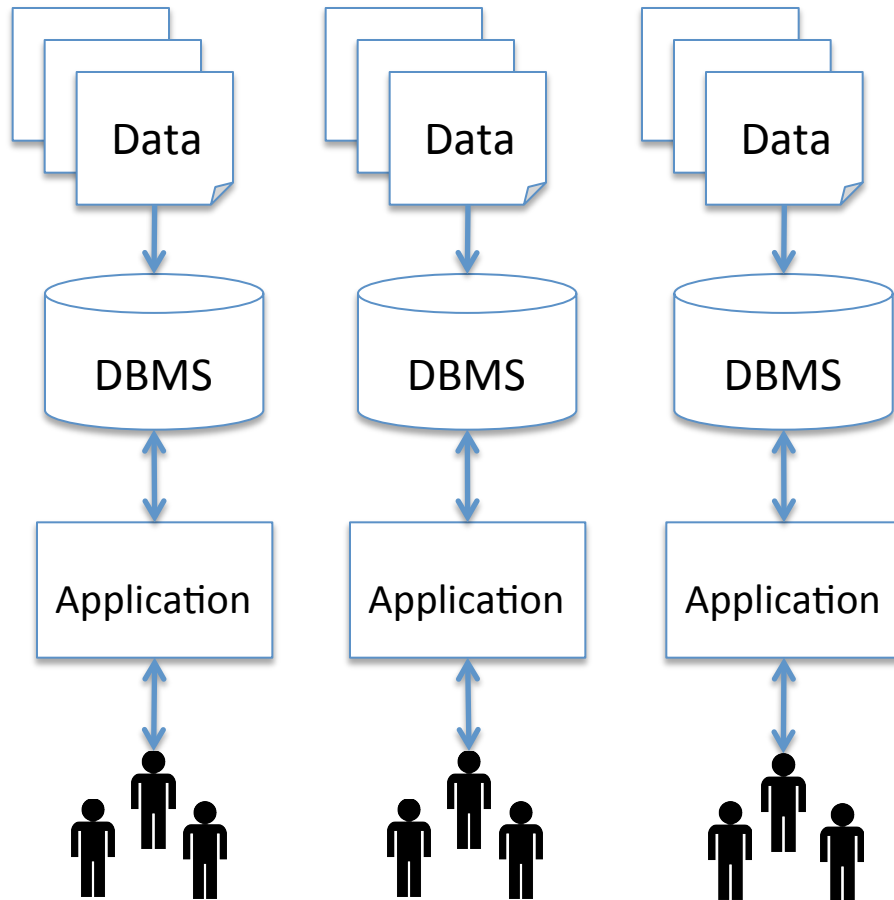
# Today's Data Environment

Secure. Scale. Adapt.

- Data poor environment in which we have long operated
- Continued inferences and limits on data insights

# The Problem

Secure. Scale. Adapt.



- Data optimized for individual business lines and specific applications
- Vertical integration reduces risks
- But, ***fractured data sets***:
  - Reduce the opportunities to drive new key insights from enterprise holdings
  - Increase the complexity of managing data and security

# Big Data Success?

Secure. Scale. Adapt.

Google

jeff lewis  
jeff lewis jenni  
jeff lewis net worth  
jeff lewis flipping out

Ad related to jeff lewis

Shop Jeff Lewis Art - 20x200.com  
www.20x200.com/Jeff-Lewis  
Limited edition art by Jeff Lewis now at 20x200.  
\$10 off \$50 orders with code NEWART - Gift Guides - Jane Mount: Ideal Bookshelf

**Jeff Lewis**  
www.jefflewisdesign.com/  
Official site provides service information, gallery, auction, biography, press, and contact details.  
Gallery - Contact - Services - About

'Flipping Out' Lawsuit: Jeff Lewis Suing Assistant Jenni Pulos Over ...  
www.huffingtonpost.com/\_/flipping-outlawsuit-jeff-lewis-jen-...  
Nov 15, 2012 - Share on Google+. Flipping Out Lawsuit Jeff Lewis Jenni Pulos.  
"Flipping out" star Jeff Lewis suing Jenni Pulos over new book. Get TV Alerts: ...

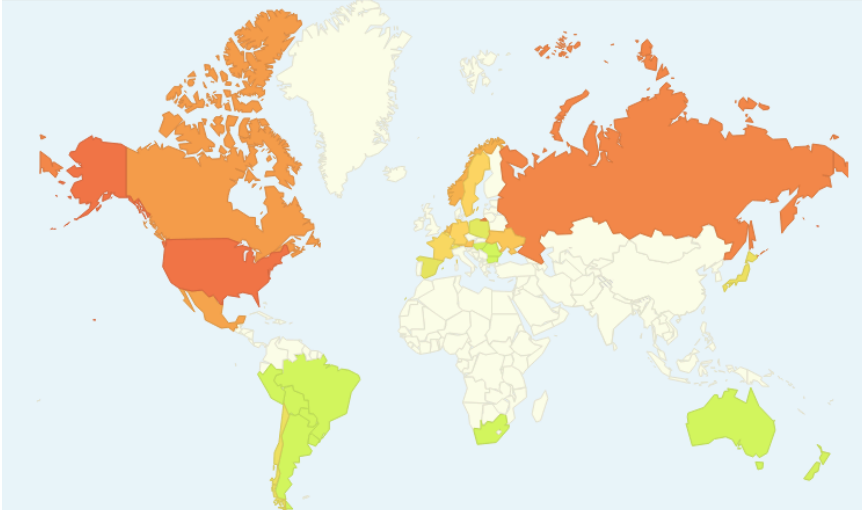
**Flipping Out**

Flipping Out is a reality television series that debuted on July 31, 2007 on Bravo in the United States and HGTV in Canada. Subsequently, it debuted on cable network Arena in Australia on September 4, 2009.

Wikipedia

First episode: July 31, 2007  
Genre: Reality television  
Cast: Chris Elwood, Jeff Lewis, Ryan Brown  
Networks: CNBC, Bravo

People also search for



NATE SILVER ON WHAT OBAMA SHOULD DO NEXT, P. 44

DON'T MOCK THE ARTISANAL PICKLE MAKERS, P. 14

A SANNNY'S VIEW OF THE WORLD, P. 47

MANAGEMENT TIPS FROM 'DOWNTON ABBEY,' P. 52

GREECE CONFRONTS ITS SPARTAN FUTURE, P. 38

10 Other new things to read in *Magazine*: Nine Americas, P. 42

## The New York Times Magazine

February 19, 2012



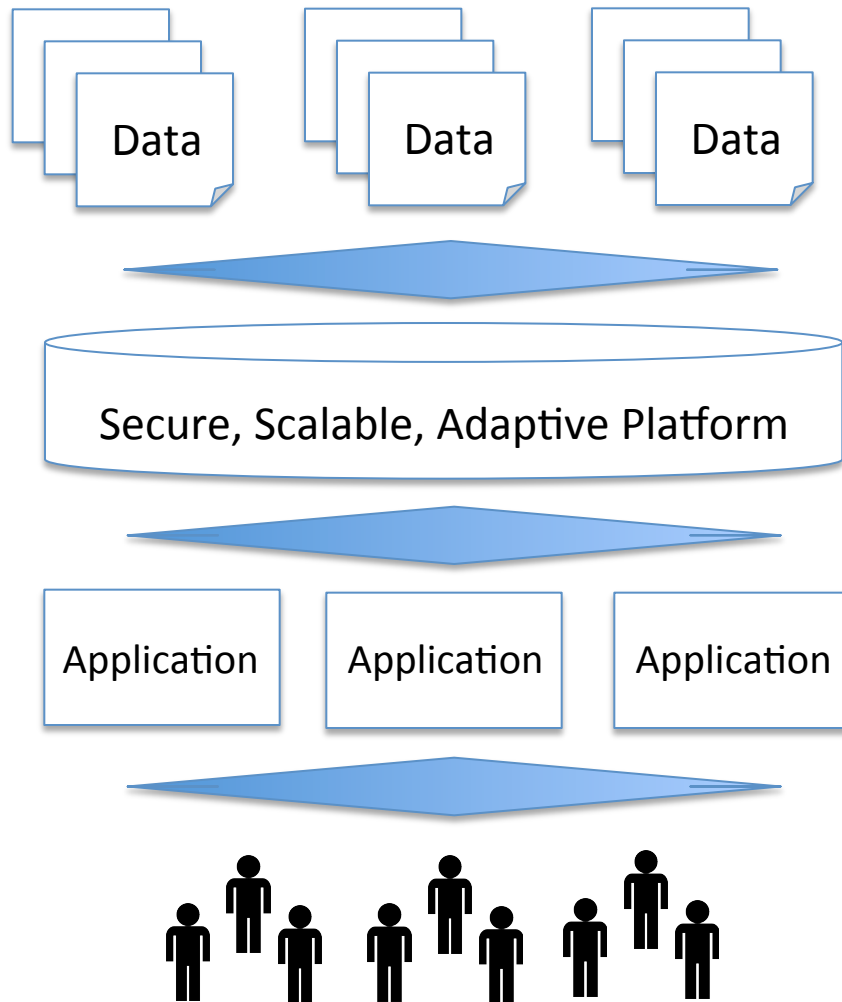
How your shopping habits reveal even the most personal information. By Charles Duhigg



- TOS and Access not Equal
- "In Europe, we consider privacy a fundamental right. That doesn't mean it is exclusive of other rights, but economic rights are not superior to privacy."
- Immature tools and talent

# Solution

Secure. Scale. Adapt.



Integration across:

- Multiple business lines
- Multiple data sets
- Multiple applications
- Multiple security, privacy, legal, policy, regulatory, and compliance constraints

New:

- Data Centric Security
- Interactive Discovery Analytics
- Rapid Application Development
- Insights and Value



- Granular access control gives data managers a scalpel instead of a sword to share data as much as possible without compromising secrecy.

## Document Labels

*Doc*<sub>1</sub> : (Federation)  
*Doc*<sub>2</sub> : (Klingon|Vulcan)  
*Doc*<sub>3</sub> : (Federation&Human&Vulcan)  
*Doc*<sub>4</sub> : (Federation&(Human|Vulcan))

## User Authorization Sets

*CptKirk* : {Federation,Human}  
*MrSpock* : {Federation,Human,Vulcan}

## Semantics

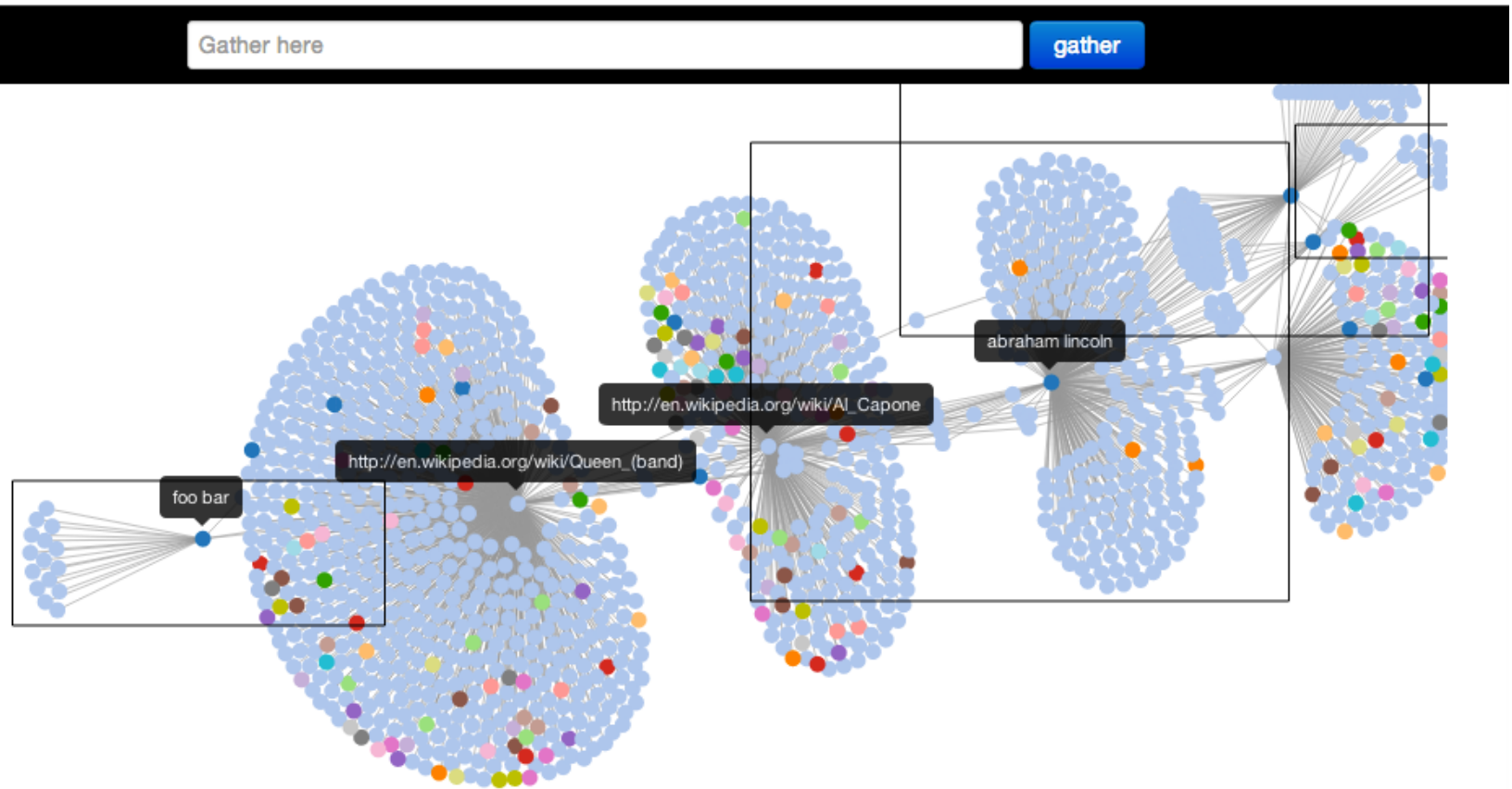
$$\frac{(T \Rightarrow \tau) \wedge (\tau \in A)}{(T, A) \models \text{true}} \text{ term}$$
$$\frac{(T \Rightarrow T_1 \ \& \ T_2) \wedge ((T_1, A) \models \text{true}) \wedge ((T_2, A) \models \text{true})}{(T, A) \models \text{true}} \text{ and}$$
$$\frac{(T \Rightarrow T_1 \ | \ T_2) \wedge (((T_1, A) \models \text{true}) \vee ((T_2, A) \models \text{true}))}{(T, A) \models \text{true}} \text{ or}$$
$$\frac{(T \Rightarrow (T_1)) \wedge (T_1 \models \text{true})}{(T, A) \models \text{true}} \text{ paren}$$

## Syntax

*WORD*  $\Rightarrow$  [a-zA-Z0-9\_]+  
*CLAUSE*  $\Rightarrow$  *AND*  
 $\Rightarrow$  *OR*  
*AND*  $\Rightarrow$  *AND & AND*  
 $\Rightarrow$  (*CLAUSE*)  
 $\Rightarrow$  *WORD*  
*OR*  $\Rightarrow$  *OR | OR*  
 $\Rightarrow$  (*CLAUSE*)  
 $\Rightarrow$  *WORD*

# Search + Graph Analysis

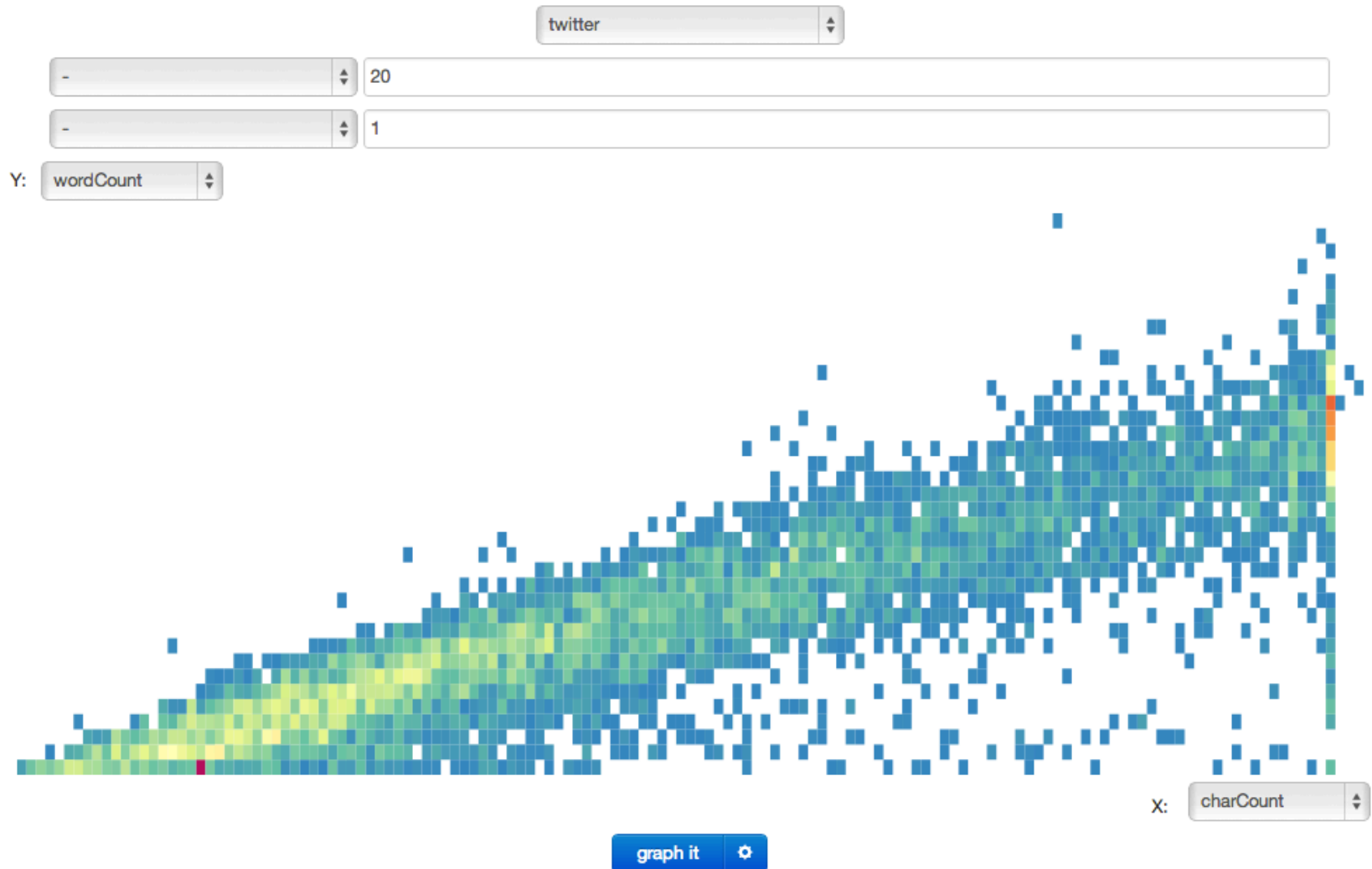
Secure. Scale. Adapt.



...because sometimes you feel like a nut

# Schema-less Statistics

Secure. Scale. Adapt.



- Secure computations in distributed programming frameworks
- Security best practices for non-relational data stores
- Secure data storage and transactions logs
- End-point input validation/filtering
- Real-time security/compliance monitoring
- Scalable and composable privacy-preserving data mining and analytics
- Cryptographically enforced access control and secure communication
- Granular access control
- Granular audits
- Data provenance

<https://cloudsecurityalliance.org/csa-news/csa-big-data-releases-top-10-security-privacy-challenges/>

- Big Data not for Individual Economic Gain, but rather for the public good:
  - Disease, Crime, World Events, etc.

# Contact

Secure. Scale. Adapt.



**sqrrl data, Inc.**  
275 Third St.  
Cambridge, MA 02142

617-520-4375  
[www.sqrrl.com](http://www.sqrrl.com)  
[@sqrrl\\_inc](https://twitter.com/sqrrl_inc)  
[info@sqrrl.com](mailto:info@sqrrl.com)