



Big Data and Insurance

Stephen Mildenhall

July 22-24, 2014

Agenda

- Section 1** Data, Big and Otherwise
- Section 2** Transforming Data into Information
- Section 3** Transforming Information into Insight
- Section 4** Communicating Insight for Action
- Section 5** The Dark Side
- Section 6** Moving Forward

Our Journey...

Data

Information

Insight

Action



Section 1: Data, Big and Otherwise

Data: the raw ingredient of decision making

[Big] Data

Three Vs
Volume
Variety
Velocity
Value | Veracity

**BIG = hard
to manage
on existing
technology**

Big data: how BIG is big?

Bytes	Example
Mega	10^6 Spreadsheet
Giga	10^9 CD-ROM, flash drive
Tera	10^{12} Upper bound of small
Peta	10^{15} Lower bound of Big
Exa	10^{18} BIG
Zetta	10^{21} Definitely BIG
Yotta	10^{24} Huge



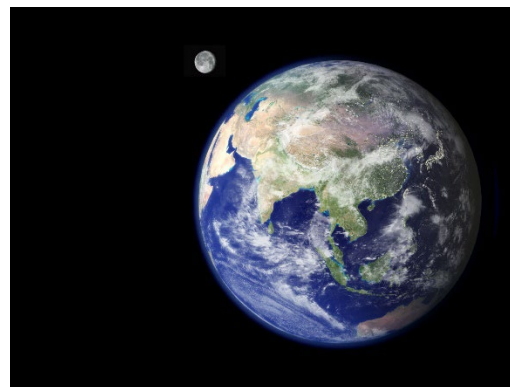
Library of Congress
All printed materials



300PB



10 grams DNA 4ZB



All stored data 500EB-3ZB
+2.5EB/day



Google's data 10 EB



Driverless car data torrent...

Driverless cars
produce 750MB
data per minute

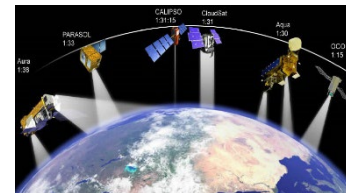
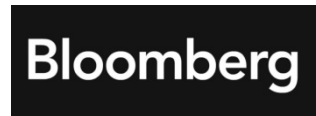
3.6ZB/yr
for US fleet

3.2T miles / yr @ 40 mph x 750 MB / min; 250M vehicles

Data: the raw ingredient of decision making

data

Big Data

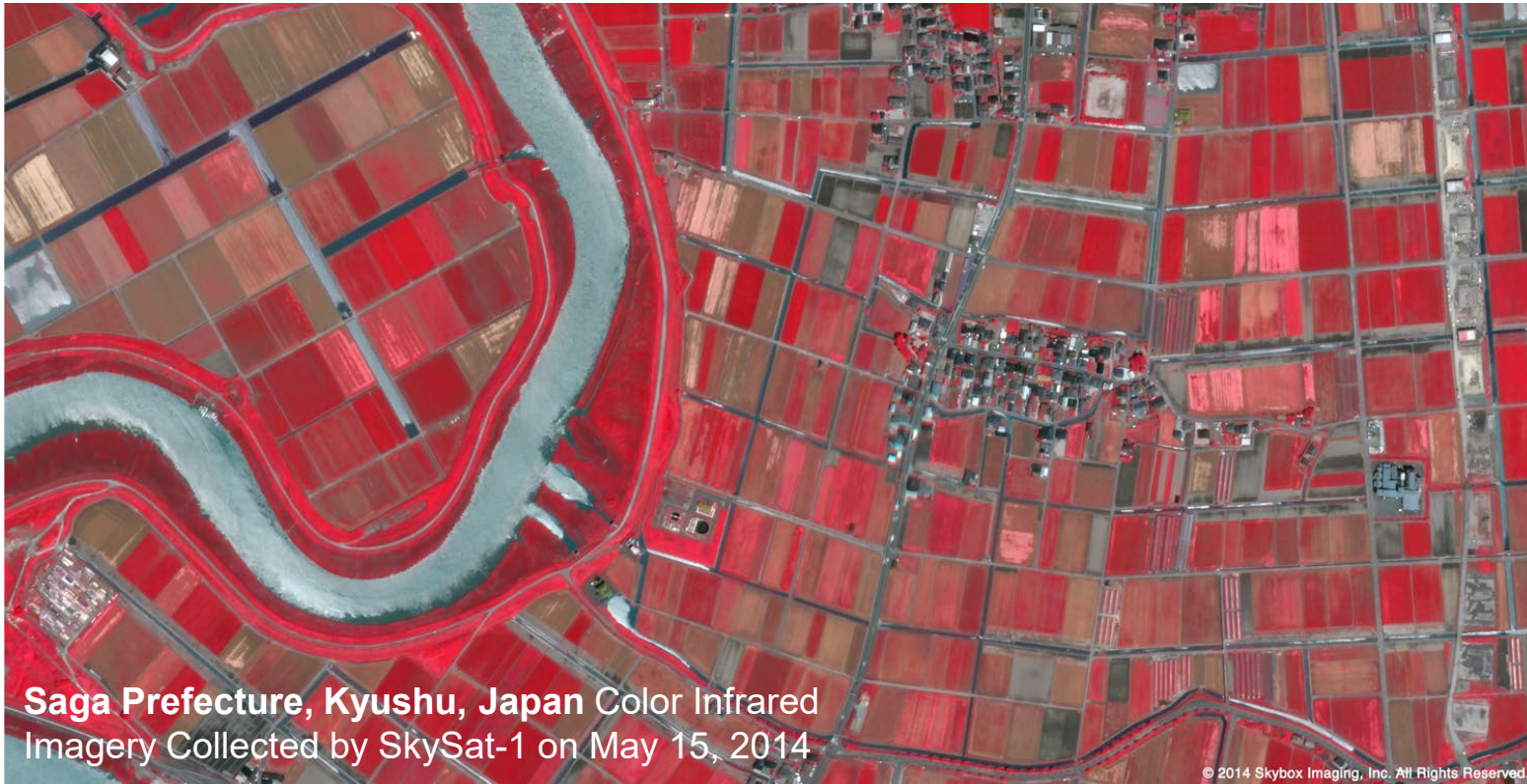


Comparison of insurance and “big” data models

Insurance Data	Internet Big Data
Sparse, expensive data	Massive, free data
Policy, claim, client	Internet of Things
Expensive to collect & maintain	Free: user contributed, digital exhaust
Static	Streaming
Structured	Unstructured - variety
Inhomogeneous within type	Homogeneous within type
Potentially high value density	Low value density
Low to medium volume	Very high volume

Successes and potential	Memes, curiosities and some value
Credit	Google Flu, translate
Progressive and GEICO growth	Amazon “Your recommendations”
Cat modeling	Click on me!!

If you don't have the data...make it



© 2014 Skybox Imaging, Inc. All Rights Reserved.

om SportsVU

AON
Empower Results®

Daily update frequency allows pre- and post-event damage assessment

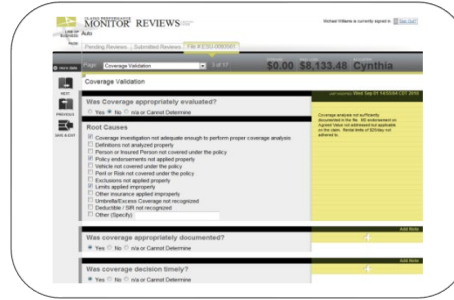


New data to better manage claims

Aon Benfield Inpoint

Accountability

Objective, outcome-focused measures and “leakage” drive behaviors

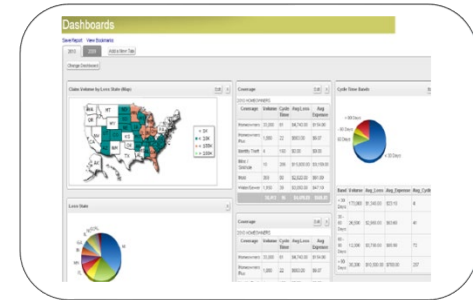


Assessment

Standardized assessment forms with claims best-practices objectively assess and measure performance

Line → Auto

Office	CPI (0-300)	Leakage %
New York	279	6%
Dallas	253	8%
San Francisco	227	12%
Chicago	217	14%
Tampa	198	15%



Actions

Actionable, root-cause level insight into areas impacting claims outcomes down to individual adjuster

Phase → Investigation

Root Cause	Opportunity
Comparative or contributory fault not considered in determination of liability	12%
Failure to properly evaluate hold harmless agreement within contract	10%
Investigation failed to adequately address relevant issues of the claim	9%
Failure to determine date of loss	6%

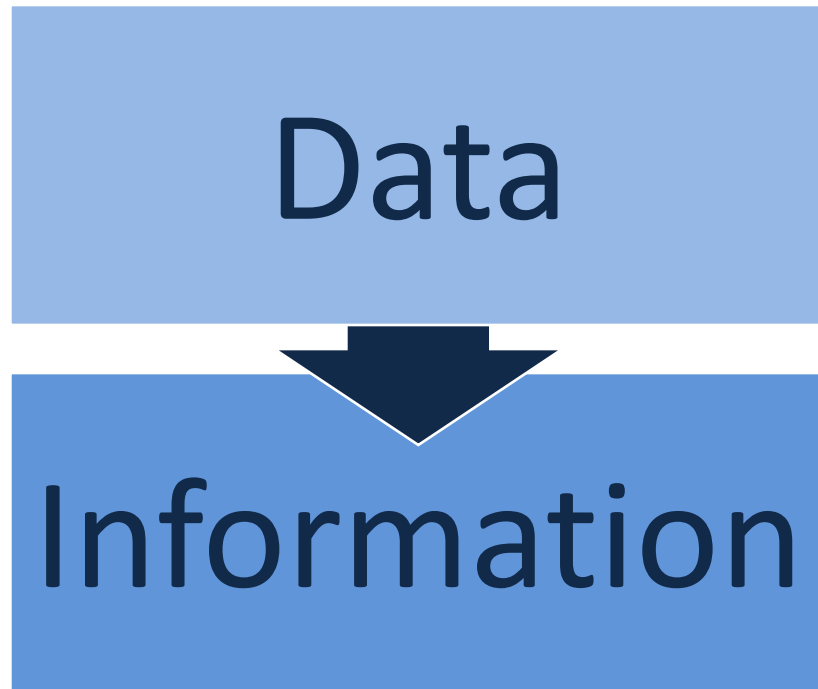
Analytics

Integrate assessment information with claims data to provide insight into and control over claims performance



Section 2: Transforming Data into Information

Work and some magic transforms data into information



**Transformation
= IT Problem**

Tools
~~Excel~~
SQL
Hadoop

Old tools and methods compared to new tools and methods

Old World	New World
Expensive, customized hardware	Cheap, commodity hardware
Model data	Dump data
Pre-optimize (index) to expected queries	Post-optimize requested ad hoc queries
Early binding types	Late binding types
Inflexible, changes slow	Flexible, built to change
Compute everything , dimensional DW	Compute what you want , on demand
Move data to the code	Move the code to the data



Magic = fundamentally different operating model, 4x speed-up

How new methods improve on old

Are two database records equal?

Record	Age	Sex	Ht	Wt	SSN	Zip	Etc.	Etc.
Memory	001010101010101001010101010100100010111...							

- Old School
 - $Age_1 = Age_2$
 - $Sex_1 = Sex_2$
 - $Wt_1 = Wt_2$
 - $SSN_1 = SSN_2$
 - Etc.
- Implement for each record type
- Slow to execute
- New School
 - Compare strings as binary numbers
- Works for ALL record types
- Quick to execute: primitive CPU function

Path to quality data remains the same...work, no quick fix



“Inhomogeneous within type”

Re-code
Re-map
Merge
Augment

- Definition of a client, product, line, geography
- Cross BU accounts
- Accounting conventions
- Pre-fix / post-fix errors
- Politics: why am I sending you this data?

Aon GRIP: global repository of insurance data

- \$20B+ premium per year
- \$90 trillion limit
- 1,330 carriers
- 66,084 clients
- 52 lines of business
- 43 source systems
- 37 countries



Managing employee benefits data

Aon Hewitt

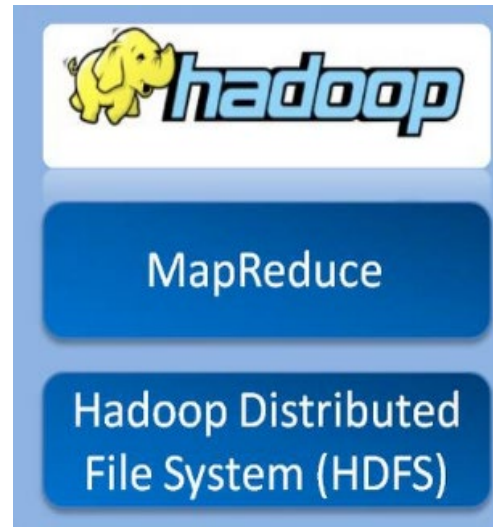
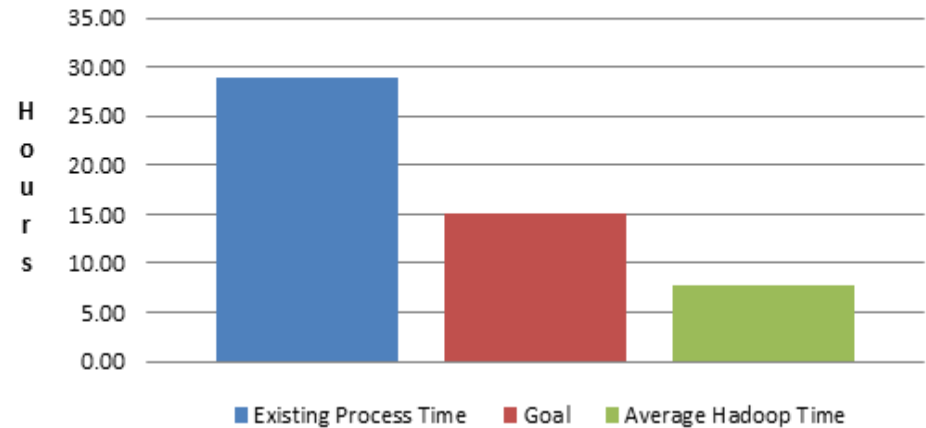
Traditional Model

250 clients' data to refactor
Two weekends per client

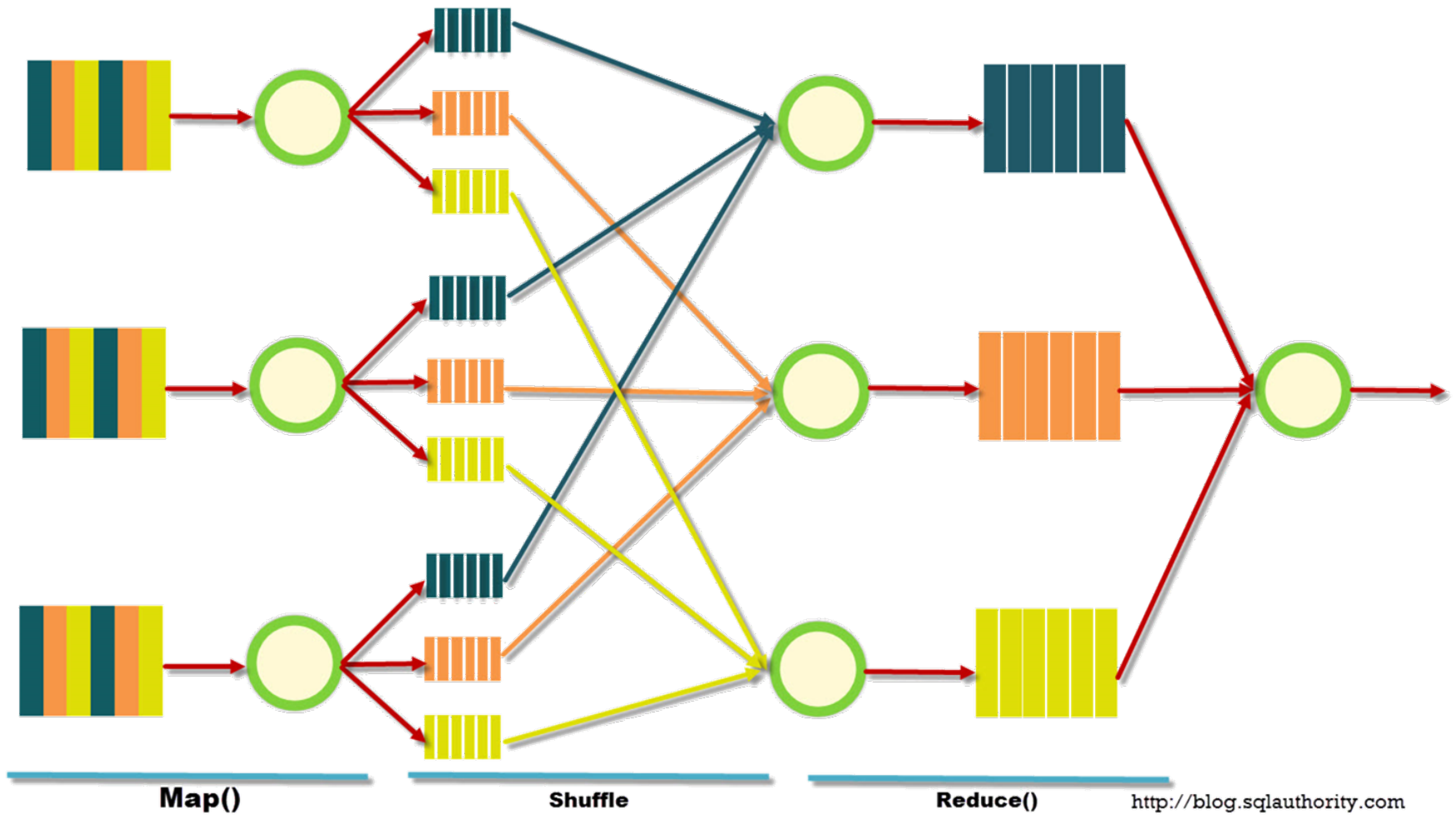
3-4x speed
0.5x cost

- Hadoop conceived as intermediate step to data warehouse
- Power and flexibility now allow it to *become* the data warehouse
 - Fast load; no pre-set data model
 - Flexible post-load querying

Loading TBA data into Benefits DW
Existing Process v. Goal v. Hadoop

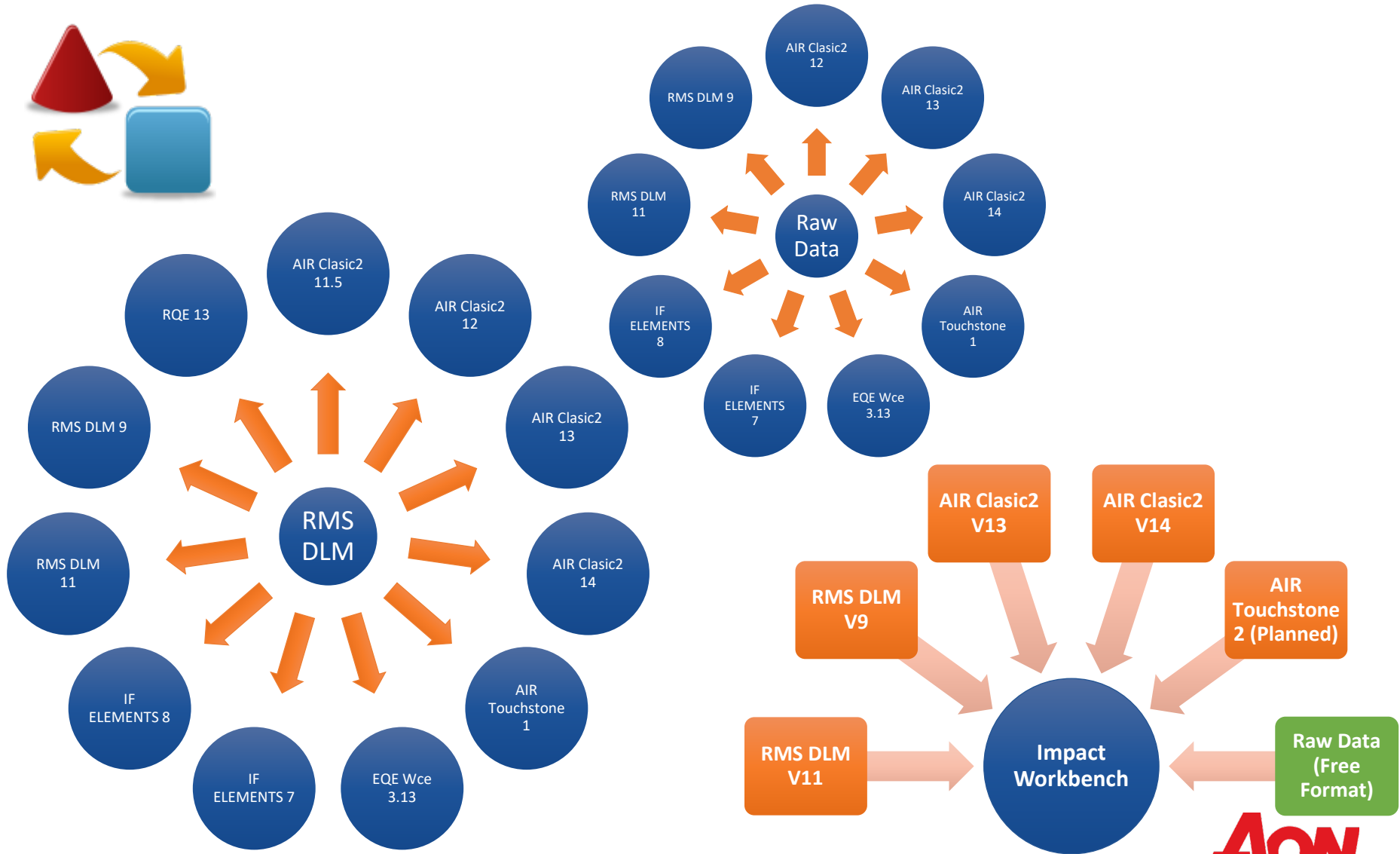


What is MapReduce?



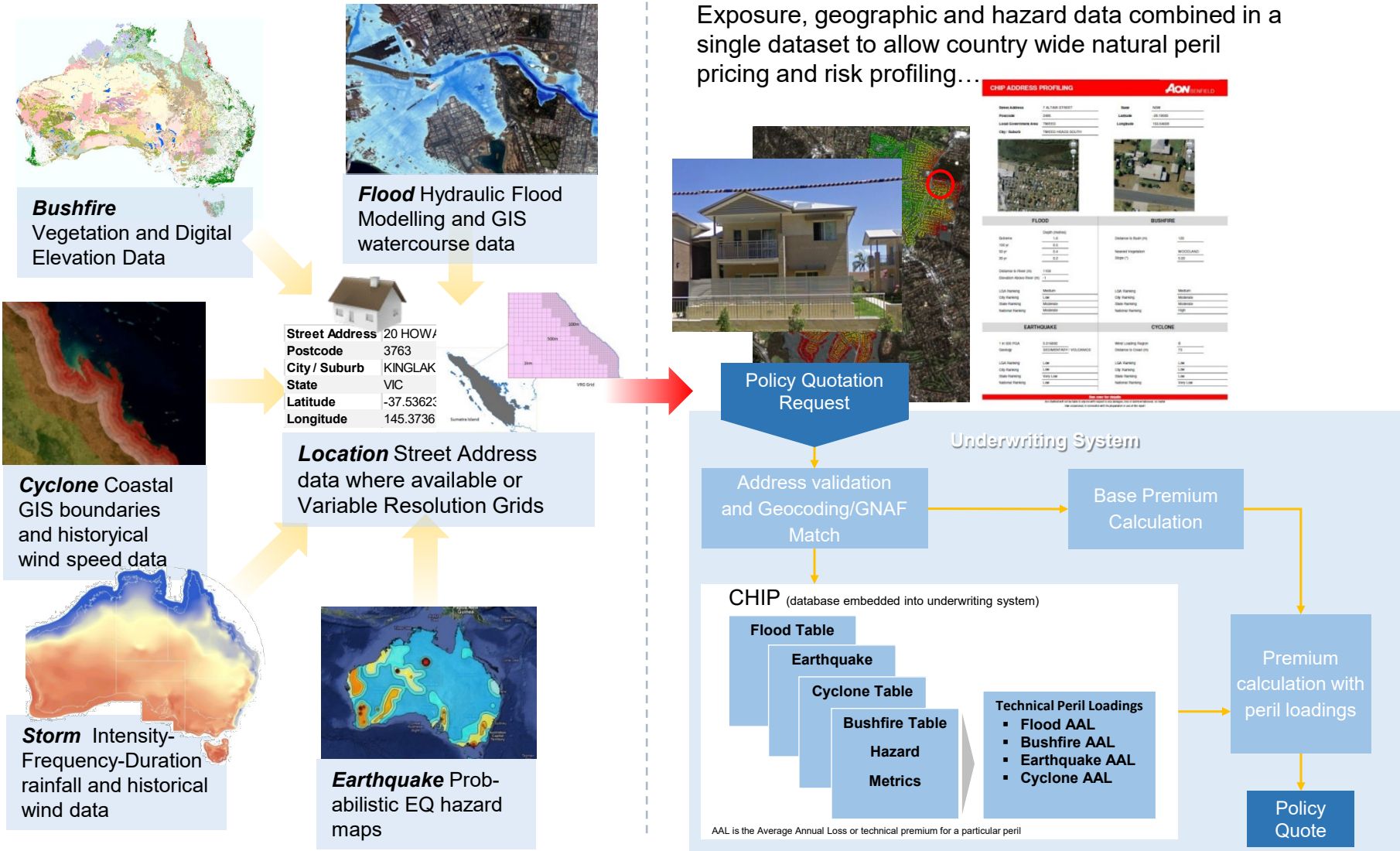
Catastrophe risk data transformation and management

Aon Benfield ImpactWorkbench



Multi-Source Catastrophe Hazard Information

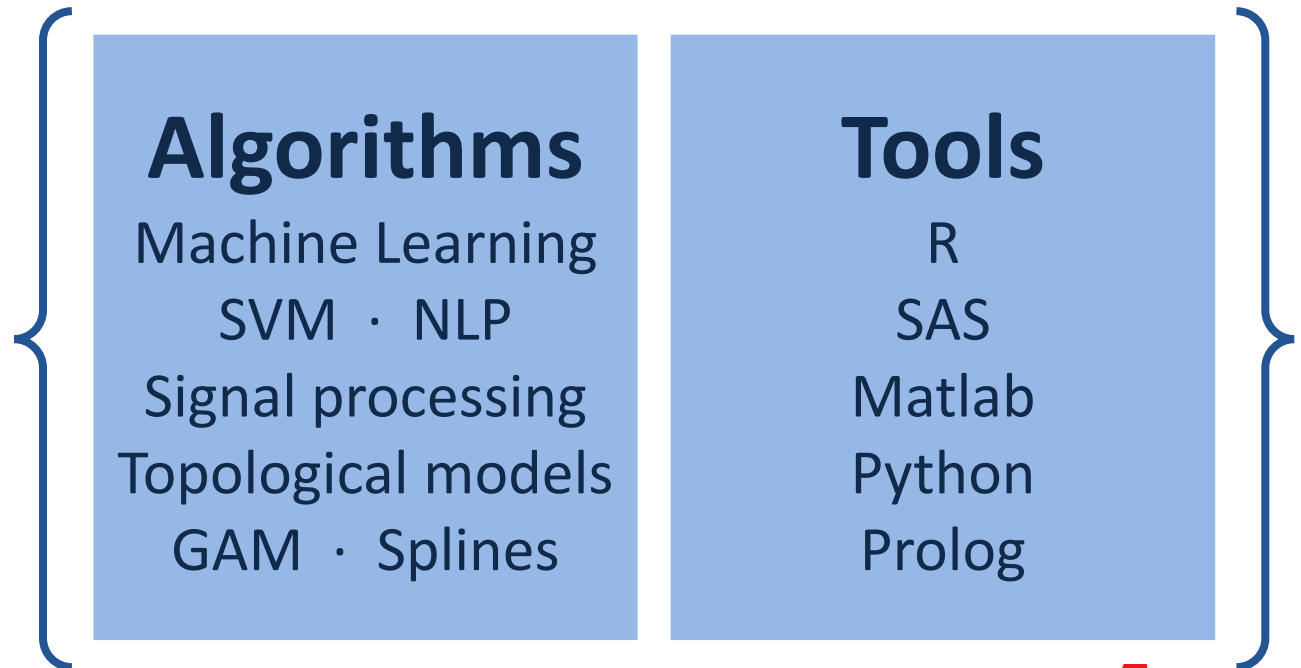
Aon Benfield Analytics Australia, "CHIP"



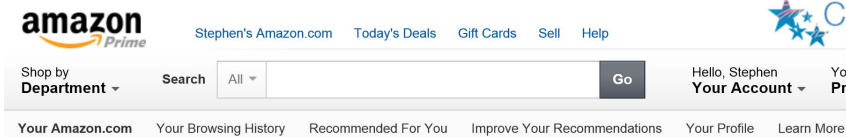
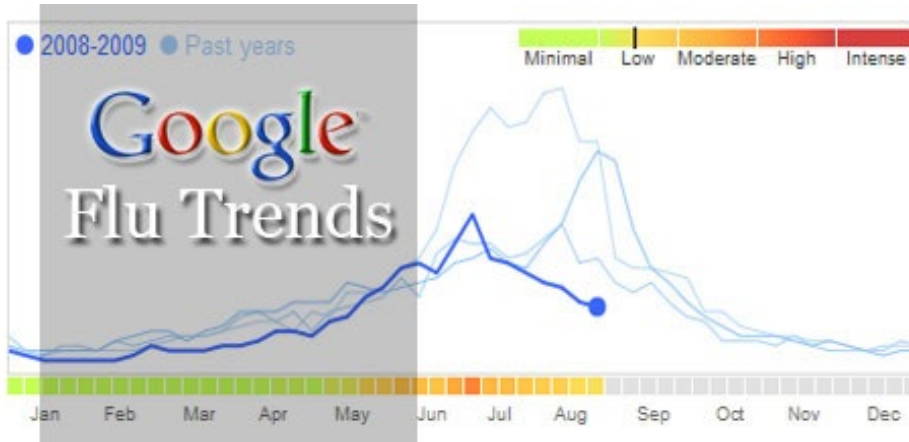


Section 3: Transforming Information into Insight

Analysis transforms information into insight



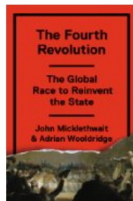
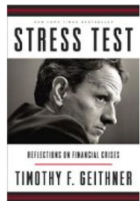
Big Data analysis memes



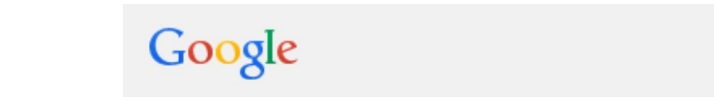
Your Amazon.com

Books

Page 1 of 34



WAL★MART®

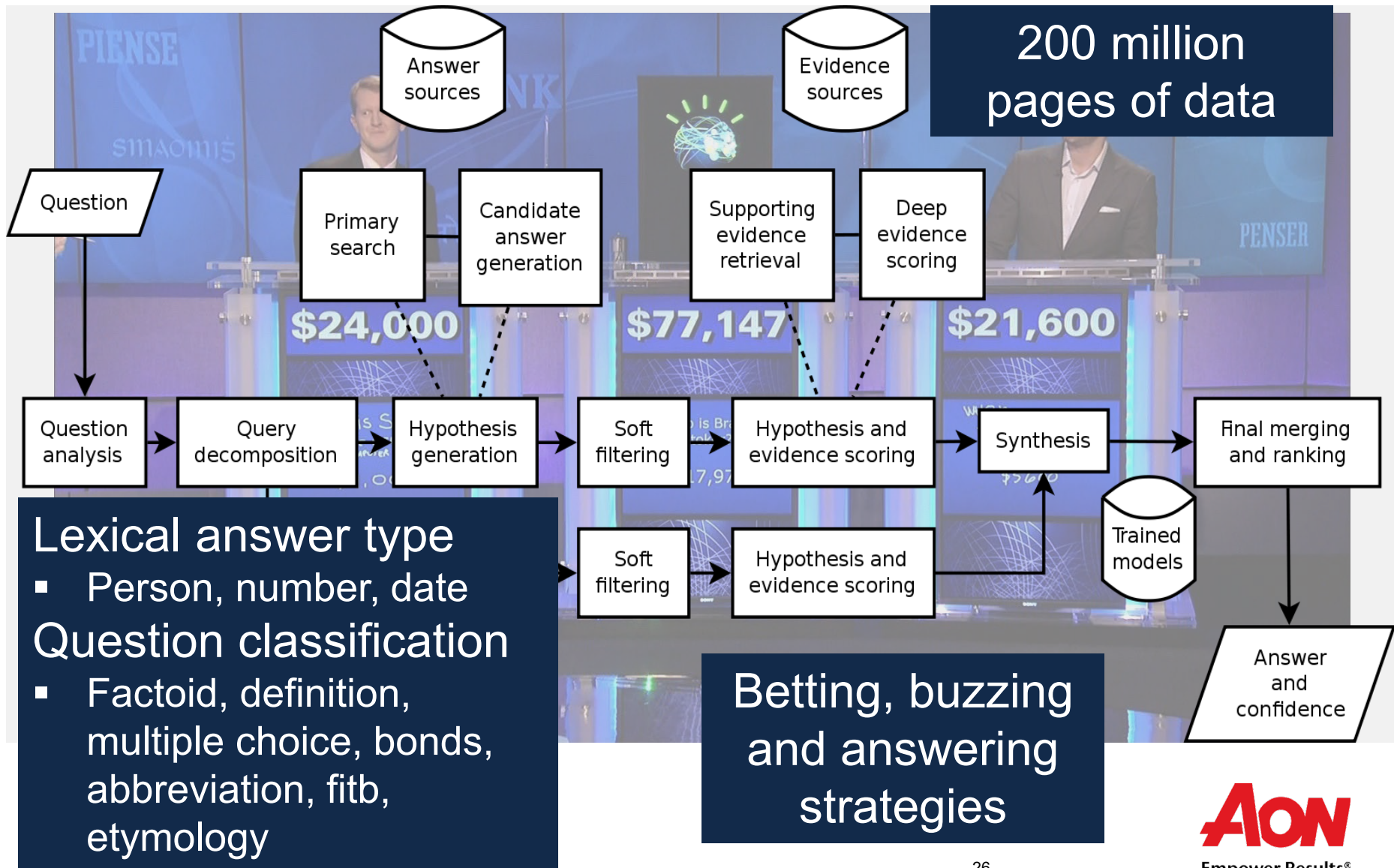


Translate

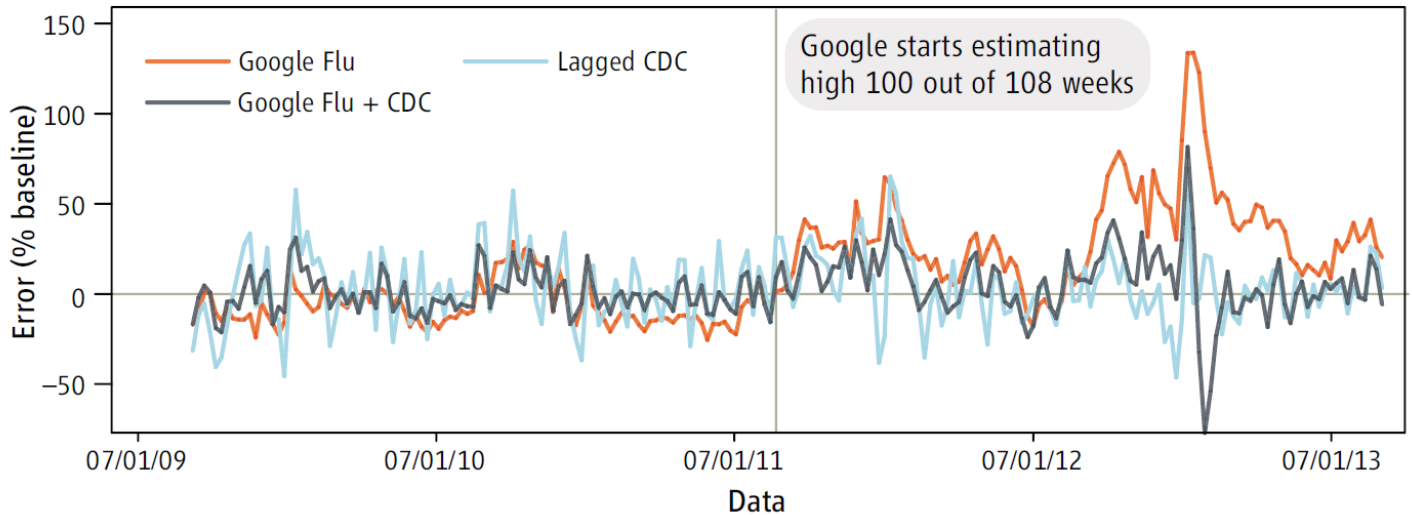
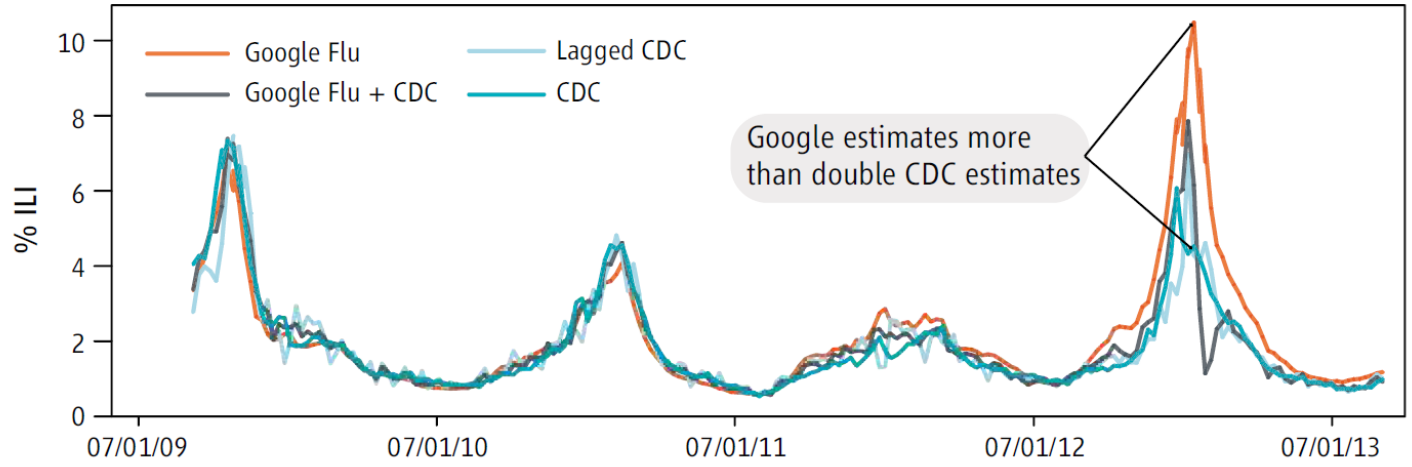
English Spanish French Detect language



Watson Wins Jeopardy! Training model a “significant effort”



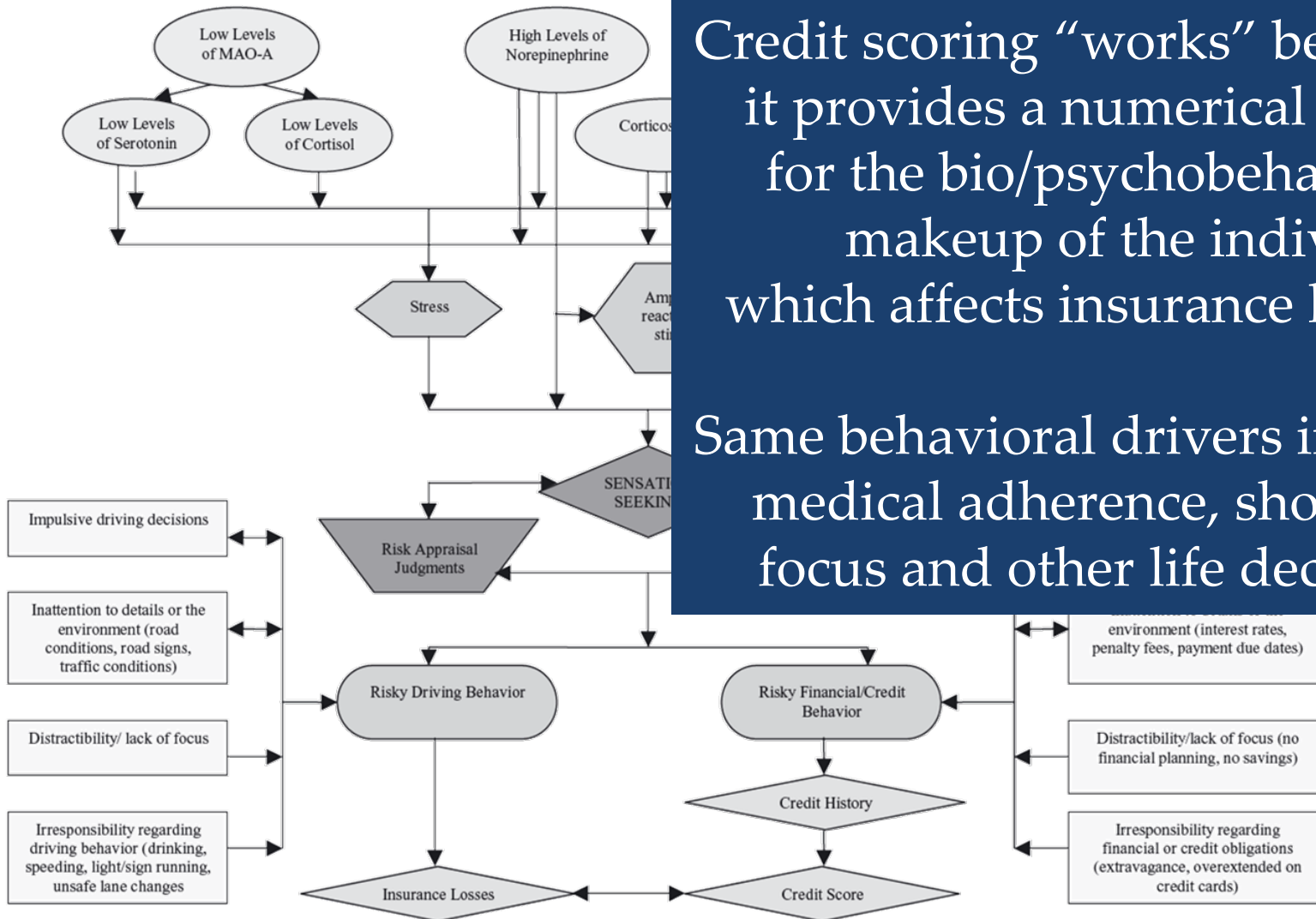
The Google Flu Trends story...the need for a theory



Need for a theory: the credit scoring model

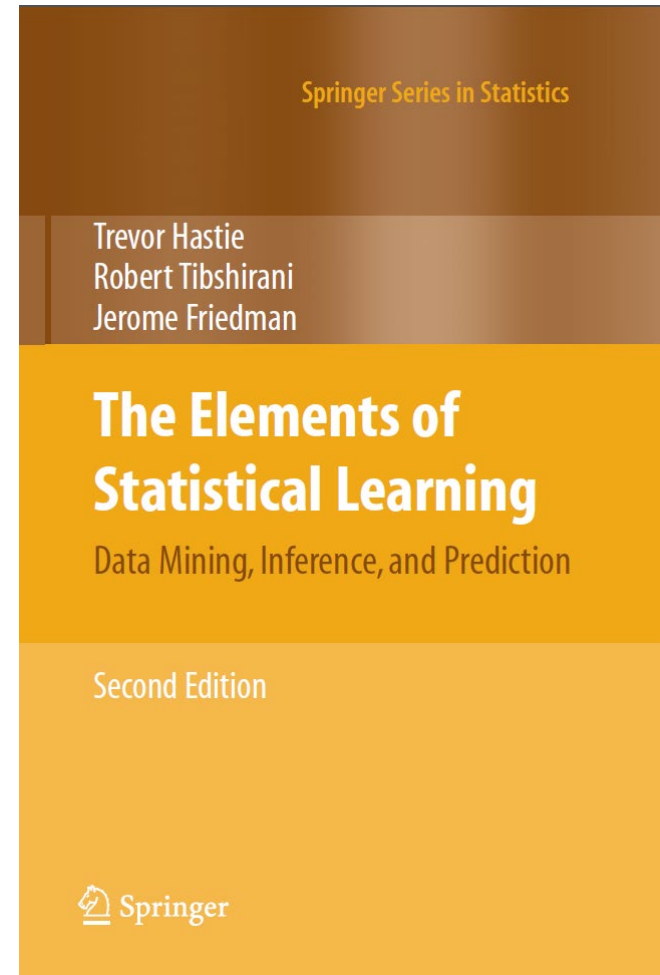
Credit scoring “works” because it provides a numerical proxy for the bio/psychobehavioral makeup of the individual which affects insurance losses.

Same behavioral drivers impact medical adherence, shopping focus and other life decisions



Techniques and algorithms

- Supervised vs. unsupervised
- Clustering
- Classification
- Machine learning
- Regression & linear models
- Support vector machines
- Naïve Bayes
- False discovery rate
- Topological data analysis
- Expansions & regularization, splines, kernels
- GAM
- Classification trees
- Boosting



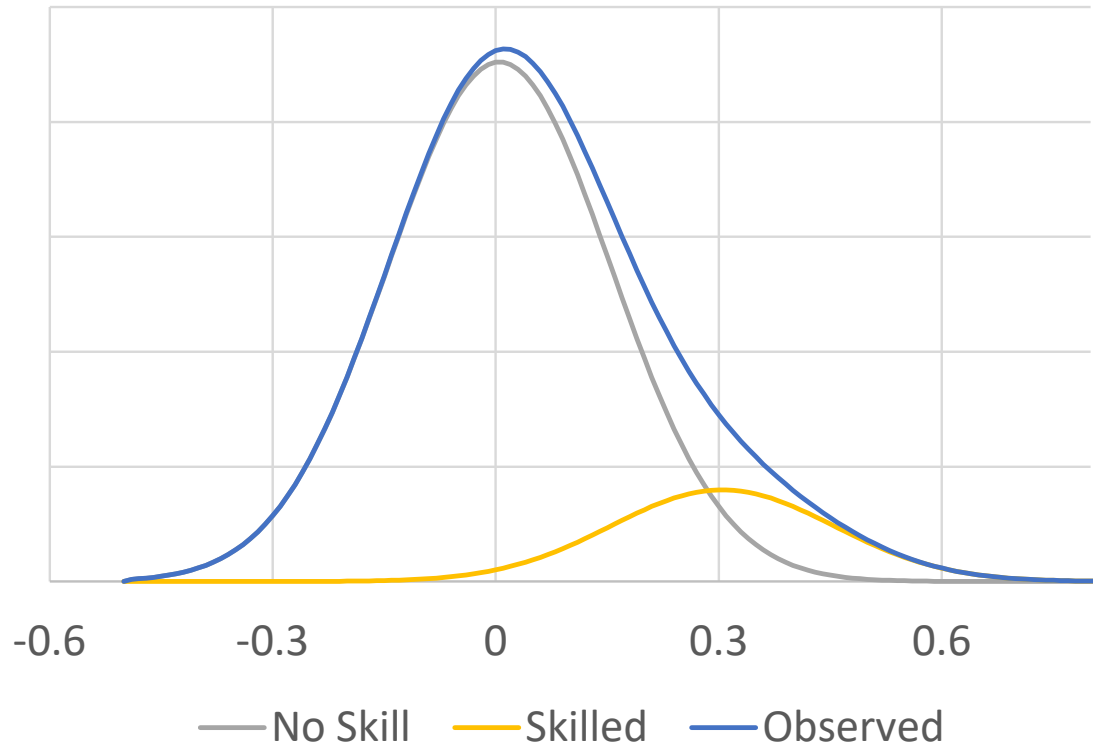
http://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII_print4.pdf

False discovery rate and investment performance

Aon Center for Innovation and Analytics / Hewitt EnnisKnupp

In 10,000 samples
100 observations
are significant at
1% level

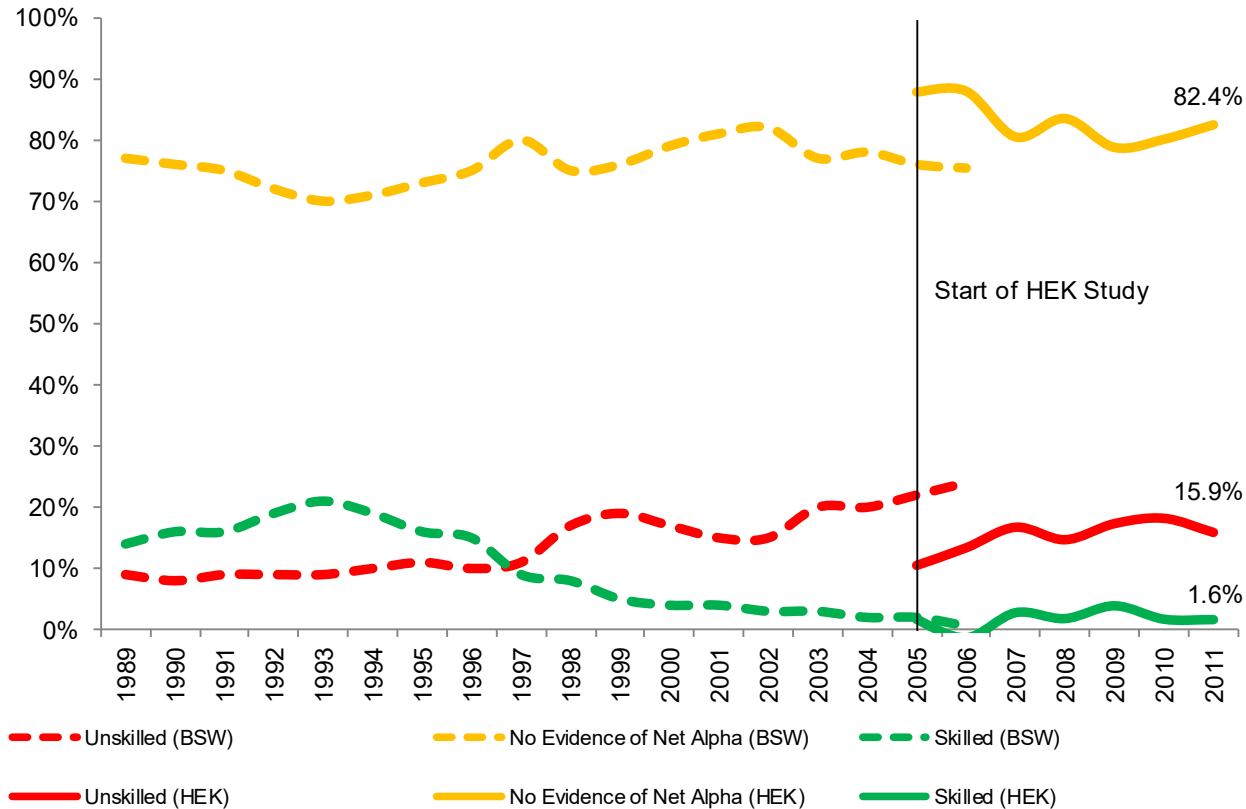
Many are
False
Positives



- With enough data you can use the empirical distribution, or parts of the empirical, to determine the null hypothesis distribution

False discovery rate and investment performance

Aon Center for Innovation and Analytics / Hewitt EnnisKnupp



Manager skill has steadily declined since the 1990s, and we estimate that only about 2% of products demonstrate evidence of true skill today. Success with active management requires a high bar

Text mining: topic classification, clustering, and sentiment

Aon Center for Innovation and Analytics / Aon Hewitt

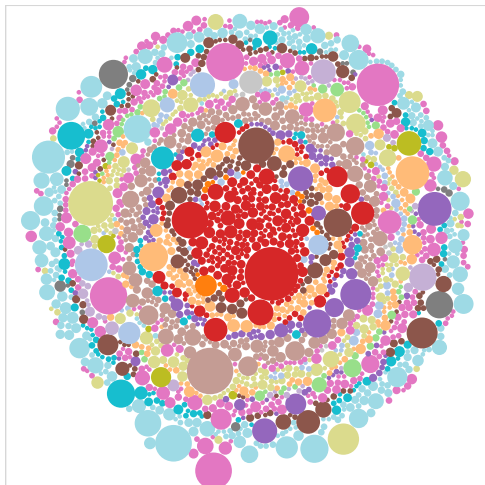
Top Layer Themes



top label

- aon hewitt
- benefit program
- bonus plan
- business unit
- career opportu..
- client teams
- continue to stri..
- cost of living
- current position
- current role
- day to day
- direct managers
- goal setting
- great job
- hard work
- laptop issues
- long paragraph
- management t..
- meet expectati..
- merit increases
- office environm..
- one group
- others
- performance re..
- personal life

Sub Layer Themes



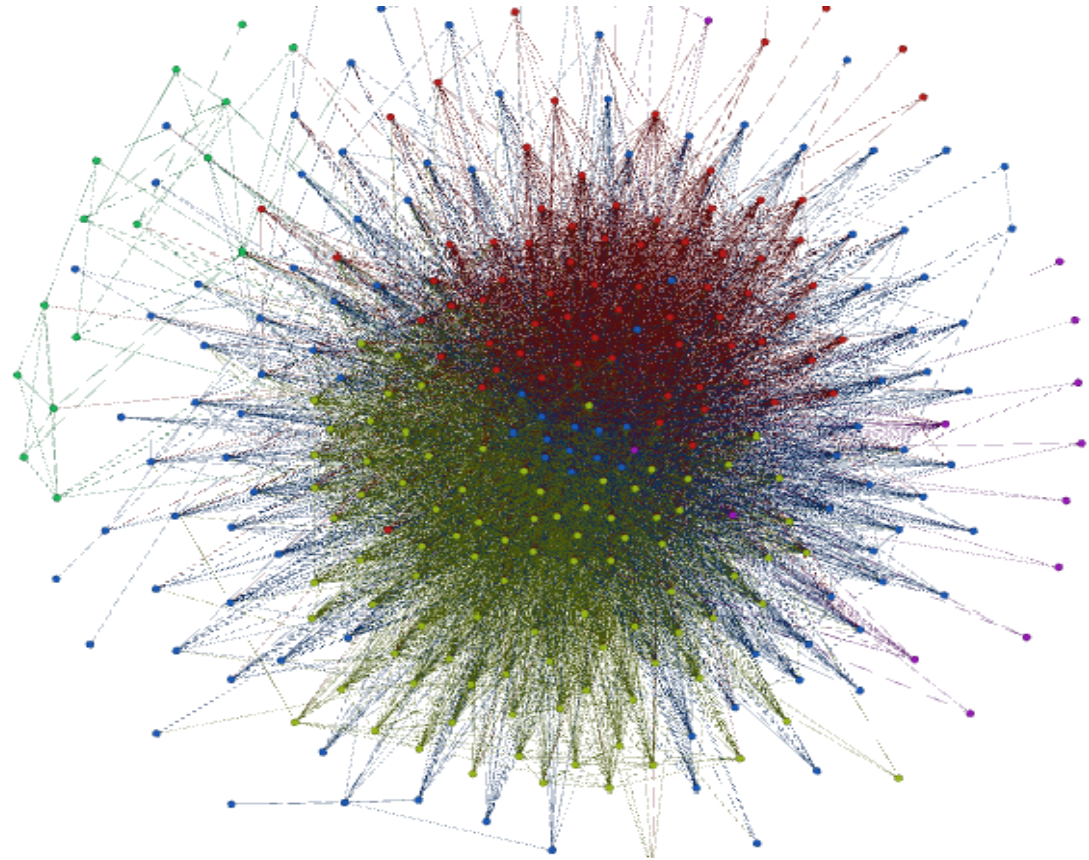
Message Details

mid	text
97116	Aon is a pay for performance environment and it in... res me to be my best a
97324	Senior Leadership is totally transparent on how the... ga...zation is doing holi... I am held to a different standard that the India asso... te... and unless Aon ho... the new business model of having them do all the... doing work is going to k... enough now. They are wonderful people to work w... meet expectati..
97664	Unless I go the standard route from BOA to DS to... M... BSM, I have no c... company. Multiple managers have said they have... dea how I can find opp... My Manager is the best Manager I ever had but ha... enny... be hard... addit...
97711	As a team manager I feel there is great flexibility to... jus... my work time if ne... time off and it is approved. The difficulty is actualy... ing able to enjoy that t... off for a personal day, to be wit... people doing t...

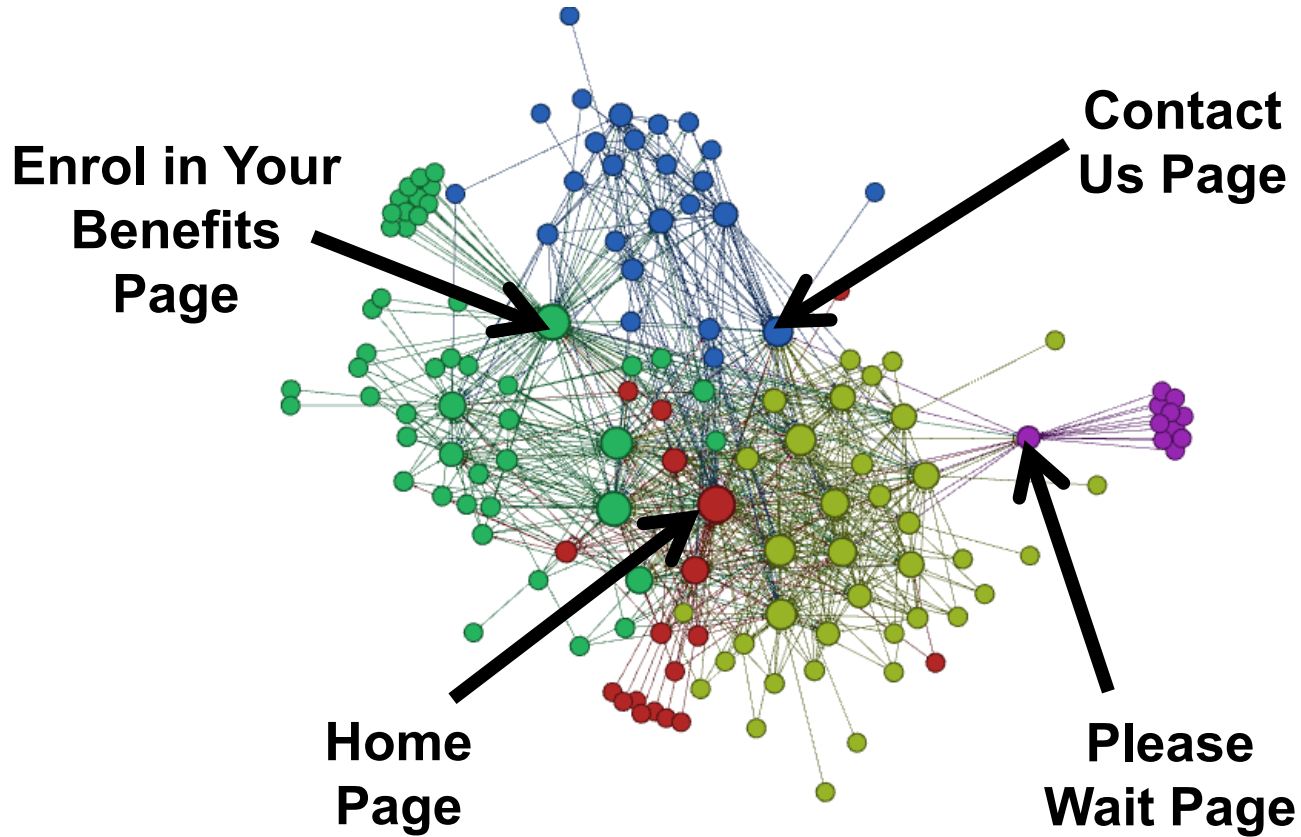
Customer experience and website usability

Aon Center for Innovation and Analytics / Aon Hewitt

- Six week annual enrolment period for large client
- 827,114 sessions
- 13,218,117 page views
- Traffic between pages



Which paths lead to the Contact Us page?

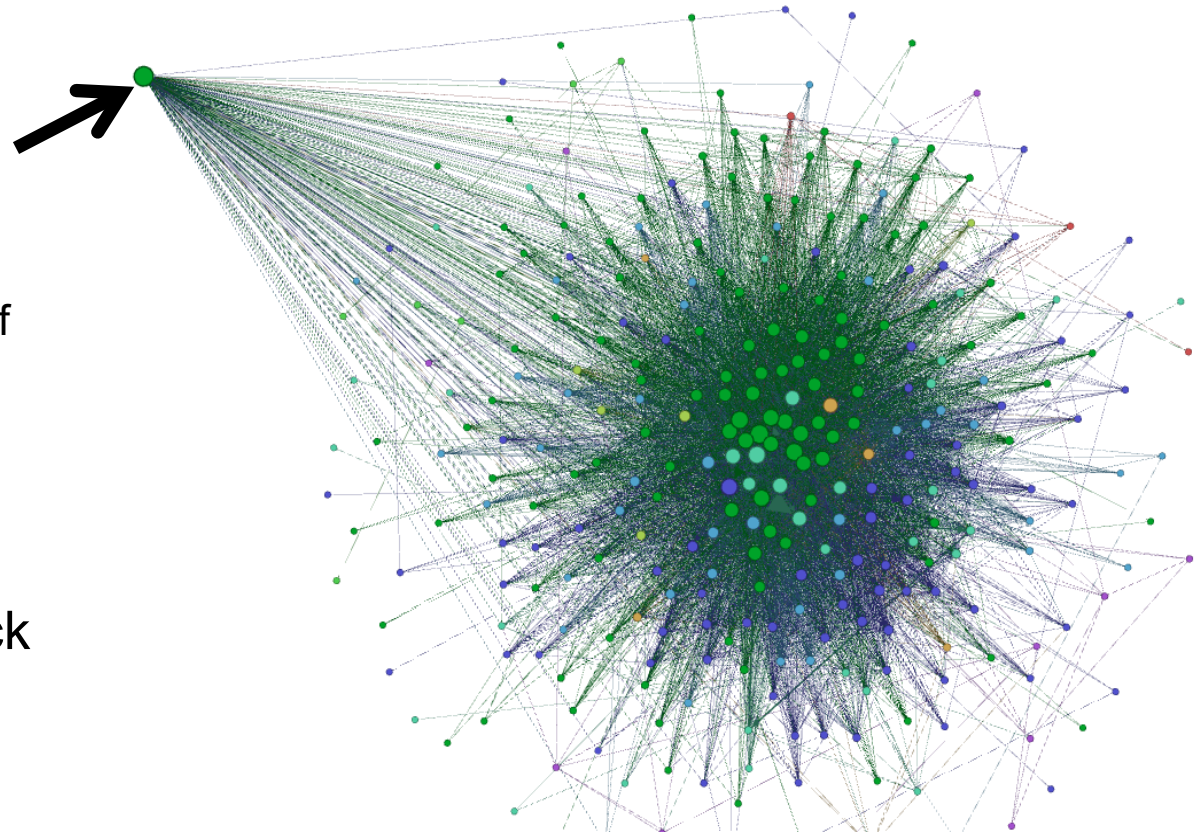


Smoking gun...

Back Button Page

moved from center of
cloud for emphasis

- 97,000 sessions involving the Back Button page
- 12%, or 11,600, resulted in a visit to the Contact Us Page



The “Back Button Cone”



Section 4: Communicating Insight for Action

Clear Communication of Insight Drives Action

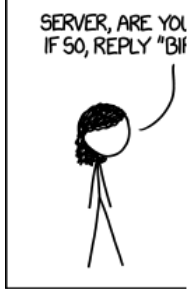
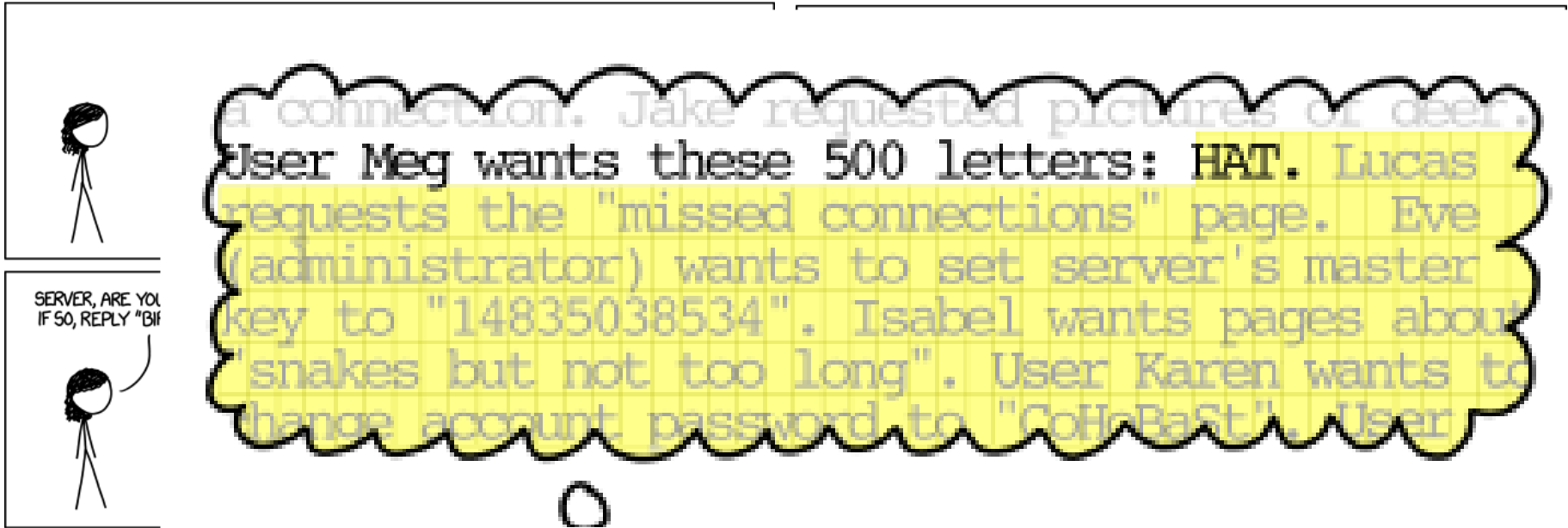
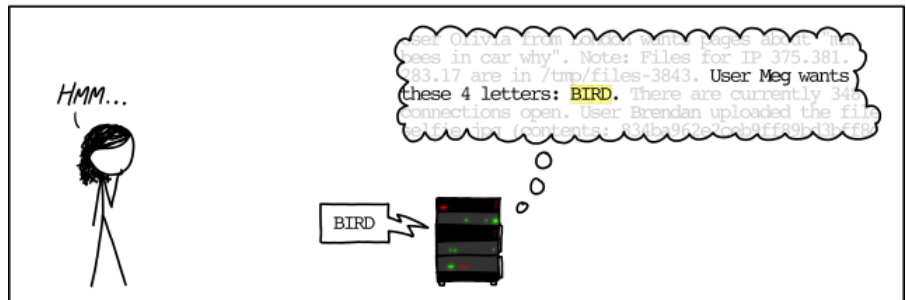
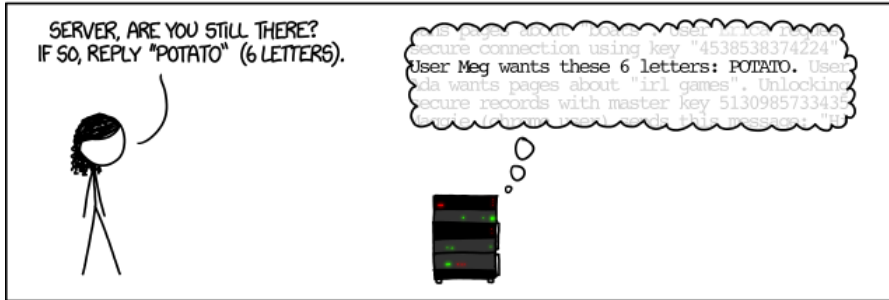




Section 5: The Dark Side

Heartbleed and data security

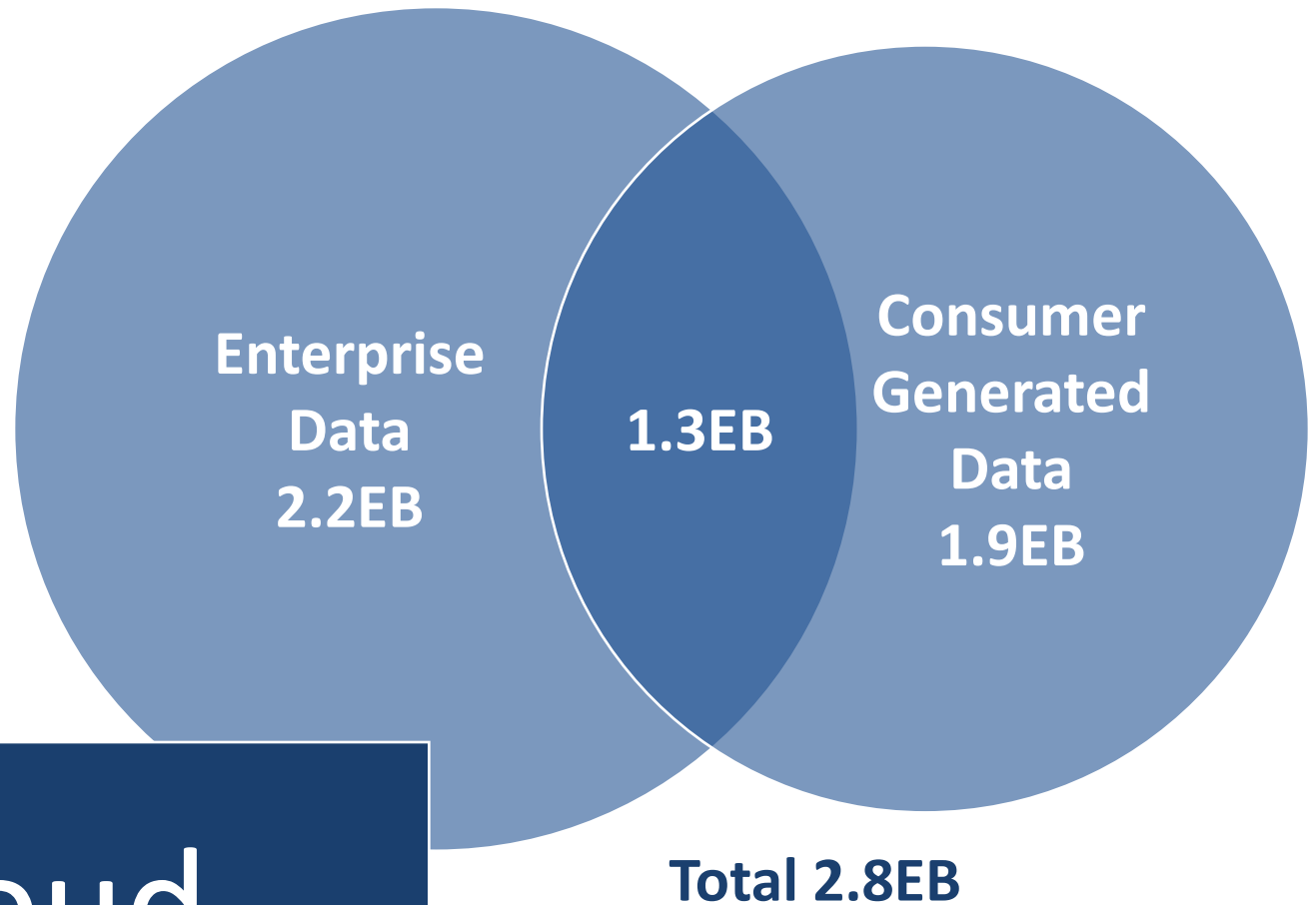
HOW THE HEARTBLEED BUG WORKS:



Privacy and security concerns

- Privacy
- Security
- Retention
- Processing
- Ownership
- Integrity
- Right to be forgotten
- Penalties based on propensities
- Uses of data that were not considered when data captured
- Opt-out creates a trace in an all-in world

Privacy and security concerns



Cloud



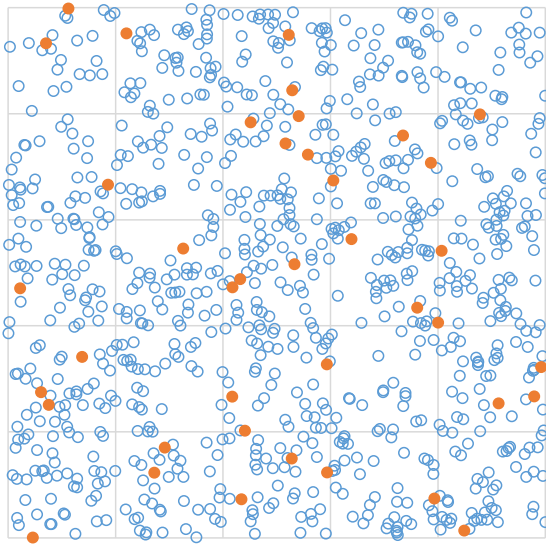
Section 6: Moving Forward

Q: Does any industry do it right?

A: Yes: insurance. Actuaries can say with great confidence what percent of people with your characteristics will live to be 80. But no actuary would ever try to predict when you are going to die. They know exactly where to draw the line.

Peter Fader “Is there money in Big Data?”
Frances and Pei-Yuan Chia Professor of Marketing
Wharton School of the University of Pennsylvania

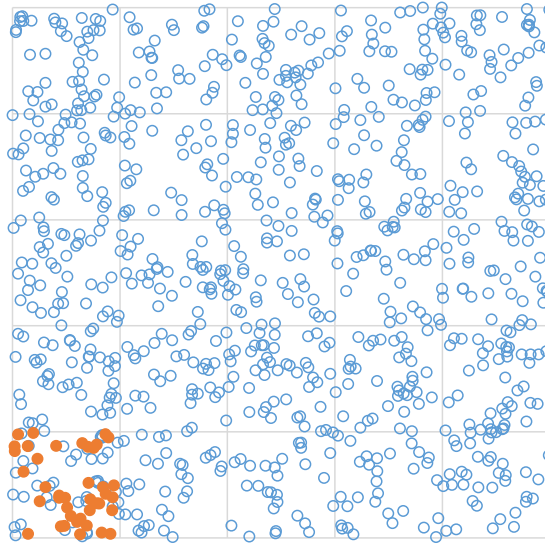
Big data and insurance: be careful what you wish for



Insurable

Old School

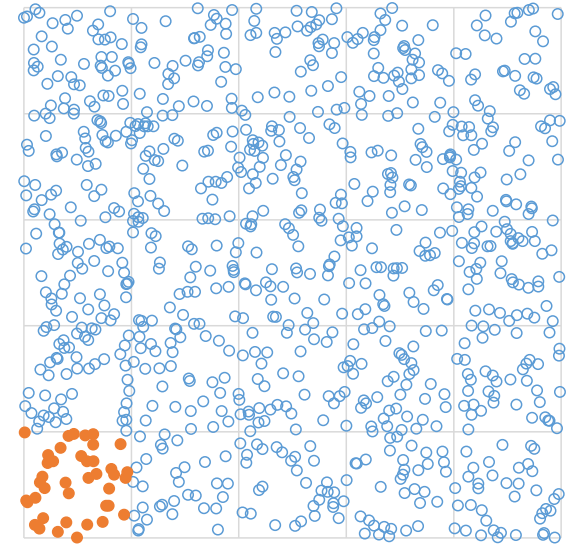
4% of 100%



Insurable, but expensive

Flood

50% of 8%



Not Insurable

Genetics

100% of 4%

If Big Data is so great why aren't projections better?



America's economy

Jobs are not enough

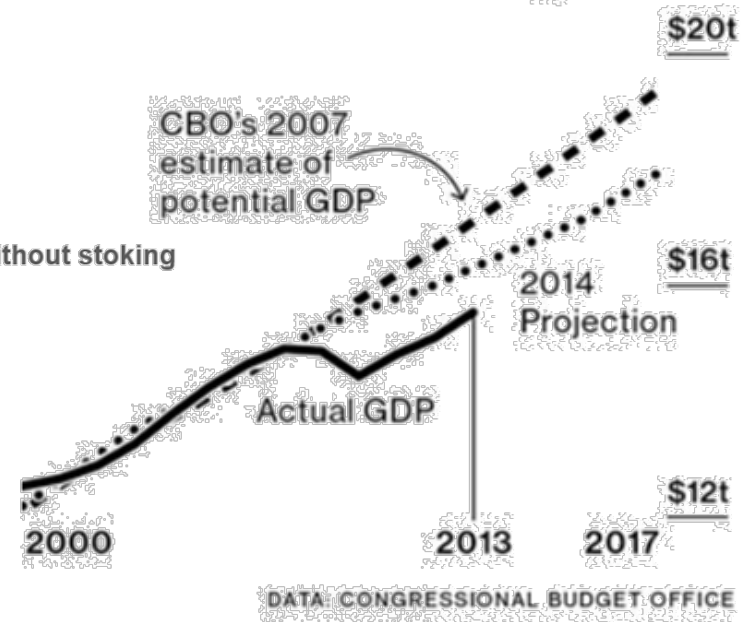
New figures show that the speed at which America's economy can grow without stoking inflation has fallen

'The best minds of my generation are thinking about how to make people click ads... That sucks.'

Jeff Hammerbacher

Founder and the Chief Scientist of Cloudera
One of Facebook's first 100 employees

Pessimism Reigns



Healthcare...

Rational expectations from Big Data

Schumpeter

Little things that mean a lot

Businesses should aim for lots of small wins from “big data”, that add up to something big

Stimulated by all the talk from consultants and sellers of data-crunching software about the transformative potential of big data, managers may have been misled into hoping it will give them massive, instant, Holy Grail solutions. But such discoveries are rare; and if they do exist, they have probably been made already. The reality is that **big data produces lots of small advances**—and that is good enough.

Lessons from our journey...

Capture ALL the data	Use the right tools	Learn yourself	Expect to do a lot of work
Identify sources: what is your digital exhaust?	Modern IT tools enable new analyses	Can't out-source institutional knowledge	IT/algorithm magic only goes so far
Omnichannel interaction data	New algorithms driving higher precision	Invest in people to learn over time	80%+ time spent in data collection
Normalize coding and fix at source	N=All: myth or reality?	Tactical partnerships	Data is competitive advantage of the future

- ABA ready to partner with clients to advance your capabilities through joint projects, workshops, peer review