

Taming the BigData Beast for Value & Insights

Usama Fayyad, Ph.D.

*Executive Chairman - **Oasis500***

*Chairman & CTO - **Blue Kangaroo***

Twitter: @usamaf

Sept 18th, 2013

Financial Times Live

Dubai - UAE



Usama Fayyad, Executive Chairman

usama@oasis500.com

Twitter: **@usamaf**



Outline

- Big Data all around us
- Introduction to Data Mining and Predictive Analytics Over BigData
- Some of the issues in BigData
- On-line data and facts
- Case studies on on-line marketing: Yahoo! Big Data
- Summary and conclusions

What Matters in the Age of Analytics?

1. Being Able to exploit all the data that is available
 - not just what you've got available
 - what you can acquire and use to enhance your actions
2. Proliferating analytics throughout the organization
 - make every part of your business smarter
3. Driving significant business value
 - embedding analytics into every area of your business can help you drive top line revenues and/or bottom line cost efficiencies

Why Big Data?

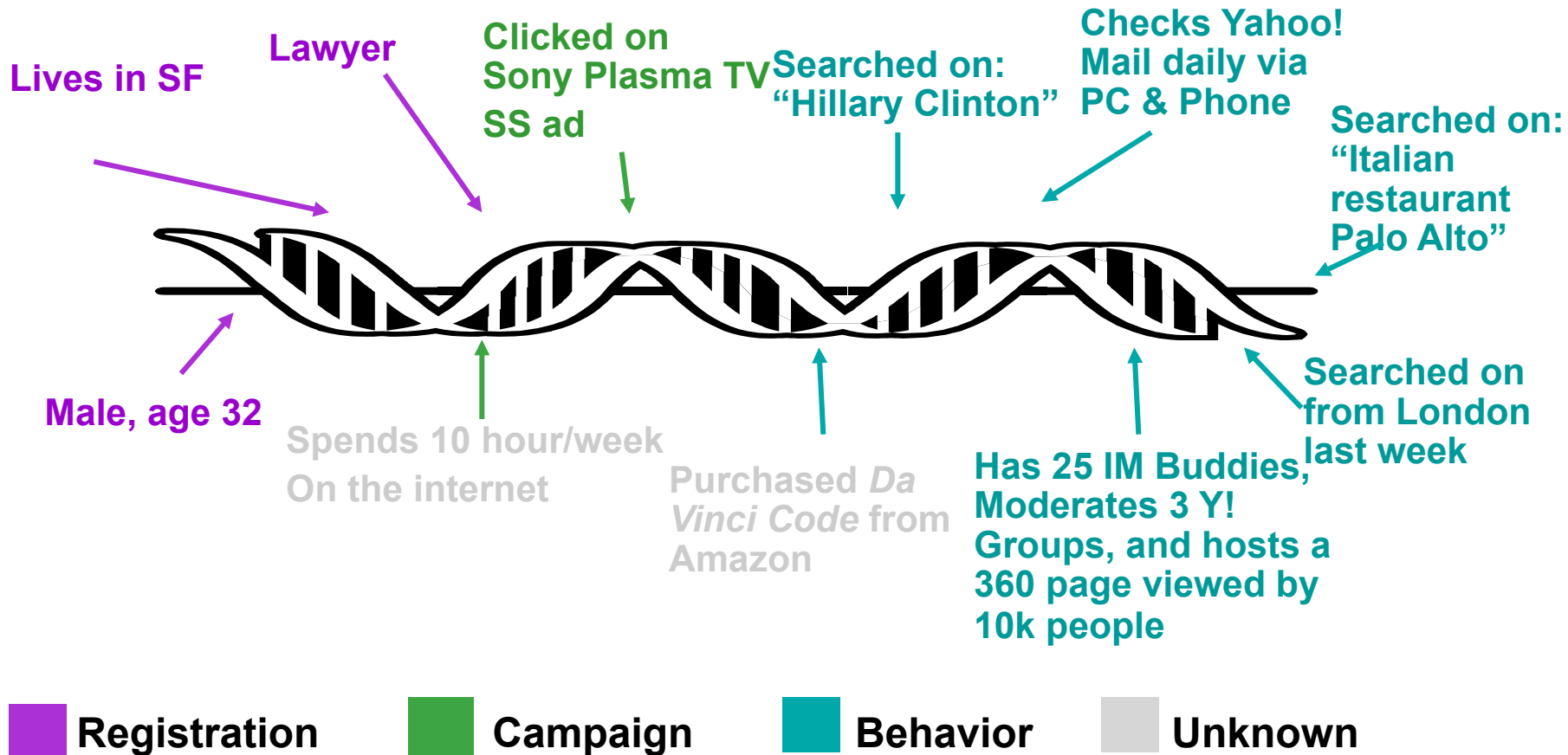
A new term, with associated “Data Scientist” positions:

- *Big Data: is a mix of structured, semi-structured, and unstructured data:*
 - Typically breaks barriers for traditional RDB storage
 - Typically breaks limits of indexing by “rows”
 - Typically requires intensive pre-processing before each query to extract “some structure” – usually using Map-Reduce type operations
- Above leads to “messy” situations with no standard recipes or architecture: hence the need for “data scientists”
 - conduct “Data Expeditions”
 - Discovery and learning on the spot

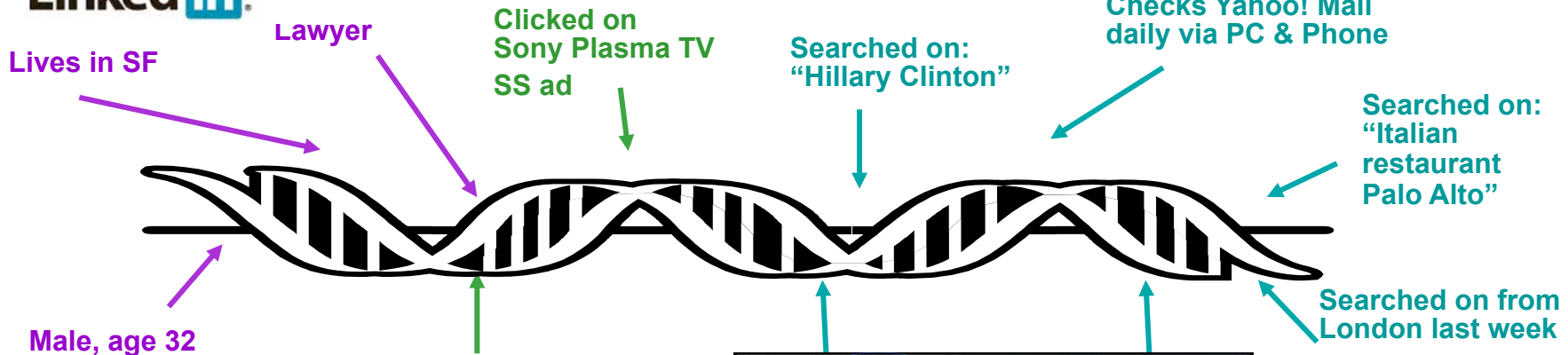
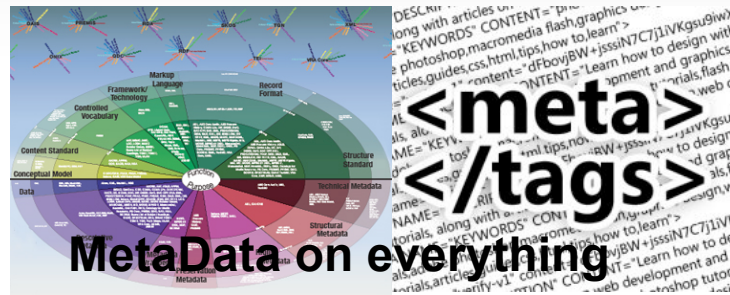
What Makes Data “Big Data”?

- **Big Data is Characterized by the 3-V's:**
 - **Volume:** *larger than “normal” – challenging to load/process*
 - Expensive to do ETL
 - Expensive to figure out how to index and retrieve
 - Multiple dimensions that are “key”
 - **Velocity:** *Rate of arrival poses real-time constraints on what are typically “batch ETL” operations*
 - If you fall behind catching up is extremely expensive (replicate very expensive systems)
 - Must keep up with rate and service queries on-the-fly
 - **Variety:** *Mix of data types and varying degrees of structure*
 - Non-standard schema
 - Lots of BLOB's and CLOB's
 - DB queries don't know what to do with semi-structured and unstructured data.

Today's Data: e.g. Yahoo! User DNA



How Data Explodes: really big



Male, age 32

Spends 10 hour/week internet

Likes & friends likes

Social Graph (FB)



ips, and
owed by

YouTube
flickr



The Distinction between “Data” and “Big Data” is fast disappearing

- Most real data sets nowadays come with a serious mix of semi-structured and unstructured components:
 - Images
 - Video
 - Text descriptions and news, blogs, etc...
 - User and customer commentary
 - Reactions on social media: e.g. Twitter is a mix of data anyway
- Using standard transforms, entity extraction, and new generation tools to transform unstructured raw data into semi-structured analyzable data

Text Data: The Big Driver

- We speak of “big data” and the “Variety” in 3-V’s
- **Reality:** biggest driver of growth of Big Data has been text data
 - Most work on analysis of “images” and “video” data has really been reduced to analysis of surrounding text

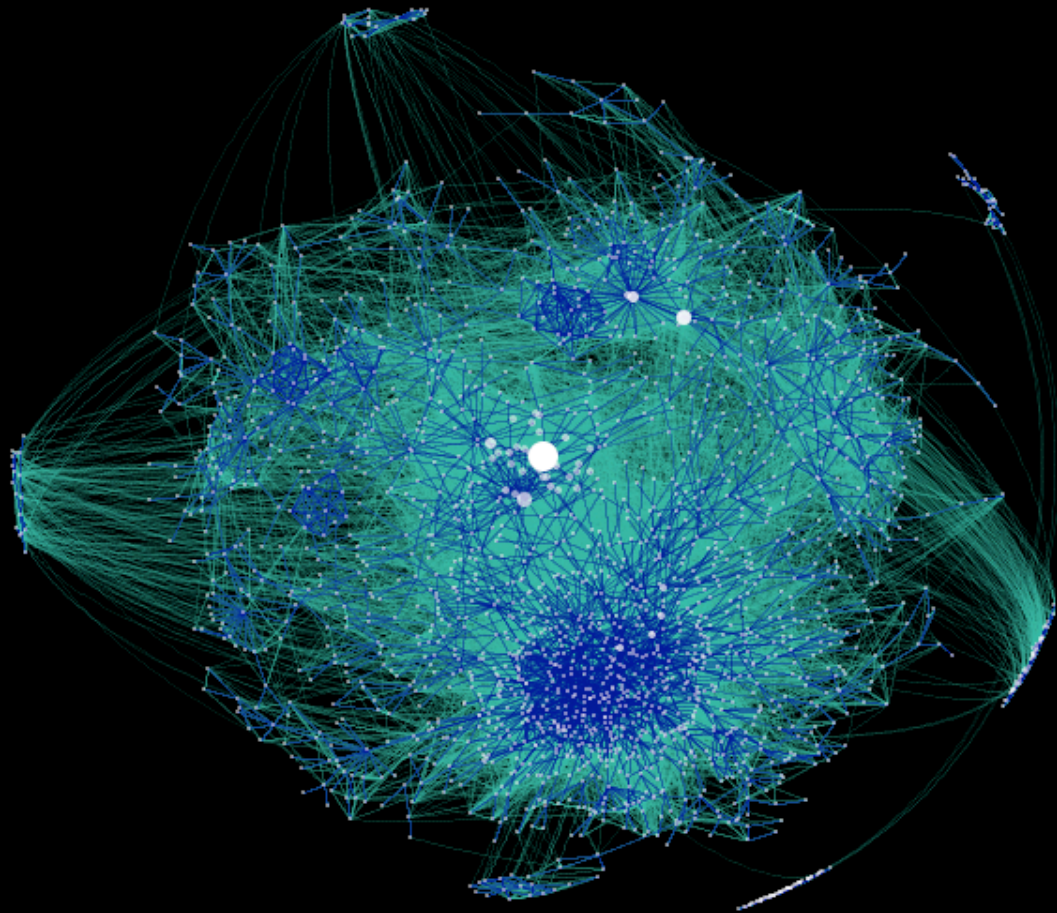
Nowhere more so than on the internet

- Map-Reduce popularized by Google to address the problem of processing large amounts of text data:
 - Many operations with each being a simple operation but done at large scale
 - Indexing a full copy of the web
 - Frequent re-indexing

Reality Check on Brand/ Reputation

What are people saying about my
brand on Social Media?

This is data of people blogging and tweeting about your products on the internet? How do you react to this?



Reality Check

Surely there are companies I can work with that can help me make this practical?

Can you analyze data from these companies and integrate the results with your strategy?

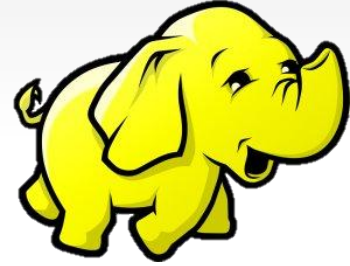


How about in real time and then change your business on the fly?

Reality Check

So what do technology people worry about these days?

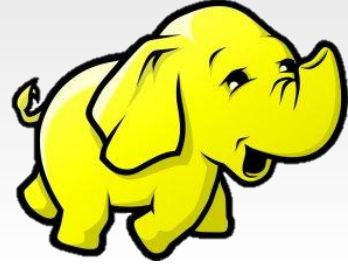
To Hadoop or not to Hadoop?



when to use techniques requiring Map-Reduce and grid computing?

- Typically organizations try to use Map-Reduce for everything to do with Big Data
 - This is actually very inefficient and often irrational
 - Certain operations require specialized storage
 - Updating segment memberships over large numbers of users
 - Defining new segments on user or usage data

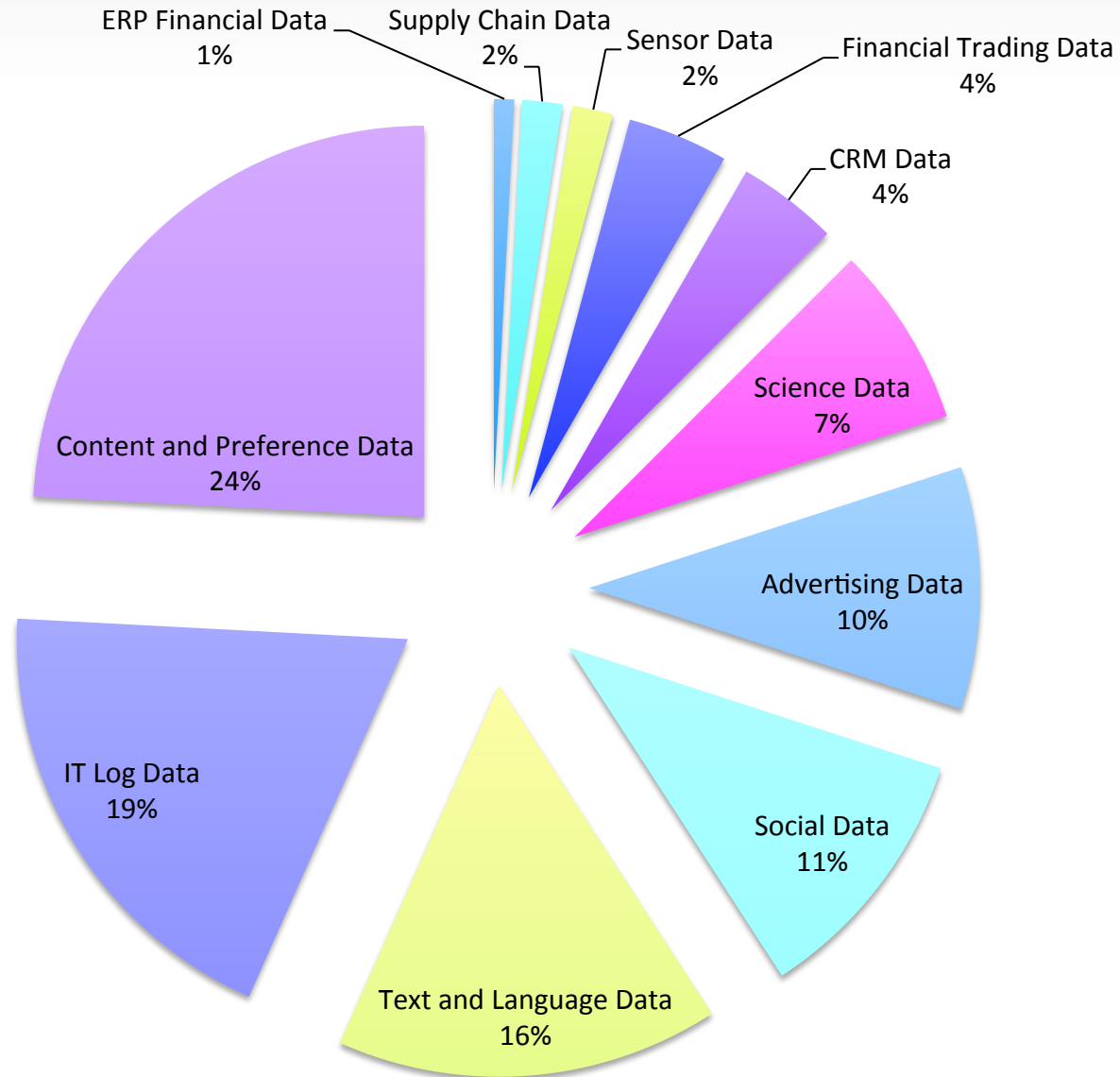
To Hadoop or not to Hadoop?



when to use techniques requiring Map-Reduce and grid computing?

- Map-Reduce is useful when a very simple operation is to be applied on a large body of unstructured data
 - Typically this is during entity and attribute extraction
 - Still need Big Data analysis post Hadoop
- Map-Reduce is not efficient or effective for tasks involving deeper statistical modeling
 - good for gathering counts and simple (sufficient) statistics
 - E.g. how many times a keyword occurs, quick aggregation of simple facts in unstructured data, estimates of variances, density, etc...
 - Mostly pre-processing for Data Mining

Hadoop Use Cases by Data Type



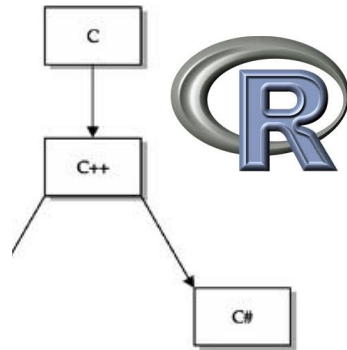
Analysis & Programming Software

SQL

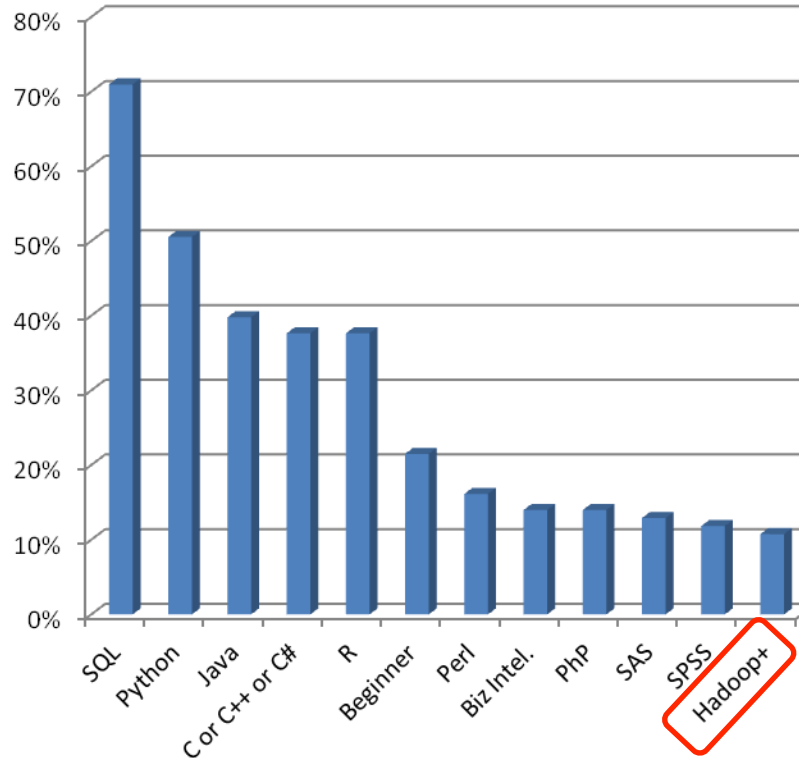
MySQL®

python

Java



Event Survey - Languages



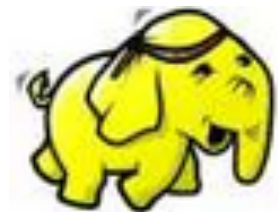
disco

massive data - minimal code

RevolutionAnalytics / RHadoop

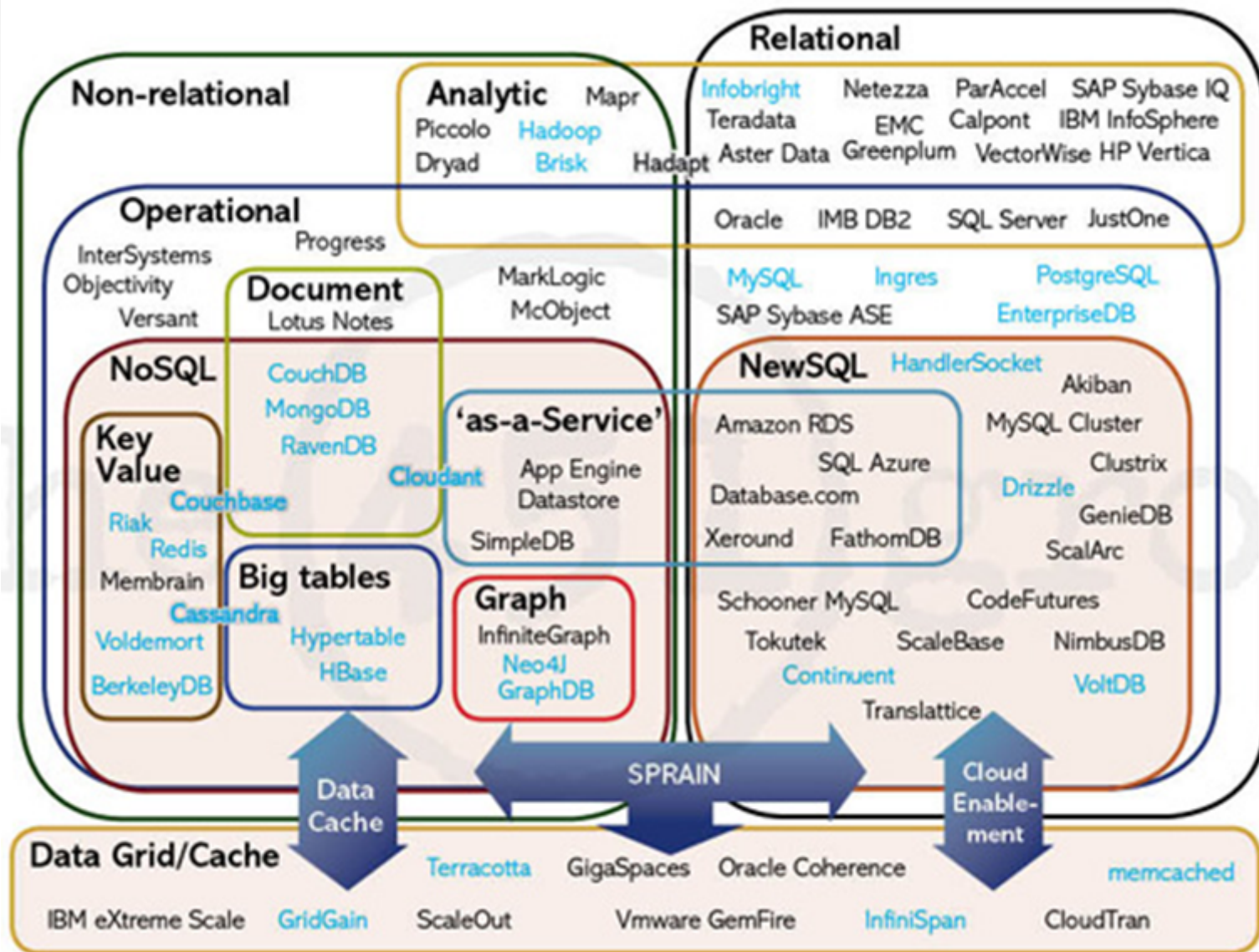


PIG



HIPI





Many Business Uses

Analytic technique	Uses in business
Marketing and sales	Identify potential customers; establish the effectiveness of a campaign
Understanding customer behavior	model churn, affinities, propensities, ...
Web analytics & metrics	model user preferences from data, collaborative filtering, targeting, etc.
Fraud detection	Identify fraudulent transactions
Credit scoring	Establish credit worthiness of a customer requesting a loan
Manufacturing process analysis	Identify the causes of manufacturing problems
Portfolio trading	optimize a portfolio of financial instruments by maximizing returns & minimizing risks
Healthcare Application	fraud detection, cost optimization, detection of events like epidemics, etc...
Insurance	fraudulent claim detection, risk assessment
Security and Surveillance	intrusion detection, sensor data analysis, remote sensing, object/person detection, link analysis, etc...

So Internet is a big place with 2B+ users and lots happening?

- ***Do we understand what each individual is trying to achieve?***
- ***Do we understand what a community's sentiment is?***
- ***Do we understand context and content?***

Social platforms 2013

- **User accounts on FaceBook:**
 - 971M
 - How many fake profiles?
 - 83M (per FB 8/2012 report)
- **How many users on LinkedIn**
 - 159.3M (as of 1/2013)
- **How many Google+ users?**
 - 343 million active users in Q4 2012

» Sources: International Business Times 1/28/2013

- <http://www.ibtimes.com/google-plus-becomes-worlds-no-2-social-network-after-facebook-knocking-twitter-1042956>

How Do People Spend Their On-line Time?

- On-line Shopping? **5%**
- Searches? **21%**
- Email/Communication? **19%**
- Reading Content? **20%**
- Social Networking? **22%**
- Multimedia Sites? **13%**

Interesting Events

- **Google:** How many searches in 2012?
 - More than 1.2 Trillion (source Google)
 - Estimates 1B to 3B per day
- **Twitter:** How many Tweets/day?
 - 500M (per CEO Dick Costolo at IAB Engage, 10/2012)
- **Facebook:** Updates per day?
 - More than 1B
- **YouTube:** Views/day
 - 4 Billion hours/month - 4B views/day in 1/2012
 - 72 hours of video uploaded every minute!
- **Social Networks:** users who have used sites for spying on their partners?
 - **56%**

**Sources: Feb.2012 - compiled from Comscore.datamine.com, Nielsen.com, thisDigitalLife.com, PewInternet.org*

Interesting Events

- Country with Highest online friends?
 - Brazil
 - 481 friends per user
 - Japan has least at 29
- Country with maximum time spent shopping on-line??
 - China: 5 hours/week

**Sources: Feb.2012 - compiled from Comscoredatamine.com, Nielsen.com, thisDigitalLife.com, PewInternet.org*

So Internet is a big place with
lots happening?

- ***Do we understand what each individual is trying to achieve?***
- ***Do we understand what a community's sentiment is?***
- ***Do we understand context and content?***

Turning the 3 V's of Big Data *Into Value*

Turning the three Vs of Big Data into **Value**

Understand context and content

- *What are appropriate ads?*
- *Is it Ok to associate my brand with this content?*
- *Is content sad?, happy?, serious?, informative?*

Understand community sentiment

- *What is the emotion?*
- *Is it negative or positive?*
- *What is the health of my brand online?*

Understand user intent?

- *What is each individual trying to achieve?*
- *Critical in monetization, advertising, etc...*

Understanding Context

Reality Check

So who is the company we think is
best at handling BigData?

Biggest BigData in Advertising?

Understanding Context for Ads

The Display Ads Challenge Today

Search: All News Advanced

Body parts delivered to Michigan home AP Associated Press


1 hour, 23 minutes ago


ELSEWHERE ON THE WEB

CNN.COM
Roadside bombs target Iraqi police, officials say

ABC NEWS
In Wealthy Santa Barbara, Some Call a Parking Lot 'Home'

THE CHRISTIAN SCIENCE MONITOR
Need barbed wire? Try the concierge.

CNN.COM
 **Offbeat News**
Get the latest news, photos and video of the funky, and weird.
» More at CNN.com

USATODAY.COM
 **Offbeat Digest**
Latest news on the strange, quirky and just plain odd.
» More at USATODAY.com

CASCADE TOWNSHIP, Mich. - Two packages containing human body parts — including a liver and part of a head — meant for a medical research lab instead were delivered to a home.

ADVERTISEMENT

The body parts, sent from China, were mistakenly dropped off Thursday at Franck and Ludivine Larmande's home by a DHL express driver who believed the bubble-wrapped items were pieces to a table.

"My husband started to unwrap one and said, 'This is strange, it looks like a liver,'" Ludivine Larmande said. "He started the second one, but stopped as soon as we saw the ear."

"Something wasn't right. It was scary, and I'm glad I didn't open them."

The couple called Kent County sheriff's deputies, who determined the preserved body parts were for medical research, Lt. Roger Parent said.

Authorities believe 28 more bubble-wrapped human organs and body parts could be dispersed across the country, The Grand Rapids Press reported. Two of five packages headed to the northern Michigan lab broke open, scattering their contents.

"There will definitely be a shock to people if they see these things, but there is no hazard to health," Parent said.

What Ad would you place here?

The Display Ads Challenge Today

Damaging to Brand?

Search: All News Advanced

Body parts delivered to Michigan home

AP Associated Press


1 hour, 23 minutes ago


ELSEWHERE ON THE WEB

CNN.COM
Roadside bombs target Iraqi police, officials say

ABC NEWS
In Wealthy Santa Barbara, Some Call a Parking Lot 'Home'

THE CHRISTIAN SCIENCE MONITOR
Need barbed wire? Try the concierge.

CNN.COM
 **Offbeat News**
Get the latest news, photos and video of the funky, and weird.
» More at CNN.com

USATODAY.COM
 **Offbeat Digest**
Latest news on the strange, quirky and just plain odd.
» More at USATODAY.com

CASCADE TOWNSHIP, Mich. - Two packages containing human body parts — including a liver and part of a head — meant for a medical research lab instead were delivered to a home.

ADVERTISEMENT

The body parts, sent from China, were mistakenly dropped off Thursday at Franck and Ludivine Larmande's home by a DHL express driver who believed the bubble-wrapped items were pieces to a table.


"My husband started to unwrap one and said, 'This is strange, it looks like a liver,'" Ludivine Larmande said. "He started the second one, but stopped as soon as we saw the ear."

"Something wasn't right. It was scary, and I'm glad I didn't open them."

The couple called Kent County sheriff's deputies, who determined the preserved body parts were for medical research, Lt. Roger Parent said.

Authorities believe 28 more bubble-wrapped human organs and body parts could be dispersed across the country, The Grand Rapids Press reported. Two of five packages headed to the northern Michigan lab broke open, scattering their contents.

"There will definitely be a shock to people if they see these things, but there is no hazard to health," Parent said.



It's UPS early morning delivery.

« Rollover for video.


The Display Ads Challenge Today

News World news Greece

Violence continues in Greece as rioters firebomb buildings

Protesters in Athens torch offices and cars amid clashes with police after memorial for teenager

Amr Dawar
guardian.co.uk, Sunday 21 December 2008 17:05 GMT
Article history



A youth assaults a police officer in Athens during a week of riots after the shooting of a teenager. Photograph: Bela Scandelszky/AP

World news
Greece

More news:

What Ad would you place here?

The Display Ads Challenge Today

News World news Greece

Violence continues in Greece as rioters firebomb buildings

Protesters in Athens torch offices and cars amid clashes with police after memorial for teenager

Amr Dawar
guardian.co.uk, Sunday 21 December 2008 17:05 GMT
Article history



A youth assaults a police officer in Athens during a week of protests. Photograph: Bela Scandelszky/AP



A larger | smaller

World news
Greece

More news:



Win a mini-break
in gorgeous
Greece...

Irrelevant and
Damaging to Brand

Completely Irrelevant

Related Topics: [make it snacktastic](#) [working it](#) [summer living](#) [summer travel](#) [summer beauty](#)

What happens to your body after you drink a soda every day, for a long time

B by Bliss Tree, on Tue Jun 7, 2011 4:58am PDT
627 Comments | [Post a Comment](#) | [Read More from This Author »](#) | [Report Abuse](#)

f Share

retweet 187

Email

Print



Sugar rushes and caffeine highs followed by a depressing energy crash are **what happens to your body if you drink a soda right now**, but plenty of BlissTree readers actually seem to be okay with that. Some of you think it's alarmist to compare a caffeine and sugar rush to doing drugs, and some just don't really care about the slump they'll find themselves in after drinking 39 grams of sugar, but **what makes us really worried about a soda-slurping**

What is Your Credit Score as of 06/15/2011



See Yours Instantly! [FreeScore.com](#)

Shine on facebook

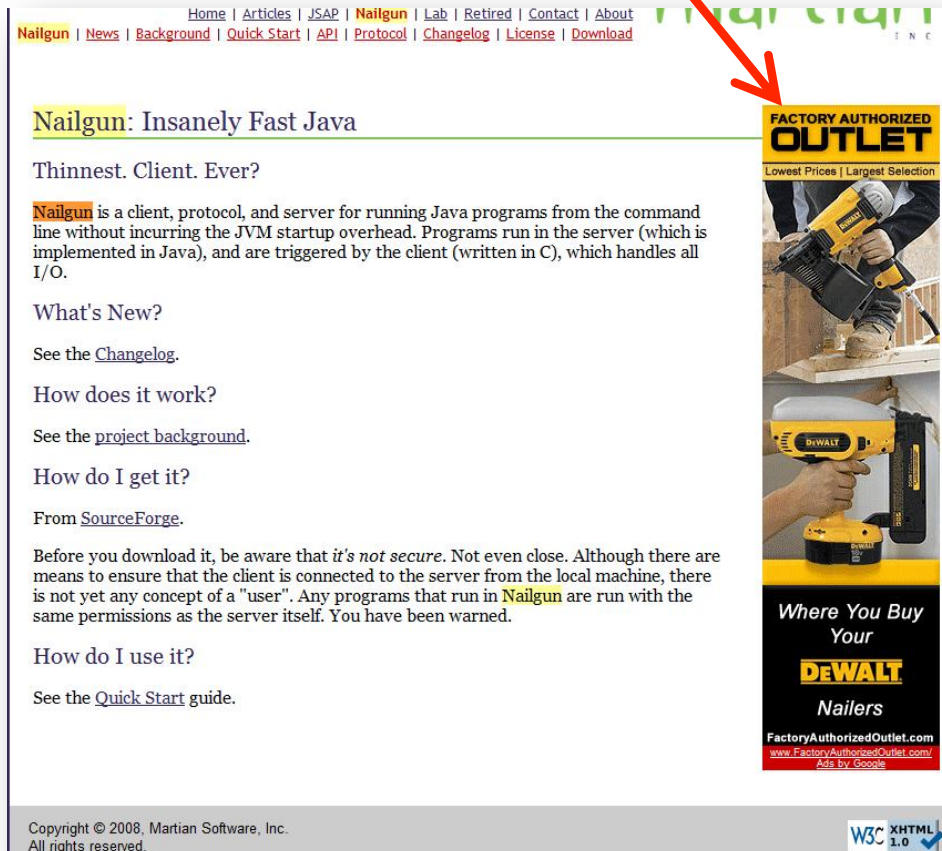
NetSeer: Intent for Display

- Currently Processing 4 Billion Impressions per Day

The image displays two web pages from NetSeer, illustrating the intent for display. The left page is titled 'TRADE-S' and features a sidebar with navigation links such as 'Home', 'FREE Display Quotes', 'Plan for Success', 'Timeline & Checklists', 'Find Trade Shows', 'Booths', 'Exhibits & Displays', 'Banner Stands', 'Design & Graphics', 'Fixtures & Lighting', 'Show Logistics', 'Staffing Strategies', 'Promotions', 'Marketing Strategy', 'Attract Booth Traffic', 'Trade Show Ideas', 'Promotional Items', 'Using Giveaways', 'Booth Entertainment', 'Event Magicians', 'Compare Prices', 'Exhibit Quotes', 'Marketing Services', 'Promo Item Quotes', 'Help & Contacts', 'RESOURCES', 'Contact / Advertising', and 'Site Search'. The main content area features a photo of Rachael Ray and a section titled 'want to meet R & be on the show share your story!'. The right page is titled 'energy fiend' and features a 'COMPLETE GUIDE TO CAFFEINE' section with sub-links for 'Caffeine in Drinks', 'Caffeine in Food', and 'Death by Caffeine'. The main article is titled 'Prevent a Stroke By Drinking Caffeine' and includes a diagram of a human head showing the brain and spine. The sidebar on the right of the 'energy fiend' page contains several advertisements, including Starbucks K-Cup Packs, 125 energy drinks for \$12.50, and a list of recently added products like TEHO Energy Drink, Speed Energy Drink, Chameleon Cold Brew Coffee, Ed Hardy King Dog Energy Mints, and Cocoa Powder (Hershey's).

Problem: Hard to Understand User Intent

Contextual Ad served by Google



The screenshot shows the Nailgun website. At the top is a navigation bar with links: Home | Articles | JSAP | **Nailgun** | Lab | Retired | Contact | About. Below this is a secondary navigation bar: Nailgun | News | Background | Quick Start | API | Protocol | Changelog | License | Download. The main heading is "Nailgun: Insanely Fast Java". The content includes sections for "Thinnest. Client. Ever?", "What's New?", and "How does it work?". A red arrow points from the text "Contextual Ad served by Google" to a yellow and black advertisement for "FACTORY AUTHORIZED OUTLET" featuring DeWalt nailers. The ad text includes "Lowest Prices | Largest Selection", "Where You Buy Your DEWALT Nailers", and "FactoryAuthorizedOutlet.com". At the bottom of the page is a copyright notice: "Copyright © 2008, Martian Software, Inc. All rights reserved." and a "W3C XHTML 1.0" logo.

What NetSeer Sees:



The screenshot shows the NetSeer search results for the URL "http://martiansoftware.com/nailgun/". The NetSeer logo is at the top right. The results are a list of keywords and phrases related to Java programming and development, including: "Programming in Java", "Java Development Kit", "Sun Java", "Java Programmers", "Java Virtual Machine", "Static IP Address", "DNS Servers", "Dynamic DNS", "Java Runtime Environment", "Java Software", "Java Programming Language", "Server Configuration", "Dynamic IP Addresses", "Free Java", "Code in Java", "Implementation of Java", "Local Machine", "Running Java Programs", "Java Language", "Java Platform", "Spring Framework", "Home Server", and "Virtual Server". At the bottom, it says "ContextLinks © by NetSeer" and "www.netseer.com".

Turning the three Vs of Big Data into **Value**

Understand context and content

- *What are appropriate ads?*
- *Is it Ok to associate my brand with this content?*
- *Is content sad?, happy?, serious?, informative?*

Understand community sentiment

- *What is the emotion?*
- *Is it negative or positive?*
- *What is the health of my brand online?*

Understand user intent?

- *What is each individual trying to achieve?*
- *Critical in monetization, advertising, etc...*

User Intent



User Intent Case Study #1

Yahoo! Behavioral Targeting

Yahoo! – One of Largest Destinations on the Web



80% of the U.S. Internet population uses Yahoo!
– Over 600 million users per month globally!

- Global network of content, commerce, media, search and access products
- 100+ properties including mail, TV, news, shopping, finance, autos, travel, games, movies, health, etc.
- **25+ terabytes of data collected each day**
 - Representing 1000's of cataloged consumer behaviors

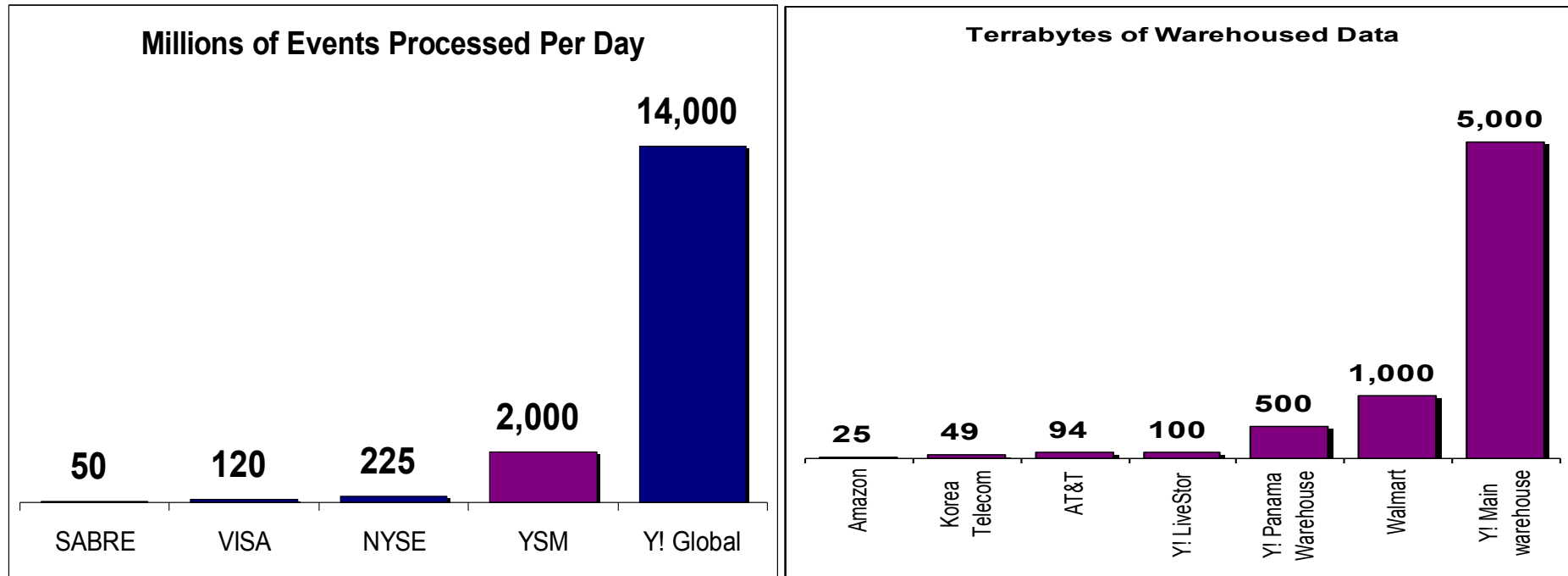
More people visited Yahoo! in the past month than:

- Use coupons
- Vote
- Recycle
- Exercise regularly
- Have children living at home
- Wear sunscreen regularly

Data is used to develop content, consumer, category and campaign insights for our key content partners and large advertisers



Yahoo! Big Data – A league of its own...



GRAND CHALLENGE PROBLEMS OF DATA PROCESSING

TRAVEL, CREDIT CARD PROCESSING, STOCK EXCHANGE, RETAIL, **INTERNET**

Y! Data Challenge Exceeds others by 2 orders of magnitude

Behavioral Targeting (BT)

YAHOO! FINANCE

Monday, June 4, 2007, 12:17

HOME INVESTING NEWS & OPINION PERSONAL FINANCE MY PORTFOLIOS

Enter Symbol(s) GET QUOTES Symbol Lookup Finance Search

Manage the Yahoo! Finance. Check it out.

Market Summary

Streaming Quotes

Stocks 4-10

2620
2615
2610
2605
2600

00m 10m 20m 40m

Symbol LAST CHANGE

Dow	13,642.97	▲ 24.14 (0.18%)
Nasdaq	2,610.91	▲ 3.41 (0.13%)
S&P 500	1,535.75	▲ 6.59 (0.04%)
10-Yr Bond	4.9310%	▲ 0.0250
NYSE Volume	768,163,000	
Nasdaq Volume	872,008,000	

Indices US - World Most Active

NEW! Personal Finance

A Surprise Reprieve for Those Struggling With Mortgage

Video: Apple iPhone Coming Soon FOX Business Now

Apple Unveils iPhone Launch Date AP

Chinese Stocks Take Big Fall AP

Electronics Buys Sealed Air in \$3.6B Deal AP

Market Overview: Mon 12:00 PM ET [Breaking.com](#)

More Top Stories...

YAHOO!

Web Images Video Local Shopping more

Search: Home Loan

My Yahoo! My Mail

Search

Page Options

YAHOO!

BT

YAHOO! SEARCH

Web Images Video Local Shopping more

Search: Home Loan

Search Results

Also try: [home loan calculator](#), [countrywide home loan](#), [home savings and loan](#) More...

- Countrywide® Home Loans**
[www.countrywide.com](#) - Less-than-perfect credit expertise from Countrywide®, 4 of 5 approved.
- Home Loans - LendingTree**
[www.lendingtree.com](#) - Refinance \$200,000 for \$667/Month. Refinance Offers, Quality Online.
- Home Loan**
[www.quickenloans.com](#) - Get a \$200,000 Home Loan for Only \$585 / Month - Apply Online Today.
- Home Loans - Ditech®**
[www.ditech.com](#) - Fast & Easy Home Loan Process from Ditech. No Hidden Fees - Apply Now.

Yahoo! Report bad results or ads. Bucket test: None [GSP2]

- Home Loans - Equity, Refinance, Mortgage & Auto | LendingTree**
LendingTree - Your ARM could increase as much as 10%. Lock in \$930/mo. 15 year Fixed. 1 simple form, get up to \$200,000. Quick Links: [Refinance Now](#) - [Fixed Home Equity](#) - [Mortgage](#) [www.lendingtree.com](#) - 84k - [More from this site](#)
- Quicken Loans - Home loans, Refinancing, Interest-only options**
Quicken Loans - Get information, check rates, and learn about refinancing your current home loan. Compare mortgage options, apply online, get pre approved and close fast. Quick Links: [Get Mortgage Rates](#) - [Compare Home Loans](#) - [Interest-Only Loans](#) [www.quickenloans.com](#) - [More from this site](#)

Yahoo! My Yahoo! Mail Help

YAHOO! FINANCE Welcome, dang_97 [Sign Out, My Account]

Web Finance Pictures Video Audio

HOME INVESTING NEWS & OPINION PERSONAL FINANCE MY PORTFOLIOS

Enter Symbol(s) GET QUOTES Symbol Lookup Finance Search

Monday, June 4, 2007, 12:26PM ET - U.S. Markets close in 3 hours and 35 minutes.

Loans Autos Education Home Personal & Small Biz

Banking & Budgeting Career & Work College & Education Family & Home Insurance Loans Real Estate Retirement

Home

Mortgage Brokers: Are They Your Friends or Foes?

Borrowers often see mortgage brokers as allies, but many brokers don't put customers' interests first. Here's what you should know... [read more](#)

HOW-TO GUIDES

- Mortgage Basics
- Financing a New Car
- Home Refinancing Basics
- College Financial Aid Programs

View more how-to guides

CALCULATORS

- Comprehensive mortgage

ADVERTISEMENT

Sponsored by: **Countrywide® HOME LOANS**

Call for a No Cost Refi**

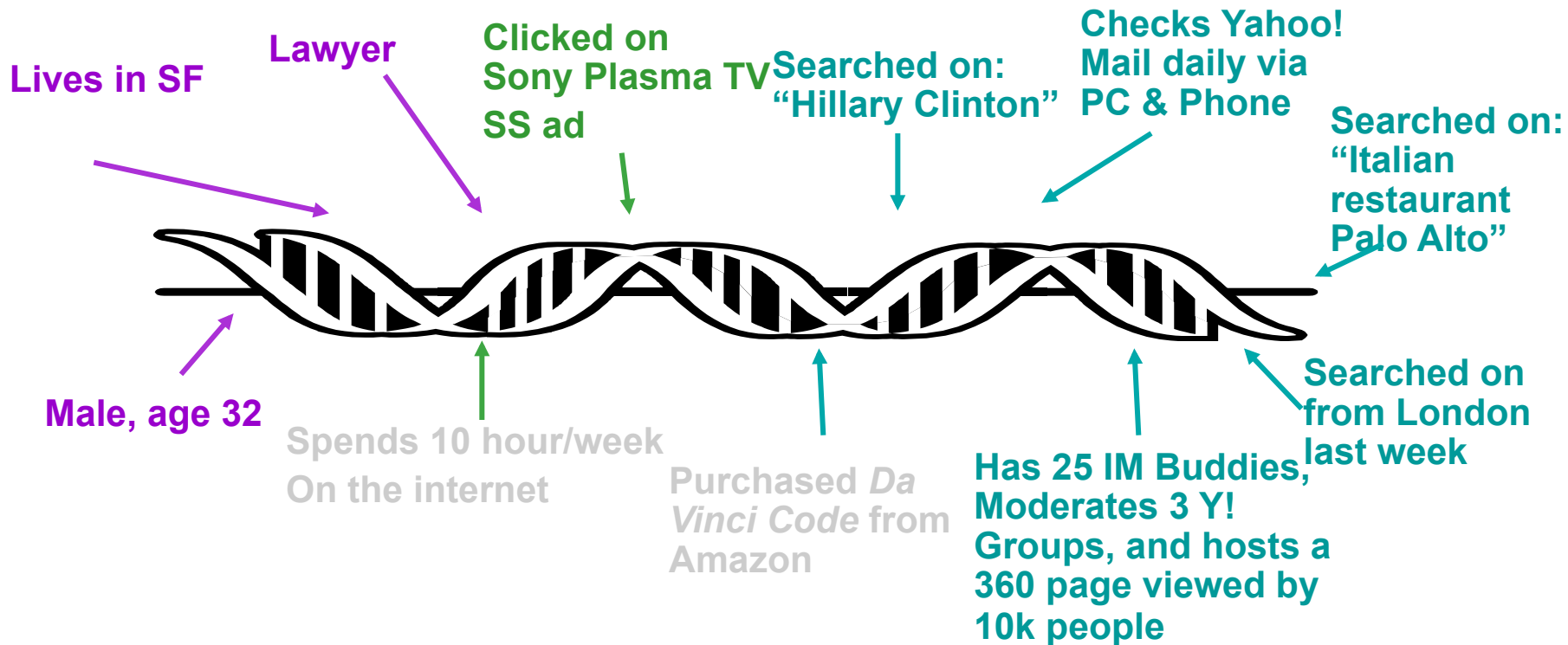
America's #1 Home Loan Lender **1-800-687-9414**

- Refinance
- Consolidate Debt
- Cash Out Refinance
- Loans for Self-Employed Borrowers
- Can I lock in a low rate now?
- How much cash can I get?
- Home Equity Loans
- Home Improvement
- Bad Credit Loans
- How much can I save each month?
- Flexible Payment Options
- Jumbo Loans

Targeting ads to consumers whose recent behaviors online indicate which product category is relevant to them

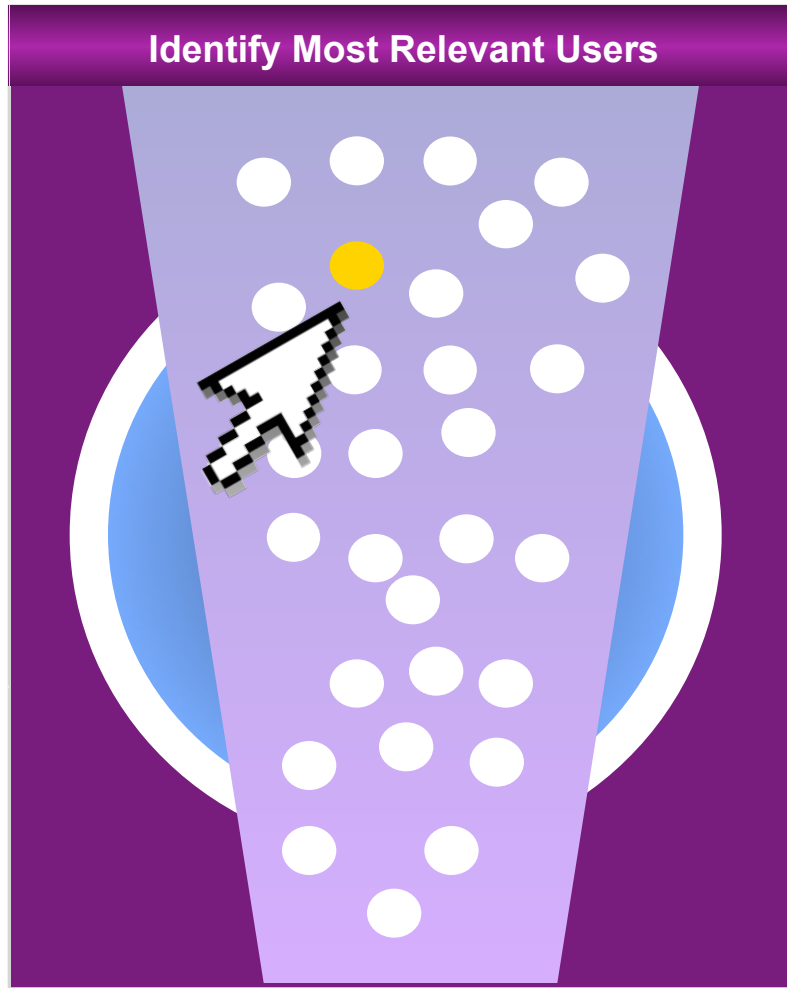
Yahoo! User DNA

Registration Campaign Behavior Unknown



- On a per consumer basis: maintain a behavioral/interests profile and profitability (user value and LTV) metrics

How it works | Network + Interests + *Modelling*



Analyze predictive patterns for purchase cycles in over 100 product categories



In each category, build models to describe behaviour most likely to lead to an ad response (i.e. click).



Score each user for fit with every category...daily.



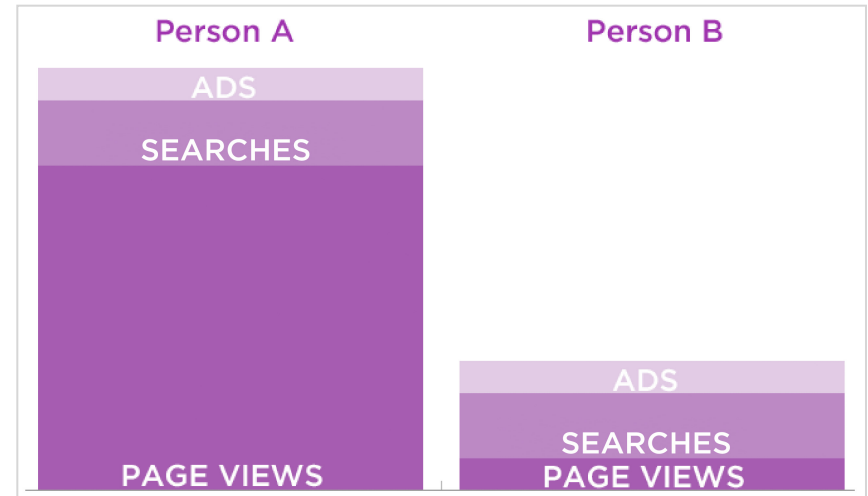
Target ads to users who get highest 'relevance' scores in the targeting categories

Recency Matters, So Does Intensity

Active now...

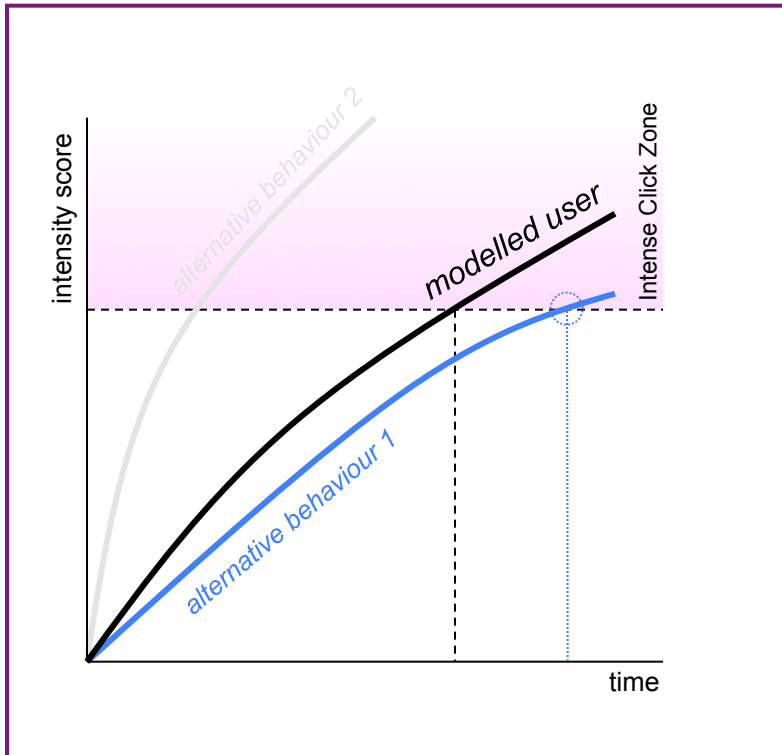


...and with feeling



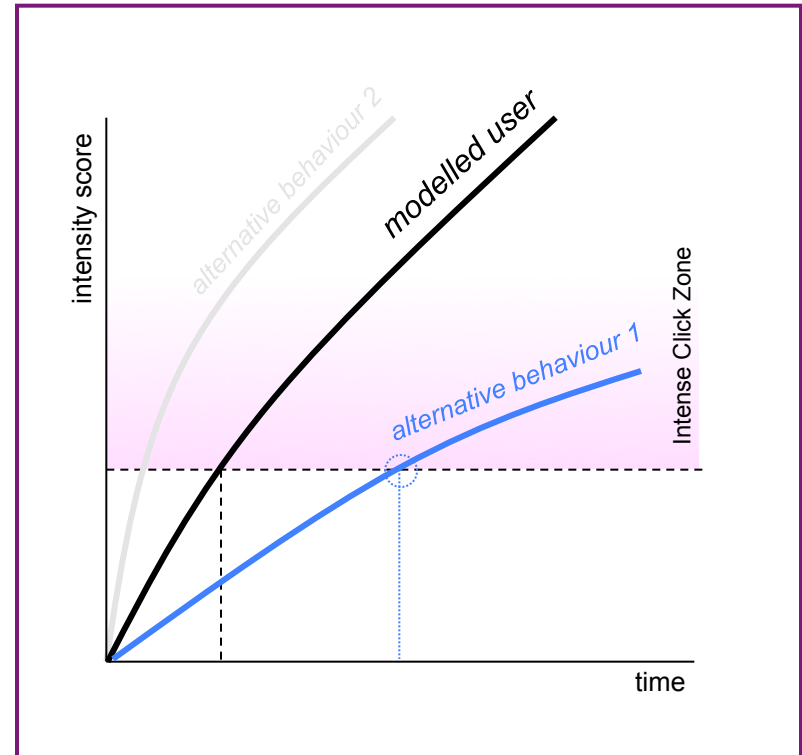
Differentiation | Category specific modelling

Example 1: Category Automotive



Alt Behaviour 1: 5 pages, 2 search keywords, 1 search click, 1 ad click

Example 2: Category Travel/Last Minute

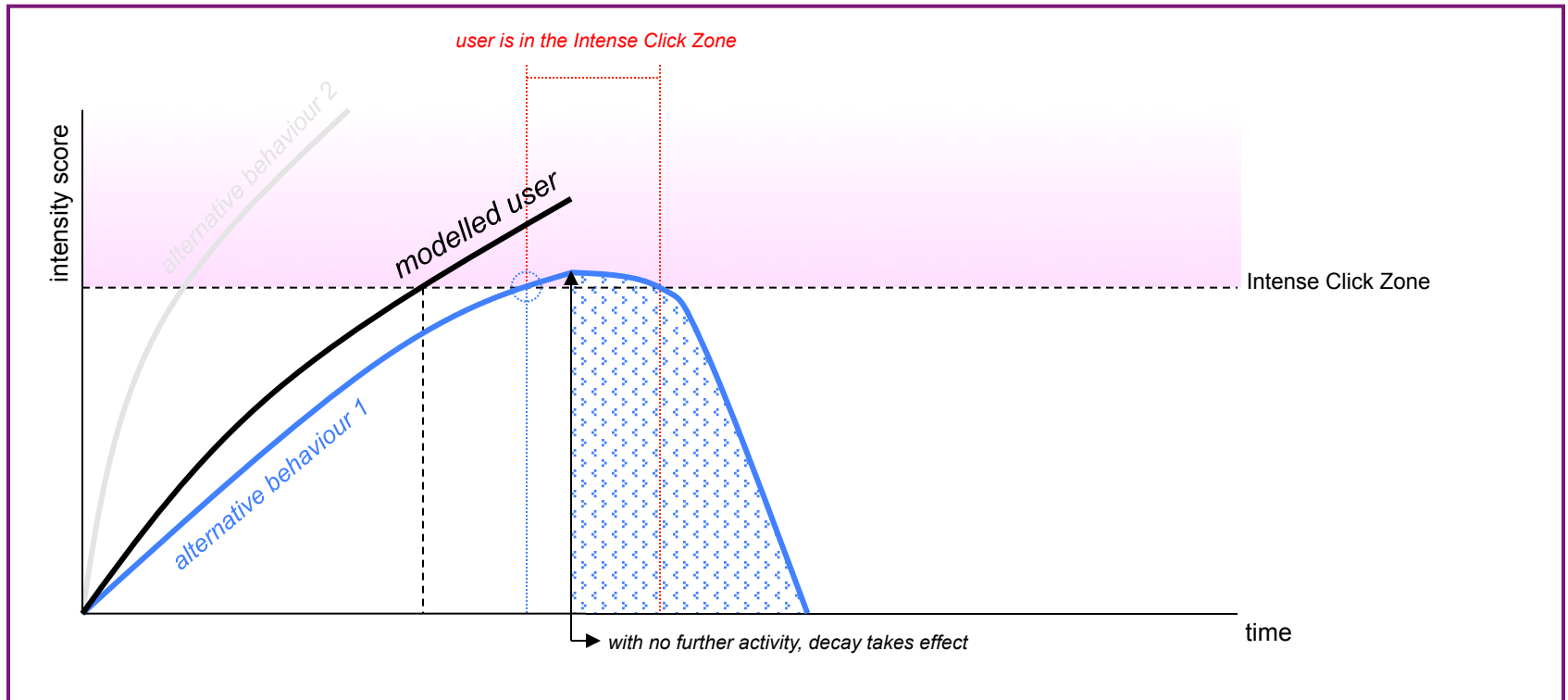


Alt Behaviour 1: 5 pages, 2 search keywords, 1 search click, 1 ad click

Different models allow us to weight and determine intensity and recency

Differentiation | Category specific modelling

Example 1: Category Automotive



Alt Behaviour 1: 5 pages, 2 search keywords, 1 search click, 1 ad click

Different models allow us to weight and determine intensity and recency

Automobile Purchase Intender Example

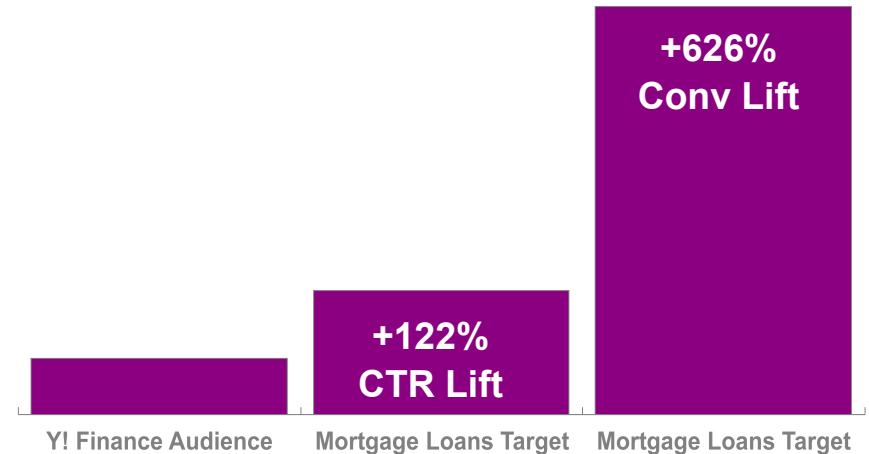
- A test ad-campaign with a major Euro automobile manufacturer
 - Designed a test that served the same ad creative to test and control groups on Yahoo
 - Success metric: performing specific actions on Jaguar website
- Test results: 900% conversion lift vs. control group
 - Purchase Intenders were 9 times more likely to configure a vehicle, request a price quote or locate a dealer than consumers in the control group
 - ~3x higher click through rates vs. control group

Mortgage Intender Example

Example: Mortgages

We found:
1,900,000 people looking
for mortgage loans.

Results from a client campaign on Yahoo!
Network



Example search terms qualified for this target:

Mortgages Home Loans Refinancing Ditech

Example Yahoo! Pages visited:

Financing section in Real Estate

Mortgage Loans area in Finance

Real Estate section in Yellow Pages

Source: Campaign Click thru Rate lift is determined by Yahoo! Internal research. Conversion is the number of qualified leads from clicks over number of impressions served. Audience size represents the audience with the highest propensity to engage with a brand or product and to click on an offer.

Experience summary at Yahoo!

- Dealing with one of the largest data sources (25 Terabyte per day)
- BT business was grown from \$20M to about \$500M in 3 years of investment!
- BigData critical to operations
 - Ad targeting creates huge value
 - Right teams to build technology (3 years of recruiting)
 - Search is a BigData problem,
- Big demands for grid computing (Hadoop)
 - Not all BigData can be handled via Hadoop
 - Spunoff BigData Segmentation data platform: nPario

Lessons Learned

A lot more data than qualified talent

- Finding talent in BigData is very difficult
- Retaining talent in BigData is even harder
- At Yahoo! we created central group that drove huge value to company
- Data people need to feel like they have critical mass
 - Makes it easier to attract
 - Makes it easier to retain
- Drive data efforts by business need, not by technology priorities
 - Chief Data Officer role at Yahoo! – now popular

BigData Analytics for Organizations

- Key to competitive Intelligence:
 - Understand context
 - Understand intent
- Key to understanding consumer trends through social media analysis
 - Brand issues
 - Trend issues
 - Anticipating the next shift

Big Picture on Big Data Analytics

Key points

Sometimes,
Simple is Very Powerful!

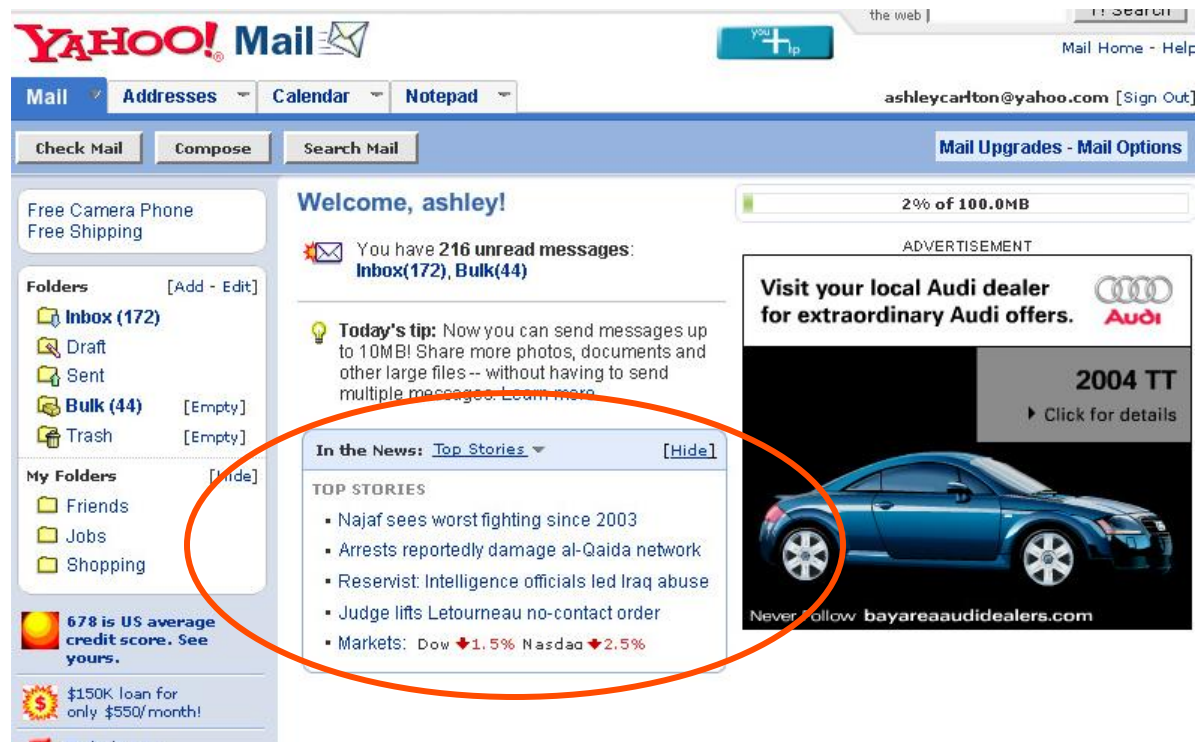
Retaining New Yahoo! Mail Registrants

Integrating Mail and News

- Data showed that users often check their mail and news in the same session
 - But no easy way to navigate to Y! News from Y! Mail
- Mail users who also visit Y! News are 3X more active on Yahoo
 - Higher retention, repeat visits and time-spent on Yahoo

“In the news” Module on Mail Welcome Page

- Increased retention on Mail for light users by 40%!
 - Est. Incremental revenue of **\$16m a year** on Y! Mail alone



Threats & Opportunities

- Data world is changing, especially in on-line businesses
- Major shifts from relational DB to NoSQL, document-oriented stores
- Connecting new world to “old” world?
 - Convenience of execution – integration with data platforms
 - Appropriateness of algorithms to BigData
 - Unstructured data algorithms:
 - Text, Semi-structured and Unstructured data
 - Entity extraction a must
 - Appropriate theory and probability distributions (power laws, fat tails)
 - Sparse Data
 - Model management and proper aging of models
 - Getting to basics so we can decide what models to use:
 - Understanding noise and distributions
 - Data tours

Thank You! & Questions



Usama Fayyad - usama@fayyad.com
usama_fayyad@yahoo.com



Twitter – [@Usamaf](https://twitter.com/Usamaf)



+1-206-529-5123



www.BlueKangaroo.com
www.Oasis500.com
www.open-insights.com



Blue Kangaroo™

