strategy&
Formerly Booz & Company

Web and social media analytics

A data and technology perspective

pwc

Contacts

Chicago

Raj Parande
Principal
+1-312-578-4675
rajendra.parande@strategyand.pwc.com

Yuri Goryunov
Principal
+1-312-578-4791
yuri.goryunov@strategyand.pwc.com

Florham Park, NJ

Ramesh Nair
Partner
+1-973-410-7673
ramesh.nair@strategyand.pwc.com

Steffen Gnegel also contributed to this report.

This report was originally published by Booz & Company in 2011.

What is Web analytics, and why is it so essential but challenging for today's businesses?

What is web analytics?

Definition: Web analytics is the measurement, collection, analysis, and reporting of Internet data for the purposes of understanding and optimizing Web usage. (Source: Web Analytics Association)

Why is it important?

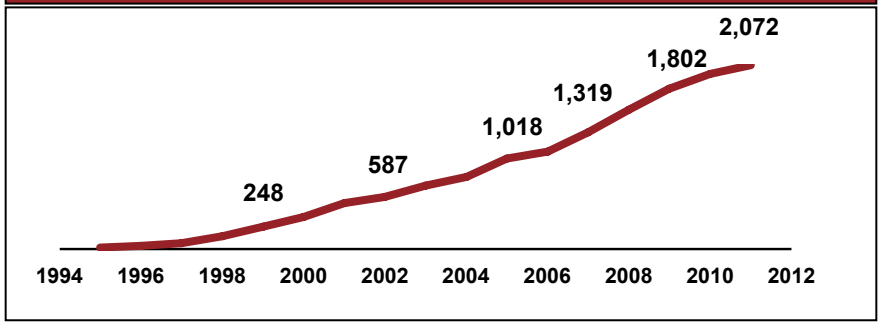
- Today, Web interactions between commercial businesses and their customers take place via e-commerce stores, customer service sites, interactive real-time chat, e-mail, and social media streams. These interactions are as important, if not more so, for a business's growth as customer touches through traditional voice and bricks-and-mortar channels
- Web data integrated with other channels provides a better picture of the customer–business relationship and helps in identifying customer trends
- It is also useful in assessing the effectiveness of marketing campaigns and optimizing marketing spend
- And it improves the customer experience through faster service, thereby driving business growth and enhancing reputation

What are the challenges?

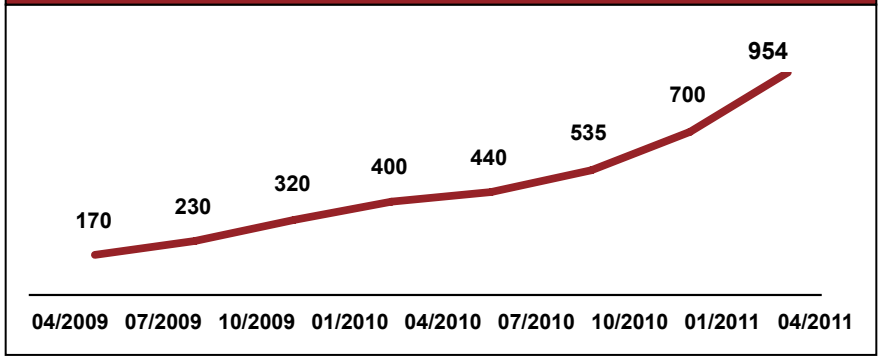
- Data volume growth is accelerating, making it cumbersome to capture and analyze Web data
- Unstructured social media data growth compounds the challenge, particularly as it must be integrated with enterprise structured data
- Multiple Web interaction platforms (PC, smartphone, tablet) further add to data capture and integration challenges
- Location and other smartphone sensor-based feeds also increase the complexity of continuous/real-time data capture
- There is no single tool available to capture and analyze all types of data

The number of Internet, social media, and mobile users tripled over the past decade, reaching a third of the world's population

World Internet users, Dec. 1994–Mar. 2011
(in millions of people)



Mobile data consumed, June 2009–Mar. 2011
(in millions of megabytes)



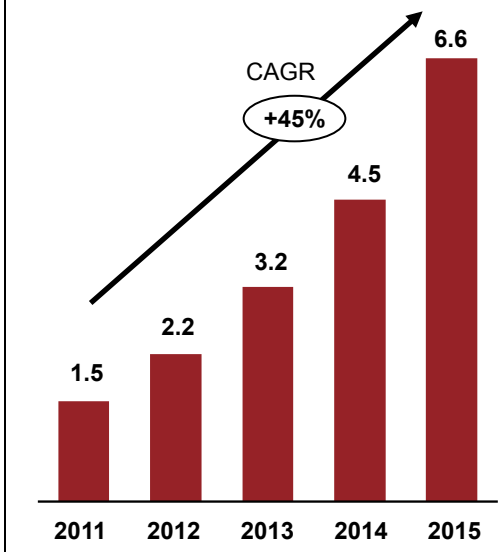
Observations

- In 2000, there were 390 million Internet users in the world
- By March 2011, there were 2.1 billion Internet users
- The number includes 78% of North Americans and 1 billion people in Asia and the Middle East combined
- More than 800 million people use Facebook, with Americans spending over 53 billion minutes a month on the site
- About 350 million currently active Facebook users access it from their mobile devices
- Twitter users are also generating more than 1 billion tweets each week
- At 140 characters per message, Twitter users alone generate nearly 500 gigabytes of information, the equivalent of 500 Encyclopaedia Britannicas, *every month*

Source: Internet Usage Statistics (www.internetworldstats.com/stats.htm); Facebook Press Room (www.facebook.com/press/info.php?statistics); Twitter Blog (blog.twitter.com/2011/03/numbers.html); Opera Software State of the Mobile Web (media.opera.com/media/smw/2011/pdf/smw042011.pdf); Strategy& analysis

In the consumer space alone, Internet-based social and commerce markets represent a multibillion-dollar opportunity

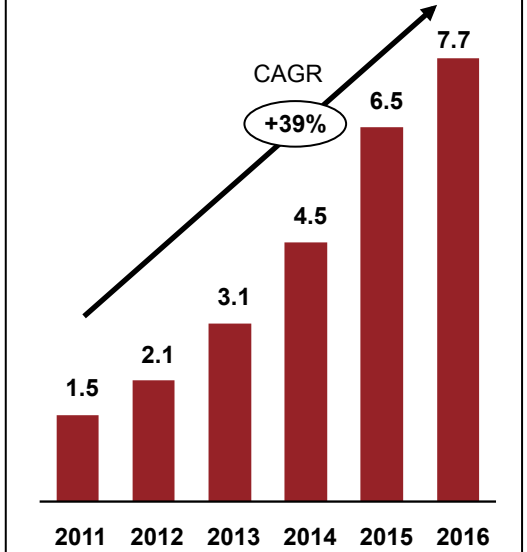
Social gaming
(in US$ billion)



Players

- Playfish
- Zynga
- Playdom

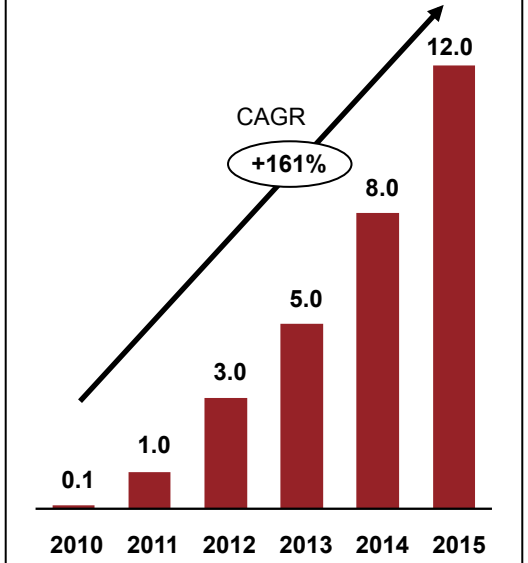
Mobile commerce
(in US$ billion)



Players

- Gilt Groupe
- eBay
- Amazon

Social commerce
(in US$ billion)



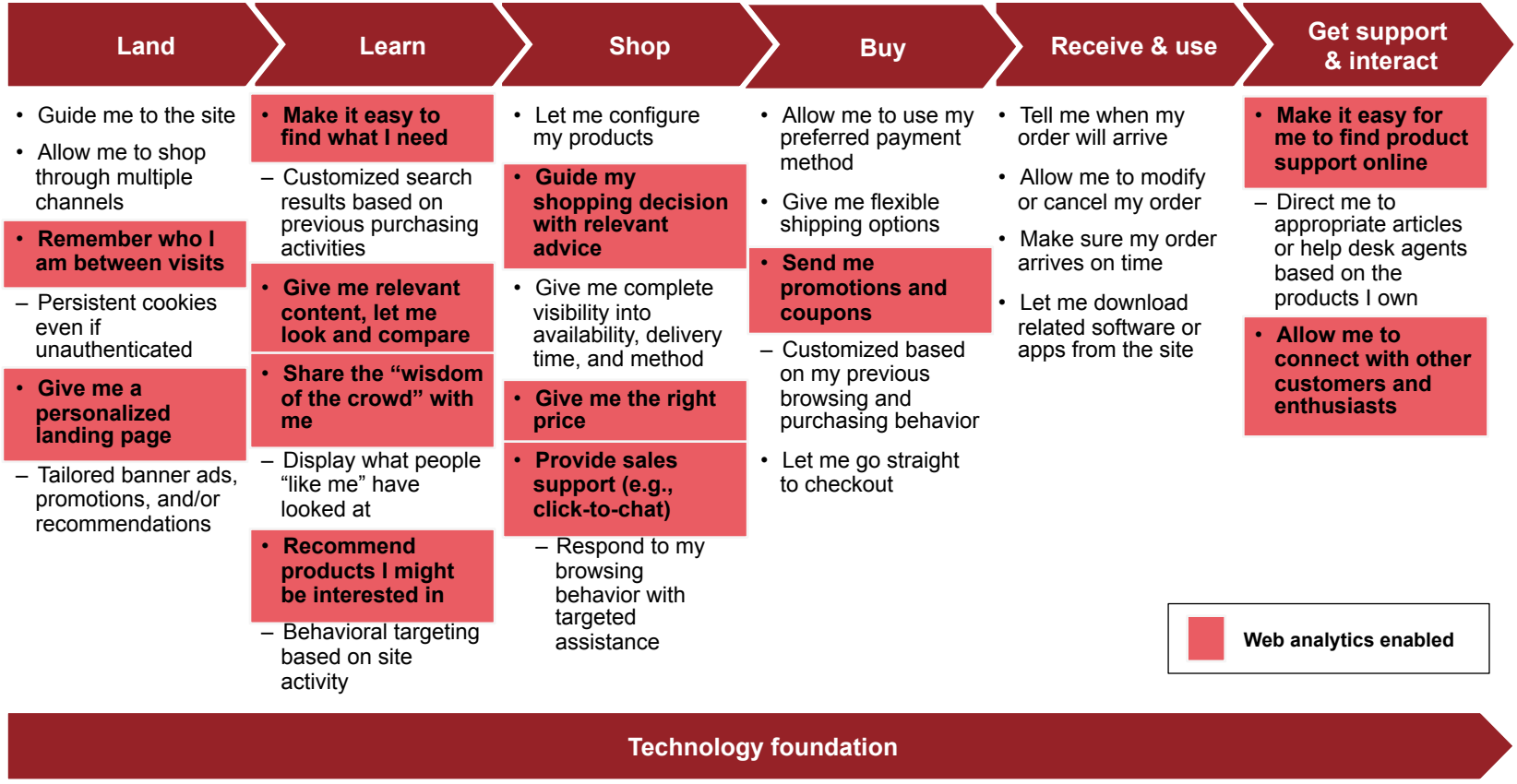
Players

- Groupon
- Yelp
- Living Social

Source: Strategy& research and analysis

Customers have very high expectations for their end-to-end online experience

Customer expectations while browsing websites



Source: Strategy& analysis

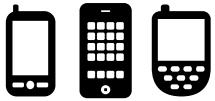
Customer analytics example: Amazon recommends targeted products based on crowd user behavior or specific user profile data

1

A prospect is shopping on Amazon for a smartphone

Customer analytics

Through insights generated from users' purchasing behaviors…



…the prospect is provided with product recommendations for similar phones

2

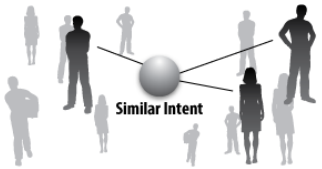
A customer logs onto his/her Amazon account

Based on previous purchase history…

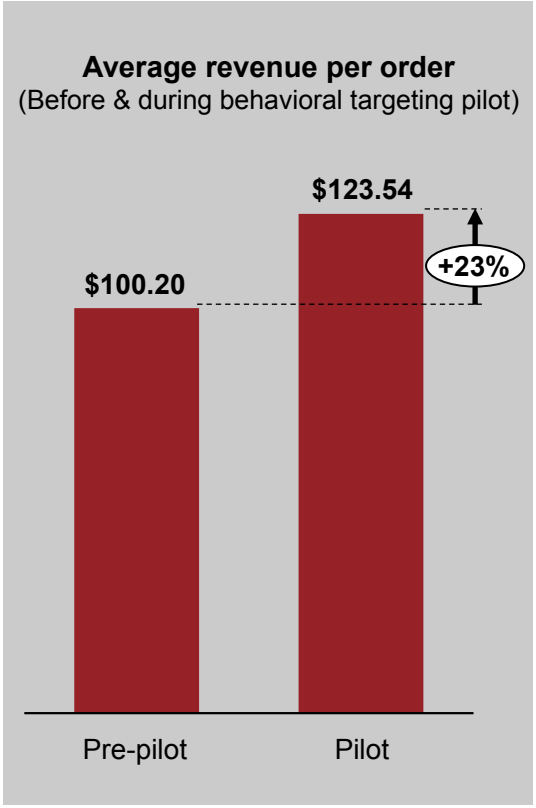
…the customer is provided with recommendations for similar products or accessories

Source: Strategy& analysis

Web analytics example: A client was able to significantly increase average order value by leveraging online data for behavioral targeting

Observe & learn	Recognize	Target
		
• Customer interactions on a webpage • Site navigation patterns • Repeat visitor activity • Search context • Purchase/conversion history • User profile data	• Individual behavioral patterns • Customer segment classification • Wisdom of the crowd – "Hot" products and content – Product affinities (customers who viewed/ bought X also viewed/ bought Y) – Popular search hits and misses	• Product recommendations – *Increased average order value (AOV)* Focus of pilot • Content personalization – *Increased conversion, loyalty* • Search personalization – *Increased conversion, loyalty*

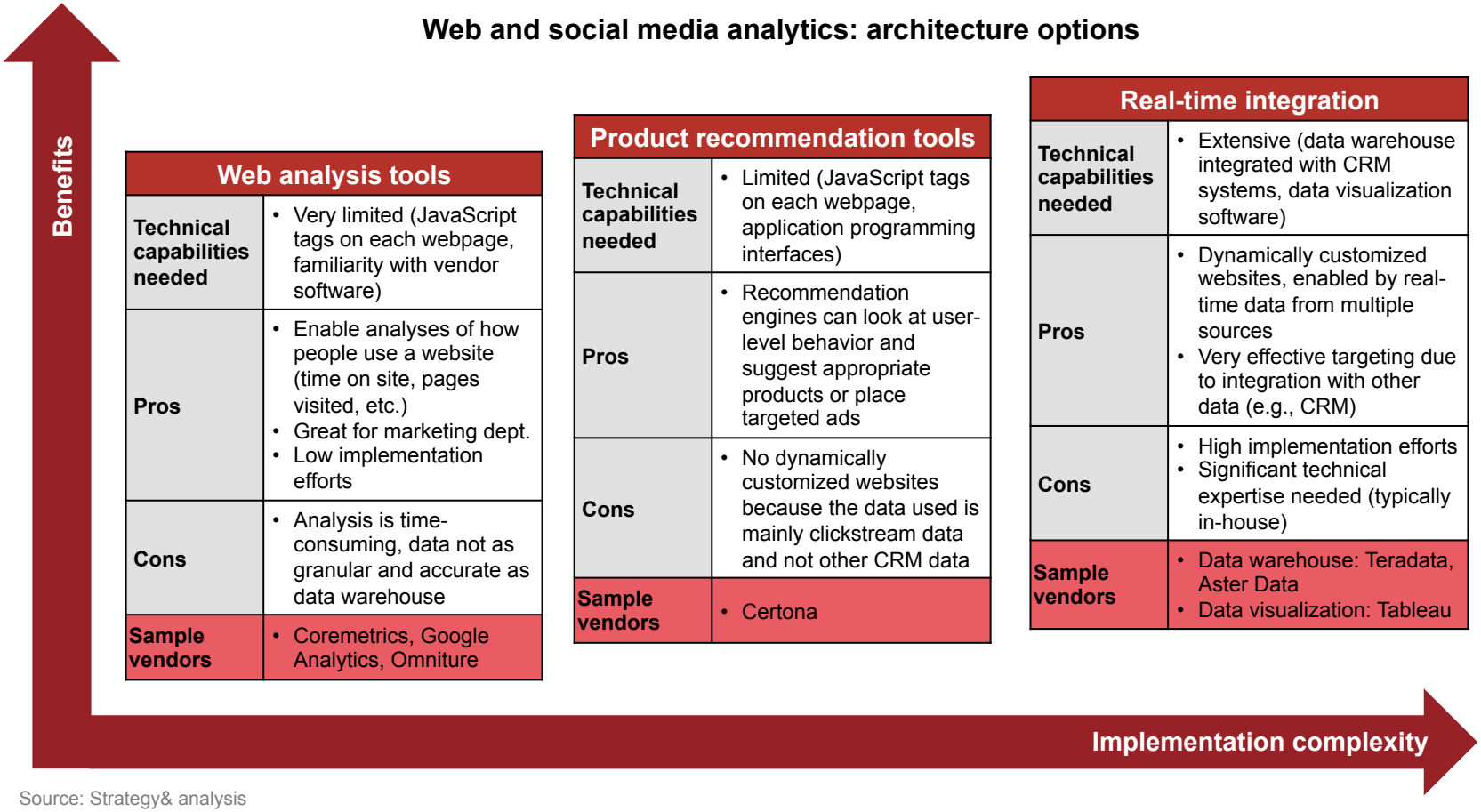
Support tools
(performance reporting, multivariate testing, site optimization)



Source: Strategy& analysis

Companies have to migrate from a Web analysis tool infrastructure to an integrated architecture to enable a customized user experience

Web and social media analytics: architecture options



Web analysis tools	
Technical capabilities needed	• Very limited (JavaScript tags on each webpage, familiarity with vendor software)
Pros	• Enable analyses of how people use a website (time on site, pages visited, etc.) • Great for marketing dept. • Low implementation efforts
Cons	• Analysis is time-consuming, data not as granular and accurate as data warehouse
Sample vendors	• Coremetrics, Google Analytics, Omniture

Product recommendation tools	
Technical capabilities needed	• Limited (JavaScript tags on each webpage, application programming interfaces)
Pros	• Recommendation engines can look at user-level behavior and suggest appropriate products or place targeted ads
Cons	• No dynamically customized websites because the data used is mainly clickstream data and not other CRM data
Sample vendors	• Certona

Real-time integration	
Technical capabilities needed	• Extensive (data warehouse integrated with CRM systems, data visualization software)
Pros	• Dynamically customized websites, enabled by real-time data from multiple sources • Very effective targeting due to integration with other data (e.g., CRM)
Cons	• High implementation efforts • Significant technical expertise needed (typically in-house)
Sample vendors	• Data warehouse: Teradata, Aster Data • Data visualization: Tableau

Source: Strategy& analysis

Providing a rich experience requires a robust analytic capability, integrating disparate sources of structured and unstructured data

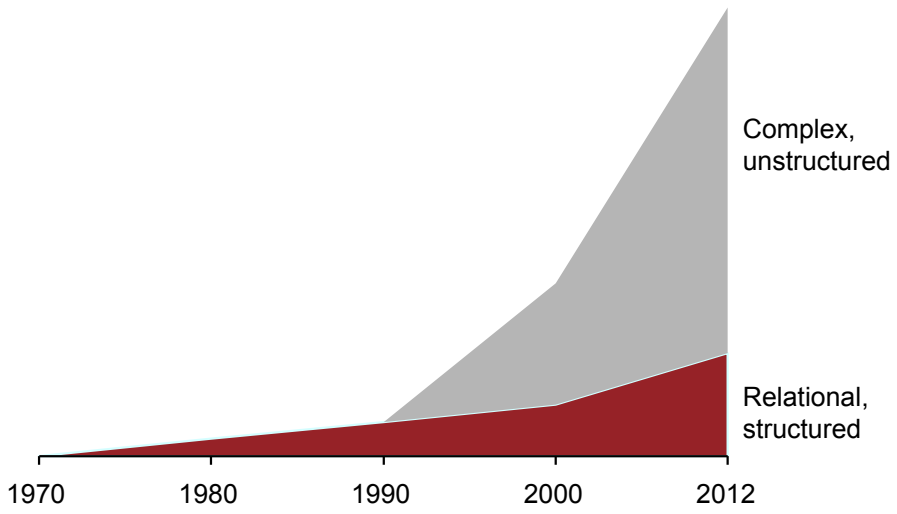
	Example analytics	Data used
Customer analytics	• Targeting promotions and personalizing offers (e.g., customized mailing, rewards, coupons) • Product recommendations	• Customer purchasing behavior • Purchase history
Marketing analytics	• Optimizing marketing mix and promotions • Pricing optimization and demand sensitivity	• Marketing response data • Pricing sensitivity data
Web analytics	• Customer online activity analysis • Sentiment analysis	• Web activity data • Customer social media posts
Operational analytics	• Demand and inventory forecasting • Localization • Supply chain analysis • Workforce optimization	• Demand data • Inventory data • Location data (Web usage, smartphone) • Distribution data • HR data
Fraud & risk analytics	• Fraud analytics • Shrinkage analysis	• Customer interaction data • Purchase returns data • Inventory data

Source: Strategy& analysis

But the explosion of unstructured data volumes requires new approaches to data consolidation and analytics applications

ILLUSTRATIVE

Structured and Unstructured Data Evolution



Relational, Structured Data		Complex, Unstructured Data	
• CRM	• Inventory	• Documents	• SharePoint
• Financials	• Sales records	• Web feeds	• Sensor data
• Logistics	• HR records	• System logs	• Audio
• Data marts	• Web profiles	• Online forums	• Images/video

Change Drivers
1. **Speed:** Data access speeds of physical storage mechanisms have not kept up with improvements in network speeds
2. **Scale:** Traditional data storage techniques like RDBMS have limited scalability to manage growing data volumes (clustering beyond a handful of servers is notoriously difficult)
3. **Integration:** Today's data processing tasks increasingly need to access and combine data from many different unstructured sources, often over a network
4. **Volume:** Data volumes have grown from tens of gigabytes in the 1990s to hundreds of terabytes and often petabytes in recent years

Source: IDC white paper sponsored by EMC and Cloudera

Unstructured data integration and analytics face multiple challenges, but they can be overcome with some new innovations

ILLUSTRATIVE

	Leading Web analytics and industry trends	Vendors and products
Capturing & analyzing multiple streams of data	• Social feed integration with Web and warehouse data for advanced customer analytics • Task- or page-targets-based unobtrusive, short, highly actionable, quick feedback data supplementing site surveys	• Google Analytics, Adobe/Omniture, IBM/Coremetrics/Unica • ForeSee
Multiple platforms for customers to interact on	• Added sources of data and complexity of integration from multiple platforms and form factors (smartphones, tablets) • Complexity of integrating structured data with unstructured feeds from Web, social media, chat, and Internet-connected televisions	• Google Android, Apple iOS, RIM BlackBerry • Google TV, Boxee, Apple TV
Increasing volume of data at a faster rate	• New analytics, storage, and processing for accelerated integration at lower costs due to exponential growth of “big data” needs • No single tool to capture and analyze massive Web data, requiring concurrent use of multiple analytic tools	• Google MapReduce, Apache Hadoop, Google Caffeine • Google Analytics, Omniture, SAS
Continuous Data streams	• Targeted offers based on customer location tracking enabled by GPS and cell-based tracking mechanisms in smartphones • Interaction opportunities from tracking customer check-ins at vendor locations using new services	• Google Latitude • Foursquare • Gowalla
Regulations	• Browser-based features for customer to opt out of tracking • Upcoming regulations like FTC “Do Not Track” initiatives • Android- and iOS-based app developers self-regulating and asking customer permission for data collection	• Mozilla Firefox, Google Chrome, Opera

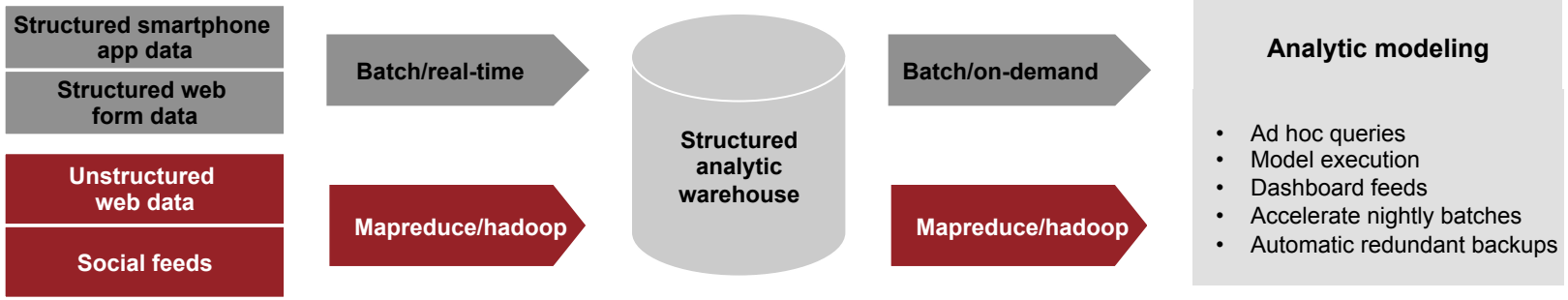
Source: Strategy& analysis

To meet the challenges and gain the benefits of integrating Web and enterprise data, multiple technology enhancements are needed

Enhancements to Web and enterprise analytics

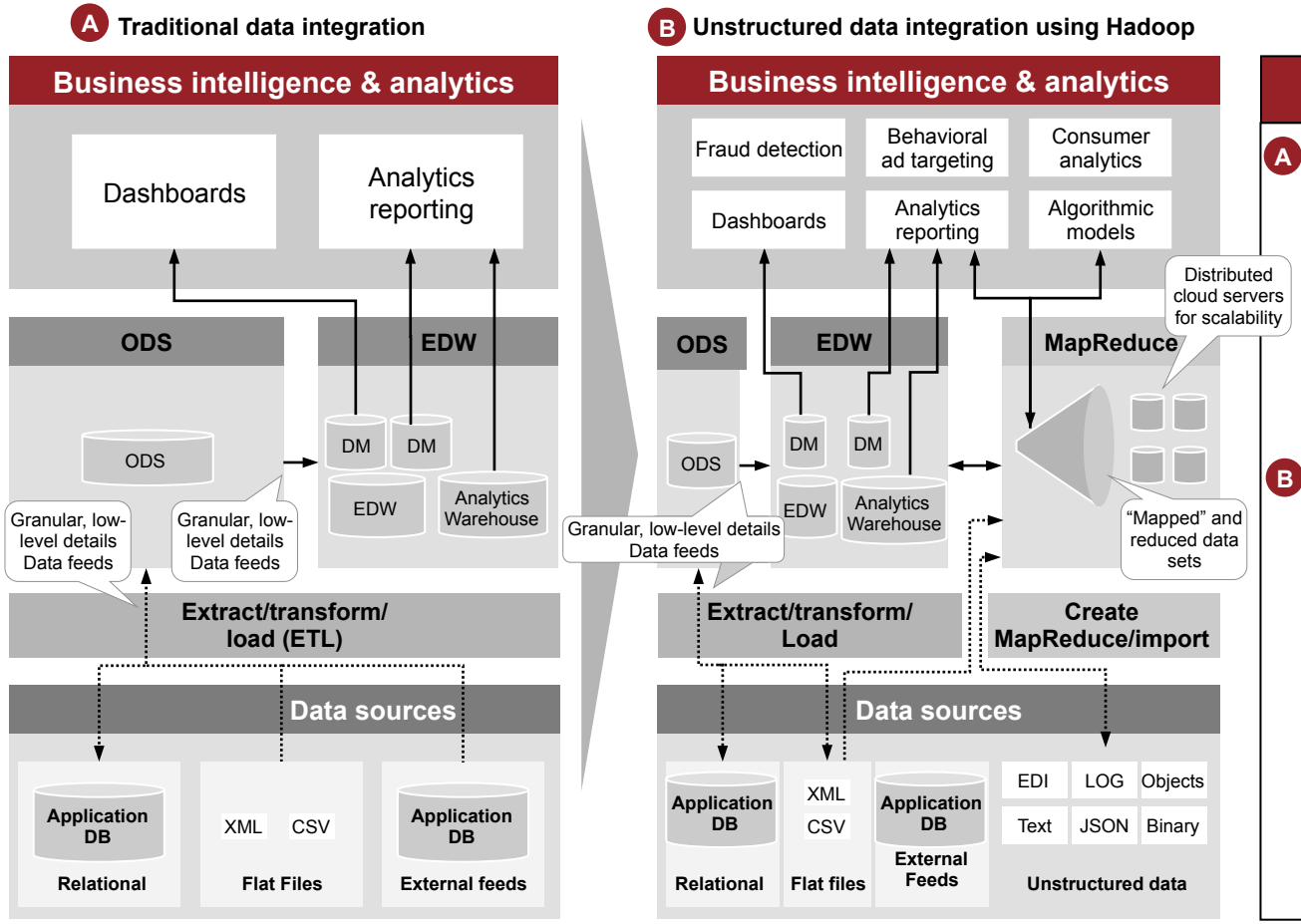
- Redesign and refine websites by optimizing site areas and page types, and rationalizing page tags to track interactions
- Upgrade infrastructure (e.g., Hadoop clusters, tag management systems) and processes to collect data from multiple streams including Web channel, social media, video, and smartphone apps
- Implement validation process and engines to ensure correct data capture
- Implement multiple Web tools (Google, Omniture, etc.) and enterprise analytics tools (SAS) to fill any gaps in data capture and enhance analytic capabilities

Integrating multi-stream data with mapreduce/hadoop



Source: Strategy& analysis

New technologies, such as MapReduce and Hadoop, can be utilized to quickly process large sets of unstructured data



Highlights

A Traditional technologies are optimized for processing structured data and presenting results for a narrow range of analytic applications. Substantial manipulations are required to process large volumes of unstructured data

B New distributed technologies such as MapReduce (developed by Google) and Hadoop (open-source Apache platform) are created for the purpose of processing large volumes of unstructured data and importing the results for use by a broad range of analytic applications

Source: Strategy& analysis

The MapReduce model does not replace traditional enterprise RDBMS; it tackles problems that could not be solved previously

Comparing RDBMS to MapReduce

	RDBMS	MapReduce/Hadoop
Data size	Gigabytes	Petabytes
Access	Interactive and batch	Batch
Structure	Fixed schema	Unstructured schema
Language	SQL	Procedural (Java, C++, Ruby, etc.)
Integrity	High	Low
Scaling	Nonlinear	Linear
Updates	Read and write	Write once, read many times
Latency	Low	High

How Hadoop complements RDBMS

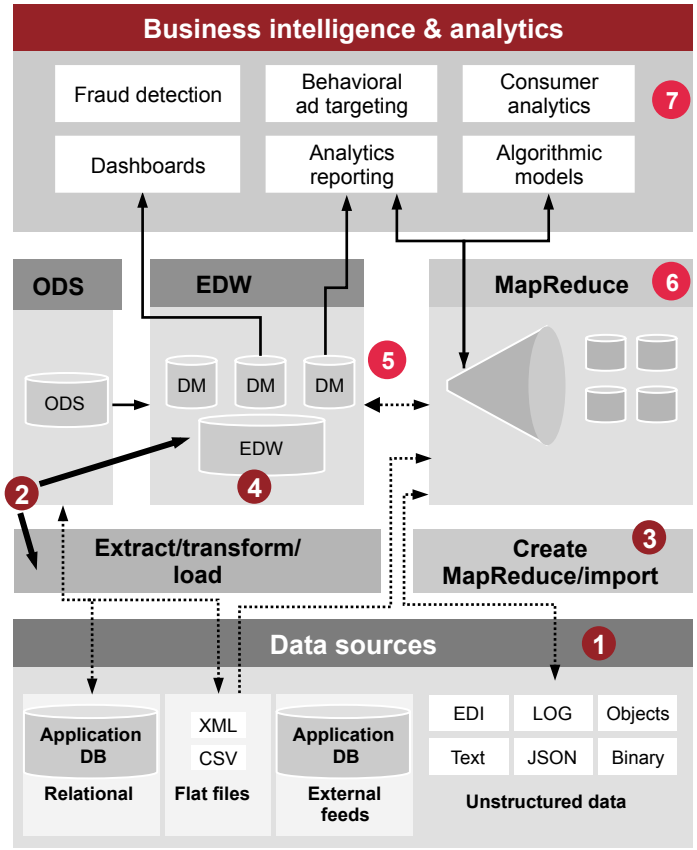
Highlights
• Storage of extremely high volumes of enterprise data
• Accelerating nightly batch business processes
• Improving the scalability of applications
• Creating automatic, redundant backups
• Producing just-in-time feeds for dashboards and business intelligence
• Use of Java for data processing instead of SQL
• Turning unstructured data into relational data
• Taking on tasks that require massive parallelism
• Moving existing algorithms, code, frameworks, and components to a highly distributed computing environment

MapReduce and Hadoop enable execution of analytics on the complete universe of data rather than on a sample set, as done traditionally in an RDBMS. This provides better analytic output for higher-quality decision making

Source: "10 Ways to Complement the Enterprise RDBMS Using Hadoop," by Dion Hinchcliffe

Successful implementation of MapReduce/Hadoop requires "heavy lifting" enhancements at every layer of data architecture

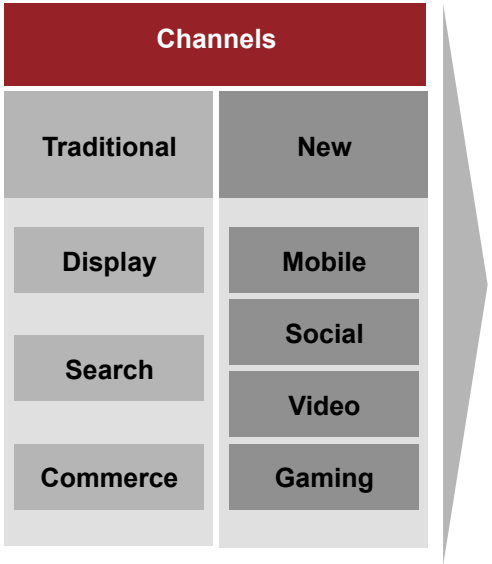
Areas of consideration in Hadoop Adoption



Source: Strategy& analysis

Considerations	
1	Data readiness of structured, external, and unstructured data sources needs to be assessed to determine which sources are suitable for the Hadoop platform and what new capabilities will be enabled or what existing capabilities will be better served
2	ETL jobs need to be reviewed and possibly rationalized, since some data is now sourced through Hadoop. Scheduling need to be rationalized while dependencies integrity is closely monitored and preserved
3	Define and implement the distributed file system while ensuring consistency of foreign keys, such as customer or product identifiers. On the infrastructure side, some nonfunctional requirements may be relaxed (e.g., backups and uptime do not need to be as strict as with conventional infrastructure)
4	Modify EDW schemas to accept data feeds from Hadoop
5	Define which data elements will be bi-directionally synchronized between EDW and Hadoop
6	MapReduce/Hadoop capabilities need to be joined with SQL so that MapReduce routines can be managed and optimized like other SQL queries. This will allow MapReduce programs to react differently depending on the data and parameters presented at run time, eliminating the need to create many versions of a program for different situations
7	Review and rationalize extract programs that shepherd data from the database to a series of downstream files, such as statistical analysis or data mining

Besides data technologies, other dimensions of the solution stack must also be considered

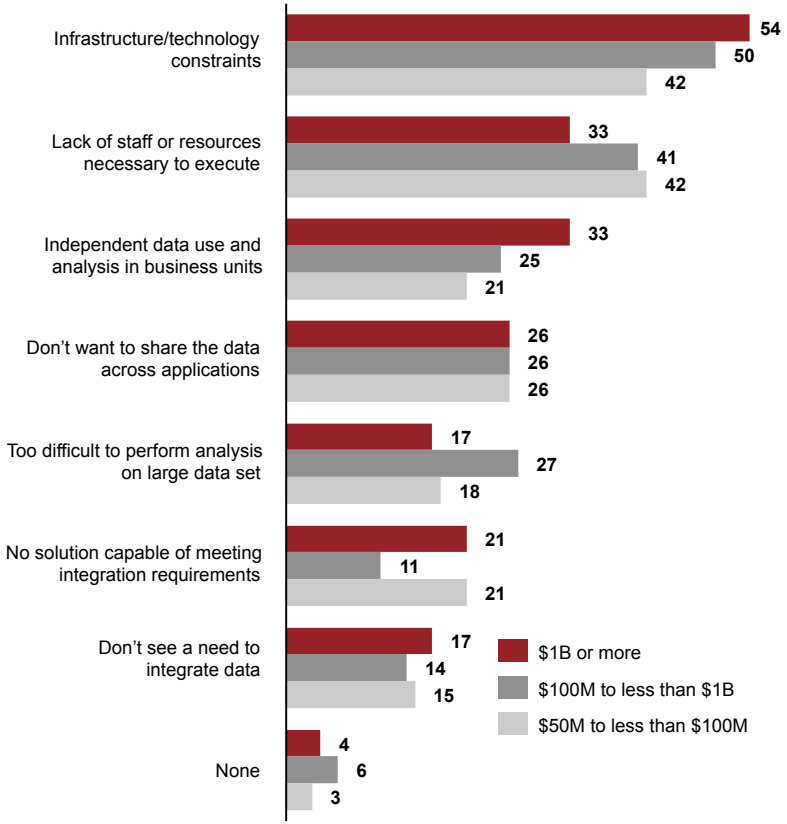


	Challenges
Navigation patterns	• Now that we better understand browsing patterns, how do we change navigation or usability? • Does the technology stack allow for appropriate and timely information updates?
Metadata tagging	• What additional content metadata can improve interpretation of clickstream data? • What metadata attributes are common across channels? • Where should those attributes be stored and maintained?
Real-time & near-real-time analytics	• What is the acceptable data lag? • Given the volume of transactions, is our architecture used in the optimal way and does it provide answers to the most important questions?
Content creation	• What new content needs to be created to address insights delivered by analytics? • How do we ensure consistency of content across channels? • How do we prevent it from becoming “stale”?
Governance & compliance	▪ In regulated industries, how can we shorten the content approval cycle while maintaining compliance? ▪ Do we have optimal workflows and effective governance?

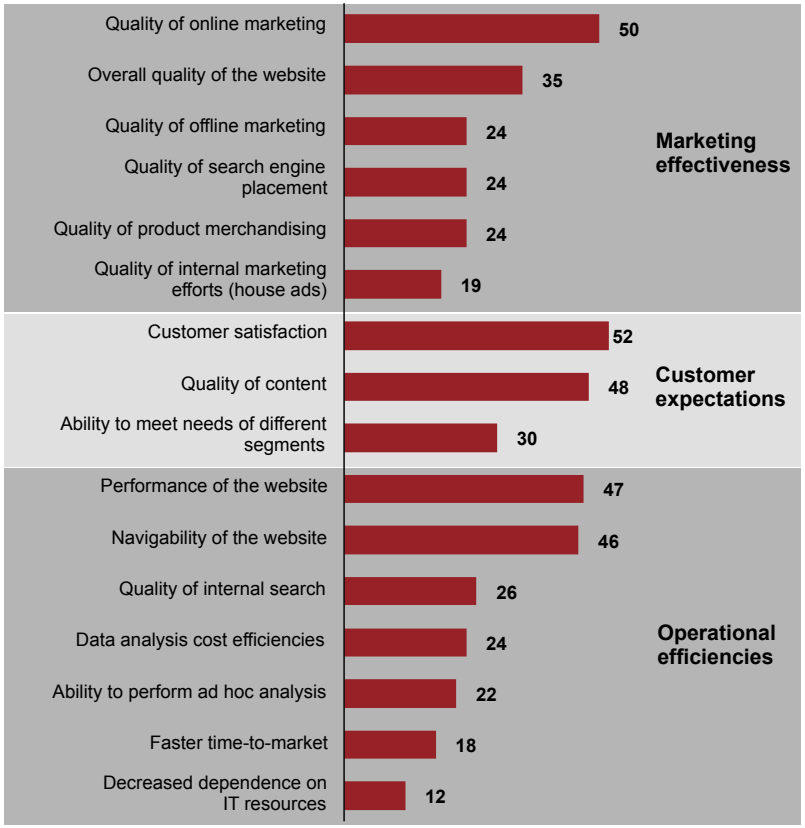
Source: Strategy& analysis

Integrating Web, social media, smartphone, and other unstructured data poses multiple challenges but provides significant benefits

Integrating Web data poses resource, infrastructure, and other challenges



Integrating Web data provides better insights and other benefits



Source: Forrester: Jupiter Research e-Rewards Executive Survey (2/08), n = 514 (small and medium-sized business decision makers, U.S.); Strategy& analysis

Strategy& is a global team of practical strategists committed to helping you seize essential advantage.

We do that by working alongside you to solve your toughest problems and helping you capture your greatest opportunities.

These are complex and high-stakes undertakings — often game-changing transformations. We bring 100 years of strategy consulting experience and the unrivaled industry and functional capabilities of the PwC network to the task. Whether you're charting your corporate strategy, transforming a function or business unit, or building critical capabilities, we'll help you create the value you're looking for with speed, confidence, and impact.

We are a member of the PwC network of firms in 157 countries with more than 184,000 people committed to delivering quality in assurance, tax, and advisory services. Tell us what matters to you and find out more by visiting us at strategyand.pwc.com.

This report was originally published by Booz & Company in 2011.

www.strategyand.pwc.com