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

<table>
<thead>
<tr>
<th>What is Web analytics?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition:</strong> 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)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Why is it important?</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ 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</td>
</tr>
<tr>
<td>▪ Web data integrated with other channels provides a better picture of the customer–business relationship and helps in identifying customer trends</td>
</tr>
<tr>
<td>▪ It is also useful in assessing the effectiveness of marketing campaigns and optimizing marketing spend</td>
</tr>
<tr>
<td>▪ And it improves the customer experience through faster service, thereby driving business growth and enhancing reputation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What are the challenges?</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Data volume growth is accelerating, making it cumbersome to capture and analyze Web data</td>
</tr>
<tr>
<td>▪ Unstructured social media data growth compounds the challenge, particularly as it must be integrated with enterprise structured data</td>
</tr>
<tr>
<td>▪ Multiple Web interaction platforms (PC, smartphone, tablet) further add to data capture and integration challenges</td>
</tr>
<tr>
<td>▪ Location and other smartphone sensor-based feeds also increase the complexity of continuous/real-time data capture</td>
</tr>
<tr>
<td>▪ There is no single tool available to capture and analyze all types of data</td>
</tr>
</tbody>
</table>
The number of Internet, social media, and mobile users tripled over the past decade, reaching a third of the world’s population.

**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*.

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

**Social Gaming**
(in US$ billion)

- 2011: 1.5
- 2012: 2.2
- 2013: 3.2
- 2014: 4.5
- 2015: 6.6

CAGR: +45%

**Players**
- Playfish
- Zynga
- Playdom

**Mobile Commerce**
(in US$ billion)

- 2011: 1.5
- 2012: 2.1
- 2013: 3.1
- 2014: 4.5
- 2015: 6.5
- 2016: 7.7

CAGR: +39%

**Players**
- Gilt Groupe
- eBay
- Amazon

**Social Commerce**
(in US$ billion)

- 2010: 0.1
- 2011: 1.0
- 2012: 3.0
- 2013: 5.0
- 2014: 8.0
- 2015: 12.0

CAGR: +161%

**Players**
- Groupon
- Yelp
- Living Social

Source: Booz & Company research and analysis
Customers have very high expectations for their end-to-end online experience

### Customer Expectations While Browsing Websites

<table>
<thead>
<tr>
<th>Land</th>
<th>Learn</th>
<th>Shop</th>
<th>Buy</th>
<th>Receive &amp; Use</th>
<th>Get Support &amp; Interact</th>
</tr>
</thead>
</table>
| Guide me to the site | Make it easy to find what I need  
– Customized search results based on previous purchasing activities | Let me configure my products  
– Guide my shopping decision with relevant advice  
– Give me complete visibility into availability, delivery time, and method | Allow me to use my preferred payment method  
– Give me flexible shipping options  
– Send me promotions and coupons  
– Customized based on my previous browsing and purchasing behavior | Tell me when my order will arrive  
– Allow me to modify or cancel my order  
– Make sure my order arrives on time  
– Let me download related software or apps from the site | Make it easy for me to find product support online  
– Direct me to appropriate articles or help desk agents based on the products I own  
– Allow me to connect with other customers and enthusiasts |
| Allow me to shop through multiple channels | Give me relevant content, let me look and compare  
– Tailored banner ads, promotions, and/or recommendations | Give me the right price  
– Provide sales support (e.g., click-to-chat)  
– Respond to my browsing behavior with targeted assistance | Let me go straight to checkout |  
| Remember who I am between visits | Share the “wisdom of the crowd” with me  
– Display what people “like me” have looked at | Recommend products I might be interested in  
– Behavioral targeting based on site activity |  
| Give me a personalized landing page |  
|  
| Source: Booz & Company 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

2. A customer logs onto his/her Amazon account

Customer Analytics

Through insights generated from users’ purchasing behaviors …

Based on previous purchase history …

… the prospect is provided with product recommendations for similar phones

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

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

- 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)

- Content personalization
  - Increased conversion, loyalty

- Search personalization
  - Increased conversion, loyalty

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

Source: Booz & Company 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

<table>
<thead>
<tr>
<th>Technical Capabilities Needed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Very limited (JavaScript tags on each webpage, familiarity with vendor software)</td>
<td></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Technical Capabilities Needed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited (JavaScript tags on each webpage, application programming interfaces)</td>
<td></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Technical Capabilities Needed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive (data warehouse integrated with CRM systems, data visualization software)</td>
<td></td>
</tr>
</tbody>
</table>

**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: Booz & Company analysis
Providing a rich experience requires a robust analytic capability, integrating disparate sources of structured and unstructured data

<table>
<thead>
<tr>
<th>Example Analytics</th>
<th>Data Used</th>
</tr>
</thead>
</table>
| **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: Booz & Company analysis
But the explosion of unstructured data volumes requires new approaches to data consolidation and analytics applications.

Structured and Unstructured Data Evolution

<table>
<thead>
<tr>
<th>Relational, Structured Data</th>
<th>Complex, Unstructured Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRM</td>
<td>Inventory</td>
</tr>
<tr>
<td>Financials</td>
<td>Sales records</td>
</tr>
<tr>
<td>Logistics</td>
<td>HR records</td>
</tr>
<tr>
<td>Data marts</td>
<td>Web profiles</td>
</tr>
<tr>
<td></td>
<td>Documents</td>
</tr>
<tr>
<td></td>
<td>Web feeds</td>
</tr>
<tr>
<td></td>
<td>System logs</td>
</tr>
<tr>
<td></td>
<td>Online forums</td>
</tr>
<tr>
<td></td>
<td>SharePoint</td>
</tr>
<tr>
<td></td>
<td>Sensor data</td>
</tr>
<tr>
<td></td>
<td>Audio</td>
</tr>
<tr>
<td></td>
<td>Images/video</td>
</tr>
</tbody>
</table>

Source: IDC white paper sponsored by EMC and Cloudera

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

### Leading Web Analytics and Industry Trends

<table>
<thead>
<tr>
<th><strong>Capturing &amp; Analyzing Multiple Streams of Data</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Social feed integration with Web and warehouse data for advanced customer analytics</td>
</tr>
<tr>
<td>- Task- or page-targets-based unobtrusive, short, highly actionable, quick feedback data supplementing site surveys</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Multiple Platforms for Customers to Interact On</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Added sources of data and complexity of integration from multiple platforms and form factors (smartphones, tablets)</td>
</tr>
<tr>
<td>- Complexity of integrating structured data with unstructured feeds from Web, social media, chat, and Internet-connected televisions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Increasing Volume of Data at a Faster Rate</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- New analytics, storage, and processing for accelerated integration at lower costs due to exponential growth of “big data” needs</td>
</tr>
<tr>
<td>- No single tool to capture and analyze massive Web data, requiring concurrent use of multiple analytic tools</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Continuous Data Streams</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Targeted offers based on customer location tracking enabled by GPS and cell-based tracking mechanisms in smartphones</td>
</tr>
<tr>
<td>- Interaction opportunities from tracking customer check-ins at vendor locations using new services</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Regulations</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Browser-based features for customer to opt out of tracking</td>
</tr>
<tr>
<td>- Upcoming regulations like FTC “Do Not Track” initiatives</td>
</tr>
<tr>
<td>- Android- and iOS-based app developers self-regulating and asking customer permission for data collection</td>
</tr>
</tbody>
</table>

### Vendors and Products

| **Source:** Booz & Company analysis |

| **Google Analytics, Adobe/Omniture, IBM/Coremetrics/Unica** |
| **ForeSee** |

| **Google Android, Apple iOS, RIM BlackBerry** |
| **Google TV, Boxee, Apple TV** |

| **Google MapReduce, Apache Hadoop, Google Caffeine** |
| **Google Analytics, Omniture, SAS** |

| **Google Latitude** |
| **Foursquare** |
| **Gowalla** |

| **Mozilla Firefox, Google Chrome, Opera** |
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**

- Structured Smartphone App Data: MapReduce/Hadoop
- Structured Web Form Data: Batch/Real-Time
- Unstructured Web Data: Batch/On-Demand
- Social Feeds: Structured Analytic Warehouse

**Analytic Modeling**

- Ad hoc queries
- Model execution
- Dashboard feeds
- Accelerate nightly batches
- Automatic redundant backups

Source: Booz & Company 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.

---

### Traditional Data Integration

<table>
<thead>
<tr>
<th>Business Intelligence &amp; Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dashboards</td>
</tr>
<tr>
<td>Analytics reporting</td>
</tr>
</tbody>
</table>

### Unstructured Data Integration Using Hadoop

<table>
<thead>
<tr>
<th>Business Intelligence &amp; Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraud detection</td>
</tr>
<tr>
<td>Behavioral ad targeting</td>
</tr>
<tr>
<td>Consumer analytics</td>
</tr>
<tr>
<td>Dashboards</td>
</tr>
<tr>
<td>Analytics reporting</td>
</tr>
<tr>
<td>Algorithmic models</td>
</tr>
</tbody>
</table>

---

**Data Sources**

- **Relational**
  - Application DB
  - External Feeds
- **Flat Files**
  - XML
  - CSV

---

**MapReduce**

- Distributed cloud servers for scalability
- "Mapped" and reduced data sets

---

**Create MapReduce/Import**

**Data Sources**

- **Relational**
  - Application DB
- **Flat Files**
  - External Feeds
- **Unstructured Data**
  - Text
  - JSON
  - Binary

---

Source: Booz & Company analysis
The MapReduce model does not replace traditional enterprise RDBMS; it tackles problems that could not be solved previously.

Comparing RDBMS to MapReduce

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

How Hadoop Complements RDBMS

- 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

Booz & Company
Successful implementation of MapReduce/Hadoop requires “heavy lifting” enhancements at every layer of data architecture

Areas of Consideration in Hadoop Adoption

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.

Source: Booz & Company analysis
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?
Integrating Web, social media, smartphone, and other unstructured data poses multiple challenges but provides significant benefits.
Contact Information

**Chicago**
Raj Parande  
Principal  
+1-312-578-4675  
rajendra.parande@booz.com

Yuri Goryunov  
Senior Associate  
+1-312-578-4791  
yuri.goryunov@booz.com

Balu Nair  
Senior Associate  
+1-312-578-4579  
balu.nair@booz.com

**Florham Park, NJ**
Ramesh Nair  
Partner  
+1-973-410-7673  
ramesh.nair@booz.com

Steffen Gnegel also contributed to this Leading Research.
Booz & Company is a leading global management consulting firm, helping the world’s top businesses, governments, and organizations. Our founder, Edwin Booz, defined the profession when he established the first management consulting firm in 1914.

Today, with more than 3,300 people in 60 offices around the world, we bring foresight and knowledge, deep functional expertise, and a practical approach to building capabilities and delivering real impact. We work closely with our clients to create and deliver essential advantage. The independent White Space report ranked Booz & Company #1 among consulting firms for “the best thought leadership” in 2011.

For our management magazine strategy+business, visit strategy-business.com.

Visit booz.com to learn more about Booz & Company.