CLICKSTREAM ANALYTICS TAKES APPLICATIONS TO THE NEXT LEVEL

RTInsights.com
Accelerate Your Business with Real-Time Insights

IBM

TABLE OF CONTENTS

Chapter 1:
The Fast-Paced Transformation of Clickstream Analytics

As web and smartphone experiences become ever-more crucial to the way we interact, businesses have been given an opportunity: They can leverage new technologies and smarter methods of securely and privately tracking consumers' behavior. By better understanding who, how, and why people are interacting with websites and mobile applications, businesses can discover how useful, powerful, and valuable these experiences are.

Clickstreams are leading the way in providing more profound insights, but, so far, only forward-thinking businesses are empowering themselves with the tools to take web analysis to the next level.

The Rise and Rapid Development of Web Analytics

Web analytics is nothing new: Omniture arrived on the scene in the late 1990s, and Google Analytics launched in 2005. Back then, web analytics consisted primarily of a long list of entries in a web server's log that detailed who requested to look at which pages and when. Rudimentary software made these server logs more readable, but they rarely took the next step—helping businesses understand what insights and actions they could glean from the numbers.

Modern web analysis is an entirely different landscape. JavaScript code can track a visitor's mouse movement, clicks, and more, providing more data to the business. Software offerings include machine learning (ML) and artificial intelligence (AI) capabilities to help enterprises gather insights, and even make predictions, based on the data they're receiving.

By combining more complex data streams on user events and interactions with these complex models, clickstreams can be directly correlated with human sentiment. This actionable insight can then be used to improve the experience.

What Is Clickstream Analysis?

A clickstream is the sequence of events, taps, and clicks that a user performs when visiting a website or using a mobile app.

Important clickstream events include:

- When and how the visitor arrived on the website or app
- The "route" the user took through the website or app
- How much time the visitor spent on each page/view
- Determining whether the user completed a given action, such as a purchase conversion
- If the visitor didn't successfully convert, uncovering when he or she left.

Clickstream analysis is the process of learning about how people interact with a website or app via these events.

The new capabilities of streaming analysis, including faster ingestion and processing than ever before, have given clickstreams a higher priority in businesses that have already garnered insights from former web analysis tools.

In fact, MarketsandMarkets expects the clickstream analysis market to grow from roughly $750 million in 2017 to $1.56 billion in 2022, a growth rate of 15.8 percent[1]. While concerns about consumer privacy and regulatory changes have the potential to reduce the growth rate of clickstreams, according to the report, the business need to benefit from more advanced analytic techniques and the massive adoption of e-commerce across nearly all verticals will far outweigh any concerns.

MarketsandMarkets expects the clickstream analysis market to grow from roughly $750 million in 2017 to $1.56 billion in 2022, a growth rate of 15.8 percent.

The Power of Clickstreams

A new Adobe Consumer Content Survey[2] shows that 47 percent of consumers will abandon a digital experience, such as a website or a mobile app, if they perceive that it takes too long to load. They crave improved digital experiences, but if they can't find what they need, they're likely to close the website or app and not return, moving on to a competitor. When those failed actions include paying a bill or getting customer support for a nagging issue, poor digital experiences can cost businesses lost revenue and customers, as well as negative reviews.

Businesses have found themselves between a rock and a hard place amidst this new consumer demand. They could install complex analytics software on their website, but that might only slow down their websites even more. On the other hand, if they fail to better understand where and why customers are abandoning, they'll get no closer to closing the loop on the aforementioned revenue-killing issues.

Clickstream analysis provides an opportunity for these businesses to discover where their customer experiences are failing, why they're failing, and the best route to fix them—all without slowing down the platform.

For example, let's say a business has a complicated six-step checkout process. With clickstreams, stakeholders in that business can watch customers arrive and flow through the checkout process. They can also view the entire journey prior to that point, such as whether customers came via an advertisement or a social media post. If they notice that a high percentage of customers are abandoning the checkout process on the fourth step, they can gather insight into how the experience could be improved.

Modern streaming platforms can scale these transactions up to hundreds of thousands of events per second. That means businesses are limited only by the complexity of the models they hope to run and the IT infrastructure that's necessary to develop new applications and take action based on the incoming insights.

---

[1] Clickstream Analytics Market worth 1,560.6 Million USD by 2022, MarketsandMarkets Research

[2] 2018 Adobe Consumer Content Survey

Why is clickstream analytics growing so quickly? It starts with the cross-vertical need for businesses to grow and improve their digital presences. Whether it's banking, retail, healthcare, e-commerce, customer support, or dozens of other industries, there is a shared desire to gain actionable insights from data at the same speeds that customers demand from their web and mobile experiences. With clickstream analytics, businesses can finally bring an omnichannel understanding of their customers into the real time. It's a faster, more responsive, and more flexible solution for everyone.

Chapter 2:
How Does Streaming Do True Real Time?

The future of clickstream analysis is complex, omnichannel, and real time. It's also going to be *aware, responsive,* and *instrumented*. Here's how businesses are going to get there.

To help tell the story, let's explore this data via the use case outlined in Chapter 1: capturing, ingesting, and analyzing clickstream events with IBM Streams from a customer who is working its way through a complex, multi-part checkout process (see Figure 1).

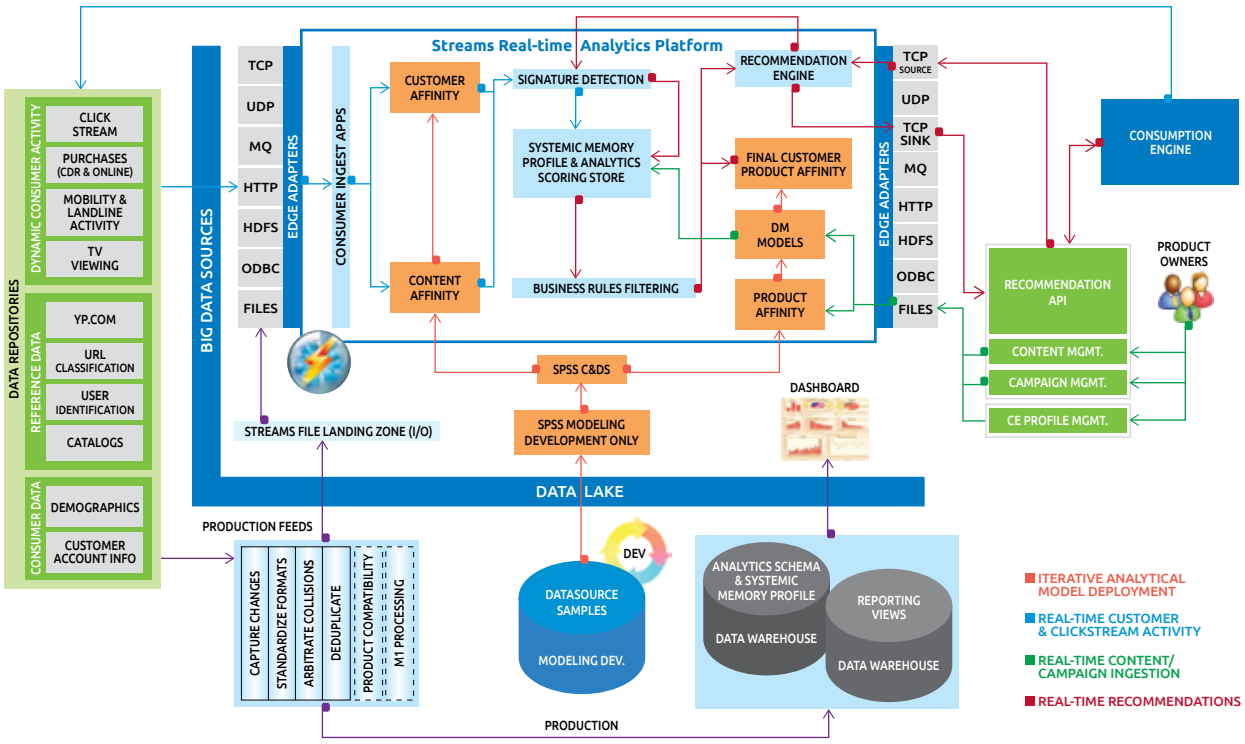

Figure 1: Analyzing Clickstream Events with IBM Streams and IBM SPSS
Source: IBM®

Clickstream data begins on the website, which acts as the **consumption engine** for the entire process. This clickstream data is just one of the many **data repositories** that can be connected to the Streams infrastructure as a whole; others can include past purchases, basic account information, browsing history, demographic details, and much more. Of vital importance is URL classification, which ensures the business knows the URL at which the most recent event occurred, such as the step in the cart process, along with the entire chain of URLs, that led the consumer to that point.

Between the data repositories and the IBM Streams infrastructure are the **source operators**, which are adapters that can read data from a variety of protocols, such as HTTP, TCP, Kafka, and more. Our clickstream has now passed through the data repository and into the **Streams runtime**.

Also heading into the runtime are **machine learning-based analytical models**. The company's developers are continually writing new models and applications to gather the right insights from prevailing trends based on information in their data lake. These new models are periodically pushed into the Streams runtime to help ensure models are accurately predicting current consumer behavior. These new models and analytical tools can include artificial intelligence and machine learning from IBM's Watson technology, which enables businesses to make predictions about future behaviors based on existing data and streaming data, all in real time.

Once inside the Streams runtime, our piece of clickstream data is passed through these analytical models, also known as **operators**. Each operator has a distinct purpose. For example, the first might log each visitor's current and past clicks on the website or app in question. Another would analyze the "routes" they've taken through the experience both now and in the past. A third might correlate all the previous information with the visitor's demographic information, and a fourth would determine if the user matches a particular journey known for abandonment.

The operator chain predicts that this user is likely to abandon the checkout process at step 5 based on his or her actions and the existing analysis about the cart's success rates on each step. Now, the business can take action. For instance, it could activate a popup that would allow the user to chat with a customer service representative to help guide him or her through the rest of the process, thus saving the conversion.

Calling streams "real time" is no mere moniker: All the ingestion, correlation, and analysis happens in 7 to 11 milliseconds, or *one-hundredth of a second*.

The Hardware That Backs Streaming Analysis Needs

At the speed streaming data needs to be processed, this complex sequence of events is performed on machines far more powerful than a typical desktop computer. At the core of any Streams implementation are three nodes, plus an instance of **Apache Zookeeper** to help maintain the application's state.

Management servers are responsible for tracking and maintaining all the processes that are occurring on the **application nodes**. Each application node runs an agent, which receives commands from the management servers, such as when to start or stop a particular process, while also sending back statistics about performance, CPU utilization, congestion, and more.

The application nodes are running many independent operators simultaneously, in what is called a **flow graph**. Each operator is positioned cyclically, meaning that it can communicate both upstream and downstream, giving the entire analytical chain the power to adapt chains of logic based on the results of parsed incoming clickstream data.

At the same time, management nodes are continuously sending the state information to the Zookeeper server, so that the entire system can adapt in the event of a node failure or other event that would negatively harm the system's performance. Because Streams is available as a service, businesses don't need to worry about maintaining **high availability** themselves, or dealing with the multiple hardware nodes that communicate amongst themselves and manage the complex network of connections.

Now that we understand how a clickstream system operates, we can begin to explore the many use cases. The power of real time is, simply put, too valuable for many industries to ignore.

Chapter 3:
From Streamlined Marketing to Targeted Advertisements and Much More

With a better understanding of clickstream analysis and how relatively small pieces of information can be converted into high-value insights, we can begin to outline the varied use cases for, and benefits of, collecting clickstream events.

The Past vs. the Present: Streamlined Marketing

With its varied use cases, clickstream analytics offers businesses a remarkable opportunity to dramatically simplify their marketing workflows.

PAST	PRESENT
Surveys collected customer feedback on a website or app experience.	Clickstreams deliver real, unbiased reports of how visitors navigated the experience, where they had difficulty navigating, and where they abandoned the process.
Data was disparate, incomplete, or statistically insignificant.	Clickstreams collect all visits and interactions, creating a large volume of data quickly, reliably, and without any manual intervention.
Website log files were difficult for non-developers to parse or understand, much less transform into actionable insights.	Visualizations help marketing teams understand exactly where experiences are struggling most and where they can create the most impact, streamlining their efforts and maximizing their impact on the company's bottom line.

The Past: Clickstream Data Without Streaming

Clickstreams of the past were slow and static: Server logs would sit in a database and would only be processed and understood long after the customer in question had left the website. Analysis was certainly not in real time, and perhaps not even streaming at all.

Using the previous example of optimizing cart conversion, the business would realize that customers weren't completing the cart only after the customers had left, leading to missed revenue opportunities.

Aside from missing out on instant insights, old clickstream data was not omnichannel in its access, which meant that it couldn't be associated and correlated with other data about the customer. Clickstream data used to be heavily siloed, and the tooling necessary to make stronger integrations didn't exist. Part of this absence of interconnectivity was due to the lack of powerful computing systems, but both businesses and analysis solutions providers, which took years to discover the potential in real-time clickstreams, played a role.

The Present: Instrumented, Aware, and Responsive Clickstreams

Every business will have different needs, and will each take advantage of various flow graphs with unique operators to their needs, but the theories are applicable across the board.

The Many Use Cases for Clickstream Analysis


Discovering at which point users have trouble getting the information they need via a website or mobile application


Optimizing e-commerce cart and checkout flow


A/B testing on marketing and design elements


Understanding which advertisements or call-to-action buttons are the highest performing


Delivering recommendations or targeted advertisements in real time


Predictions of users' behavior based on their past and current actions


Discovering new mediums by which consumers are finding a product or service


Identifying prevailing trends in customer behavior or product preference

Today, one IBM Streams client is running a real-time targeted advertising campaign using clickstream data. The campaign is based on the assumption that different consumers might want to receive different advertisements on the business's website based on their current needs, demographics, and more. Based on the aggregate of data—both clickstream events and an omnichannel view of the consumer—the system could determine, as an example, whether a Hyundai or BMW ad is more appropriate.

First and foremost, the client needs a definitive method of capturing and communicating clickstream events *(instrumented)* and additional omnichannel information to maintain accurate customer profiles. Clusters of machines, spread across two locations, provide high availability and disaster recovery, to ensure continuous service.

The client also needs a way to detect events of interest *(aware)*, for which Streams provides an under 10-millisecond response time—from ingesting the clickstream data, to joining it with the customer's profile, evaluating models, and returning a relevant advertisement. The system updates and saves tens of millions of customer profiles in-memory every day using data from over 200 sources.

Finally, the customer needs a way to act on the detected clickstream events *(responsive)*. With over 1,500 models running concurrently in the system to select relevant ads, the best ad is suggested to a consumption engine, which sends the suggested ad along with content the customer has requested.

The campaign has already had a successful five-year run. Perhaps most importantly, the system has experienced no unscheduled application or Streams downtime to become a reliable method of determining prevailing trends or changing demographics, and then responding by delivering more relevant content.

Clickthrough rates and revenue have increased 50 percent over the previous segmentation-based approach, which might have only used a single point of data when determining whether to show a BMW or Hyundai ad, as from our previous example. Residents of certain ZIP code might have a high average net worth, but that doesn't mean they're all in the market for a $50,000 car—clickstream analytics goes above and beyond segmentation by correlating demographics with far more detail on the customer, and in a fraction of a second.

With dozens of other practical applications (see "The Many Use Cases for Clickstream Analysis," page 9) that can be put into production use today, clickstream analysis has the potential to remake many of the ways that businesses manage their online presence. Delivering static content in real time has always been easy, but more powerful streaming data and real-time analytics unlocks new potential in dynamic, intelligent, and personalized customer experiences.

Delivering static content in real time has always been easy, but more powerful streaming data and real-time analytics unlocks new potential in dynamic, intelligent, and personalized customer experiences.

RTInsights is an independent, expert-driven web resource for senior business and IT enterprise professionals in vertical industries. We help our readers understand how they can transform their businesses to higher-value outcomes and new business models with IoT real-time analytics. We provide clarity and direction amid the often confusing array of approaches and vendor solutions. We provide our partners with a unique combination of services and deep domain expertise to improve their product marketing, lead generation, and thought leadership activity.

IBM Streams analyzes the broadest range of streaming data, making decisions while events are happening. It brings meaning to fast-moving data streams and helps organizations in a wide variety of industries. Now organizations can subscribe to device data to provide advanced analytics using Streams with IoT platform capabilities.

A key component of the IBM Cloud Platform, Streams offers a computing platform that helps organizations turn burgeoning, fast-moving volumes and varieties of data into insight. It delivers a programming language and an integrated development environment (IDE) for applications, a runtime system that executes the applications on a single host or a distributed set of hosts, and analytic toolkits to help speed development. You can use Java and Python to develop applications for deployment to the runtime. Streams can ingest, filter, analyze, and correlate massive volumes of continuous data streams.

These data streams can originate from any of the following:

- IoT devices and sensors.
- Text files, spreadsheets, images, video, and audio recordings.
- Email, chat, and instant messaging; web traffic, blogs, and social networking sites.
- Financial transactions, customer service records, telephone usage records, and system and application logs.
- Satellite data, GPS data, smart devices, sensors, network traffic, and messages.