



# Secure Web Analytics

*Understand your web visitors  
without web logs or page tags and  
keep all your data inside your  
firewall*

**Metronome Labs, LLC.**  
425 First Avenue  
Pittsburgh, PA 15219  
+1 (412) 434-4911  
[www.metronomelabs.com](http://www.metronomelabs.com)

---

# Secure web analytics

## Understanding your web visitors

Understanding your web visitors is critical to the effectiveness of your site. Never before has there been such a wealth of data on the way your visitors interact with your web site and react to your sales and marketing strategies.

Web analytics is becoming increasingly important for companies that sell or market through the web. In essence, web analytics packages are simply a set of pre-packaged reports. What differentiates them is the way they collect the data. Initially, data was obtained from server log files and this is still the most popular method. But log files do not give the whole story, so page tags have become popular, especially on larger, more sophisticated sites. They provide more information about your visitors but the data is often sent to a third party site which raises concerns about security and privacy. Because your web data is in a remote site it is difficult to correlate with your in-house sales and marketing databases.

But there is a better way. Metronome Capture traffic collection provides the richness of tag data with the security of log data inside your firewall. Metronome Explain analytics package extends the solution using WebAbacus to give you a complete view of your visitors.

## Web traffic overview

When a visitor goes to your web site, his browser sends an HTTP request packet. This is routed over the internet to your server which then replies with an HTML page carried by the HTTP protocol. On busy sites which have many servers (server farm), a load balancer routes the request to the least busy server. When the visitor's browser receives the HTML page, it loads it and reads all the links it contains. Every graphic on the page is a separate file which must be requested from your server farm. The initial request for the HTML page is typically called a "page view" and each request for an object such as a graphic is called a "hit."

Each page view usually results in about 5 to 20 hits. Each page request from a particular visitor may be directed by the load balancer to a different server so the pages for one visitor session may be served by many servers in your server farm.

---

## Logs and tags

### **Log files**

All web servers output log files in a standard format, although the actual content may differ slightly. They contain information about your visitor but the data is essentially about what the server is doing. If you use server log files to track your visitors, your analytics software has to gather the logs from all your servers, merge them together and then try to organize the page views and hits into visitor sessions. Server logs contain the hits for graphics which are usually uninteresting and so there is a huge amount of additional data that must be filtered out. All of this takes a lot of time and expensive computer power and storage. Typically, processing is performed each night so you have to wait a day to get your information. Some low end analytics packages do not even attempt to organize the page views into sessions so you cannot follow the path a visitor took. They can only provide overall statistics such as the number of requests for a particular page or the number of hits in a given period.

Web logs miss important data because servers do not see the underlying network protocol and they do not know when the page they sent actually got there. They don't know when it is complete with all its objects loaded ready to view. A web log does not show that a visitor clicked to a different page while the first page is on its way.

Outputting a web log slows your servers down and reduces your site capacity. If you can turn web logs off, you can save money on server hardware and software.

But one advantage of web logs is that the data they collect is secure inside your firewall and can be joined with your enterprise sales and marketing data to get a more complete view of your visitor.

### **Page tags**

Page tagging is now in vogue for larger sites. It works by placing a one pixel dummy graphic on a page. The visitor's browser will request this dummy graphic from a server. Typically, the page has a script embedded in it that will gather information about the visitor's machine and add it as parameters to this request. The request is usually directed to a third party managed site where the parameters are collected in this site's web server logs and then processed into a data warehouse. The data can then be viewed over the Internet through a portal.

Page tags are essentially visitor oriented and tell you much more about what your visitors are doing. Because the tags are operating from the visitor side, it is easier to relate the page views to visitor sessions and eliminate all the unwanted hits for graphic objects. There is less post-processing, so the data may be available sooner. In theory, you can even track the visitor's keystrokes and mouse-clicks. In practice, sites have many pages that are changing often and it is not practical or cost effective to maintain custom tags on every page.

Web logs need extensive filtering and processing to be useful.

They slow your servers down and do not really tell you what the visitor is experiencing.

A page tag is a link and code embedded in your page that sends data to a different server, usually at a third party vendor.

Page tags look a lot like spyware.

---

Managed services send your data to a remote site where it is difficult to correlate with your enterprise databases.

The solution is standard tags placed there automatically. This makes maintenance easier but reduces the quality of the data. In any case, a tagging solution requires that you make changes to your pages or the servers and be prepared to maintain them as your site changes.

Because page tags work from the visitor's browser, they can miss some important server events. For example, if a page has a server error, it never gets to the visitor and the page tags do not fire, so you get no data on this important event.

Then there are the security and privacy issues that have prevented many financial and government institutions from employing page tags. Most tagging solutions work by embedding scripts in the web pages which then send data about their actions back to a server. This looks a lot like spyware to security and privacy officers. The tag data is usually sent to a third party ASP site where it is warehoused with all the other clients' data. Sending potentially sensitive information off-site is often unacceptable.

Today, web sites are seen as one customer touch point in an integrated marketing and sales strategy. To get a complete view of your visitor, web data must be joined to data in your corporate, sales and marketing databases. But data from web tagging is collected in a remote 3<sup>rd</sup> party database and there is a vast amount of it. Your corporate data is too sensitive to send off site. To make the join, data must flow across the Internet. Do you send your sensitive data to the third party or do you download and store huge amounts of web data that you are paying someone else to manage?

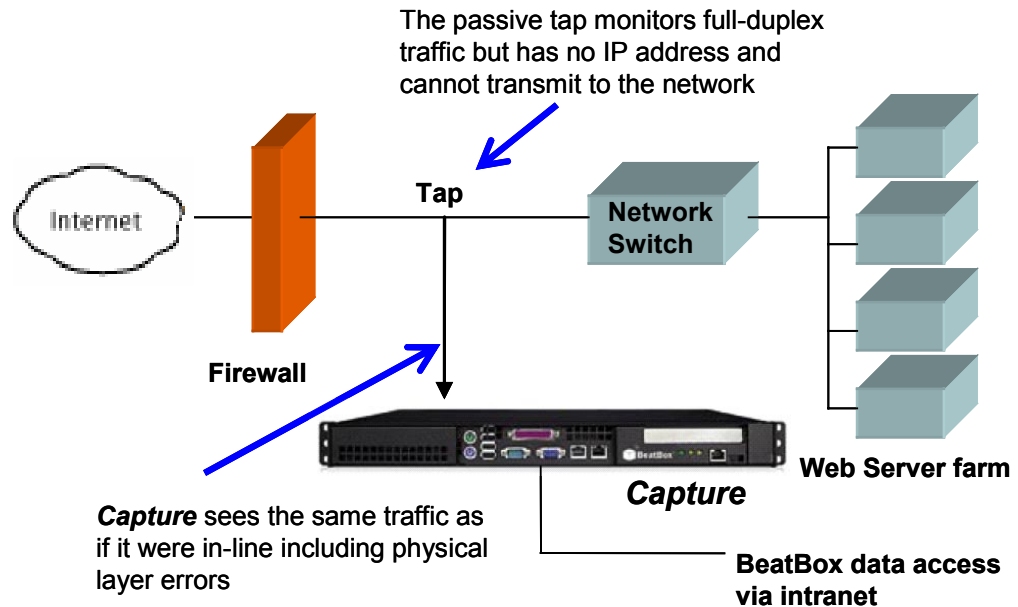
### **Metronome Capture – rich, secure and convenient data capture**

Metronome Capture collects the visitor and server activity, filters and sessionizes the data and keeps it all inside your firewall.

The Metronome Capture is placed inside your firewall before your load balancer. It passively listens to all the traffic to and from your site regardless of which server actually handles the request. It collects all your clickstream data at one central location, even when you have multiple servers and domains. It produces a single log file (or data stream) for your whole site with the data already filtered and organized into sessions.

Metronome Capture sees all of the traffic flowing between your visitors and web servers (including the IP packets) so it sees the acknowledgements to requests plus the low level errors that the server never sees. This enables Metronome Capture to calculate every detail of the transaction including precise load times for the HTML and each of its components.

Metronome Capture automatically groups page views and hits into sessions using a sophisticated algorithm and links them together with a unique session ID. The data is available as soon as the session completes, or sooner if you like.



### ***Filter and transform***

Web data is cleaned, filtered, transformed and organized into visitor sessions in one log file.

Metronome Capture has a sophisticated rule engine that is easily configured to give you the data you want and remove the data you don't. You can filter out hits you do not need based on any criteria. You can decide which fields you want and determine the format of your log. You can perform translations on the data and create custom fields to your specifications. For example, you could look for and extract a specific string from your cookie. You can categorize traffic by website, domain, etc. The rules are executed when the transaction occurs, so you get the results in exactly the format you want with no post processing required. Metronome Capture can even extract tags and data from the HTML pages, perform transformations and add them to the page view or hit logs.

Metronome data is ready to use without any post processing.

Cleaning data in real-time consumes less storage space, less computer power and makes the data immediately available.

### ***Data channels***

Channels enable you to deliver different views of the data to different user communities such as IT, Marketing, etc. The combination of rules and channels enables you to feed analytics, load databases and perform traffic analysis any way you want.

Data collected by Metronome Capture is sent to one or more channels. Channels allow you to filter, clean and aggregate data in different ways, and allow you to deliver the results to different locations. You could create a "session" log that contained one row per session with information that does not change over a session, another log for page views and maybe a third for specific hits you care

about. A channel typically sends the data to a log file, but it will also stream the data to an IP address for processing by another computer on your network.

Information collected and managed by Metronome Capture, including all custom reports and errors, can be requested as XML via the HTTP protocol.

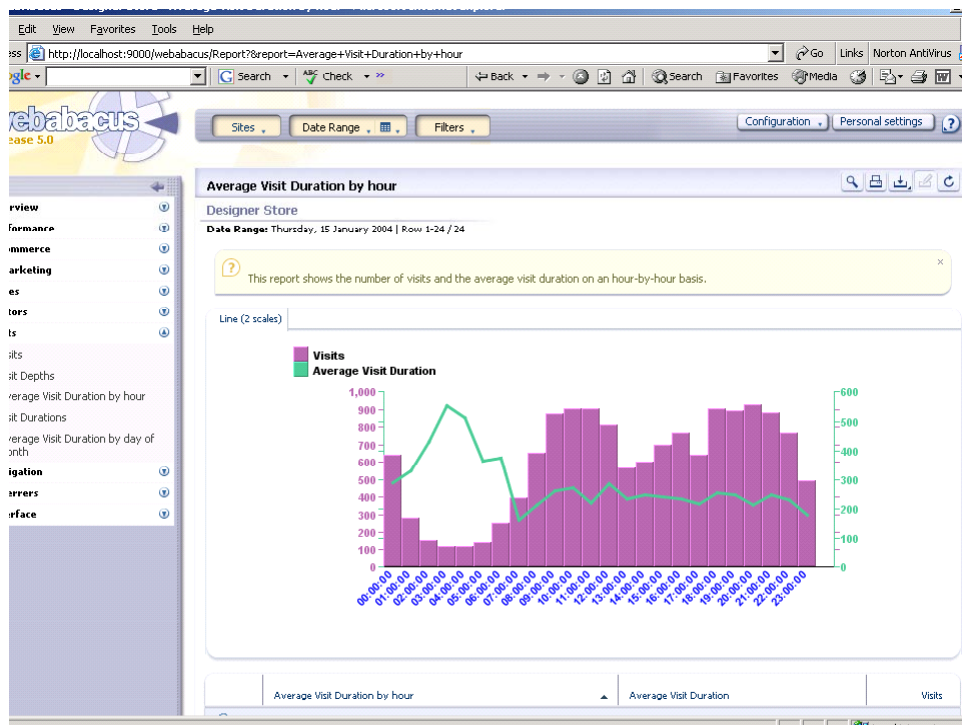
Metronome Capture has a Java-based database module that uses JDBC to store your information to most popular databases, including Oracle®, Microsoft SQL Server™, DB2®, Sybase® and MySQL™. This allows you to easily integrate data into your data warehouse and use standard reporting tools like Crystal Reports®.

If you are currently using web logs and want to keep your current analytics package, Metronome Capture can mimic the log format while creating just one pre-filtered log file. If you need great analytics, read about Metronome Explain with WebAbacus in the analytics section.

### Secure Web analytics – Metronome Explain uses WebAbacus analytics

Metronome Explain integrates WebAbacus™ analytics, a powerful and flexible analytical software package to give you new insights into site performance and visitor behavior. You get a complete view of your visitors from the Metronome data combined with your enterprise, sales and marketing data. Metronome Explain analytics provides a configurable dashboard for a quick view of your site. Metronome Explain includes extensive reports by visitor, visit, page views and hits with extensive drill down available at each level.

Using powerful analytics, tightly integrated Metronome Explain gives new insights into site performance and visitor behavior.



---

Metronome Explain loads the captured data into its datastore. You can integrate data in the Explain datastore with data from your own databases. Data can be imported via ODBC and most file formats. You can import the data into the datastore for continuous use or just access it at the time the report is generated.

You get the richness of Metronome data with analytics that enable you to view your traffic at the visitor, visit, page view and hit level without any changes to your site. You can adapt the analytics to your needs by configuring Metronome to capture additional data and build your own reports in Metronome Explain analytics.

### **More advantages**

#### ***Data encryption***

You can load your master encryption key file onto the Metronome platform. Since Metronome Capture is secure behind your firewall, there is no security risk. Metronome Capture collects the encrypted master secret when it is sent to your web server and decrypts it using the key. It can then decrypt the secure communications between your visitor and your web servers.

Metronome supports sites that use multiple RSA keys. There is also an SSL acceleration module available for sites with large amounts of encrypted traffic.

#### ***Triggering Events***

Metronome Capture supports three types of events (report, error and session events). An event is triggered whenever the channel it is associated with allows transaction data to pass through its filtering rules. All of the channel's data cleansing rules apply to any data used by the event. Events may also have their own data filtering and cleansing rules, allowing you to reuse a single channel for multiple events.

An error event allows you to define custom errors that can trigger SNMP traps and that can be tracked and managed within the Metronome web interface. All errors can also be stored to database tables or custom log files, requested as XML, and referenced via SNMP tables.

#### ***Beacons and event sequences***

The unique Metronome beacon feature enables you to detect and track sequences of important business events that occur within unique sessions. For example, you may want to track the particular sequence of placing an item in the shopping cart, viewing the shopping cart and then the item being removed. The sequences of events that have been detected so far for a particular session may be analyzed using the x-beacon data identifier. The complete sequence may be placed in a log variable and used to generate an event.

By allowing Metronome Capture to decode secure data, trigger on events, and track sequences of events in your site you will develop a powerful understanding of key events like purchase or shopping cart abandons.

Instantly know where your visitors are coming from.

Track network latency, detect suspicious visitors.

### **Geo location**

Metronome Capture uses an integrated database from Quova® to instantly pinpoint each visitor's physical location (country, state, city) and identify their connection information (ISP, network carrier and connection speed). You will know where your visitors are coming from. You can monitor connection latency from different cities to troubleshoot response problems. You can extend this with events for real-time fraud detection applications.

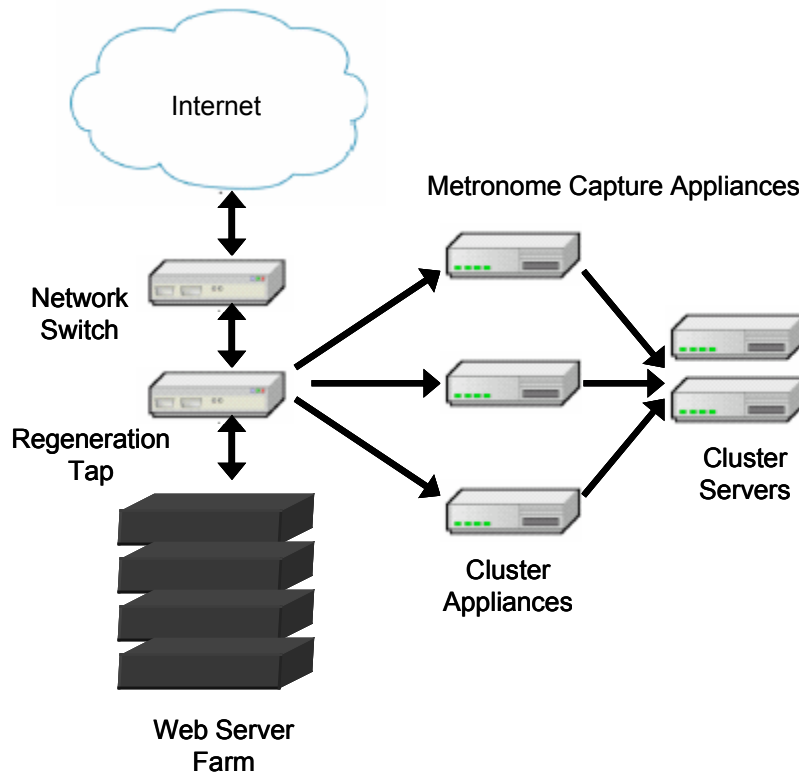
### **Metronome Web Console**

You can create real-time reports from any of the standard or custom fields that are being collected. Reports typically show aggregated data about what is happening on your web site now. This might be the number of visitors currently on-line, the number of visitors per hour over the last few hours, longest page load times, etc. The reports show up-to-the second information and can be refreshed every few seconds. They are viewed through your web browser.

Reports can also be triggered on events.

### **Clustering and failover**

On busy sites, the data collection Capture functionality can be logically split into the functions that handle the packet capture, reassembly and filtering (appliance) and the functions that handle channeling and events (server).



**Clustering**

---

Plug in Metronome Capture appliance and it immediately begins collecting relevant data without any changes to your server or site software.

Metronome Capture's dynamic clustering allows appliances to automatically share the load. When an appliance is added or removed the others dynamically adjust to share the load evenly. Metronome appliances sense when a member of the cluster fails and reconfigure themselves automatically. Metronome clusters are currently installed on some of the busiest retail sites. Regeneration taps can be used to send data simultaneously to multiple Metronome network appliances.

One cluster server can act as a warm backup, constantly monitoring the primary cluster server and taking over if it fails.

### ***Extensible***

Metronome Capture has an extensible Java layer that listens to an IP socket and receives data from a Capture channel. This layer can be loaded on a Metronome appliance or a different machine. By extending the Java classes, you can distribute channel data any way you want. Currently, there are standard plug-ins to load the data into a database and send alerts over email.

### ***Technical***

Metronome use high-quality passive network taps to promiscuously collect packets. Its multi-threaded architecture distributes work evenly across multiple processors, allowing a single appliance to scale to the full line speed of both copper and gigabit fiber networks. The appliance ports cannot transmit data and have no IP address so they cannot be interrogated.

A network tap is typically inserted between the load-balancing switch and an edge router. This tap maintains a hard-wired connection between the two devices so that the flow of traffic is not delayed and a failure of the tap (e.g. due to a power loss) will not cause a network outage. Since the tap prevents routing into the appliance, it does not introduce a security risk. Metronome also supports the use of spanning ports and repeating hubs.

Metronome Capture is a Linux application that is shipped in a dual processor server configuration. Metronome Capture and Metronome Explain can also be supplied as a software application to load on your own hardware.

---

Plug in  
Metronome  
Explain to  
understand  
what is  
happening on  
your web site  
and what your  
visitors are  
experiencing.

## **A Revolution in Web Analytics**

Web logs were never intended to be used in analytics, so it is very difficult and expensive to extract information from them. Building any but the most basic report from web logs takes hours or even days. Web logs also slow down your servers by as much as 20% and offer virtually no insight into a visitor's actual experience.

Embedding page tags into your web pages allows you to eliminate web logs and receive reports faster. Page tags raise privacy and security issues and have to be maintained on your site. They only work on pages that actually get loaded, not on the ones that break.

Metronome Capture eliminates web logs and page tags by analyzing and extracting information from your network traffic inside your firewall. There are no security or privacy concerns and little maintenance. Plug it in and you are up and running. Metronome Explain extends the solution to provide powerful and sophisticated reporting that gives you a complete view of your visitors.

## **About Metronome Labs**

Based in Pittsburgh, Metronome Labs LLC was formed by some of the management team of ClickCadence/BeatBox Technologies, the original creators of the BeatBox Capture appliance. Mercury Interactive acquired BeatBox Technologies LLC in late 2005. Mercury has licensed Metronome Labs to distribute BeatBox Capture as Metronome Capture and to incorporate the BeatBox technology in developing value-added products including solutions for web analytics, IT forensics and web data capture and loading. There are about 250 Capture appliances installed, including major retail sites like QVC. For more information, visit the web site at [www.metronomelabs.com](http://www.metronomelabs.com) .