Access Logging in Confluence

I have been looking at logging user / anonymous access in Confluence, and there are several ways to do this, but very few with "the full monty"

To compare, I have made a "list" after best effort on what we are getting/not getting from each method.

A few observations (read Filtering in Confluence Access Logging):

Monitoring Tools lige Nagios, Zenoss, Datadog etc can seriously fuck up the access logging, so be sure to have some form of either awereness or control over the hits from the monitorsystem, for example by:

- Frequency
- Source IP
- URL or a special Page that the monitoring hits

Robots and Spiders etc. can seriouly fuck up the access logging, give false results (or true).... But in general, these are not so interesting to access log.

See some tips at http://www.limov.com/library/do-not-believe-your-web-stats.lml

Logging options

- Logging options
- Apache/NGIX Access Log
- Tomcat Valve Logging
- View Tracker
- Confluence Event Logging
- Google Analytics (GA)
    - No valid Data Center license found
- Client Side Javascript

Apache/NGIX Access Log

If You have an Apache or NGIX (or similar) in front of Confluence, grapping the log from here is typically straight forward, as it more or less contains something like this sample:

```
62.145.36.18 - - [06/Feb/2017:15:14:11 +0100] "GET /display/ATLASSIAN/JIRA+as+CMDB HTTP/1.1" 200 18790
"https://www.google.nl/" "Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/55.
0.2883.87 Safari/537.36"
62.145.36.18 - - [06/Feb/2017:15:14:13 +0100] "GET /s/e052e137f250dc11172248580574573a-CDN/en_GB/6441
/c568f796f3f8ace564a3b6ddb68509c75e50e3a9/d542c7242aba64cb6167bf236f7afc02/_/download/contextbatch/css/_super
/batch.css?atlassian.aui.raphael.disabled=true HTTP/1.1" 200 90179 "http://www.mos-eisley.dk/display/ATLASSIAN
/JIRA+as+CMDB" "Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/55.0.2883.87
Safari/537.36"
62.145.36.18 - - [06/Feb/2017:15:14:13 +0100] "GET /s/en_GB/6441/c568f796f3f8ace564a3b6ddb68509c75e50e3a9/479/_
/styles/colors.css HTTP/1.1" 200 2923 "http://www.mos-eisley.dk/display/ATLASSIAN/JIRA+as+CMDB" "Mozilla/5.0
(Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/55.0.2883.87 Safari/537.36"
62.145.36.18 - - [06/Feb/2017:15:14:13 +0100] "GET /s/d41d8cd98f00b204e9800998ecf8427e-CDN/en_GB/6441
/c568f796f3f8ace564a3b6ddb68509c75e50e3a9/5.1.5/_/download/batch/com.refinedwiki.confluence.plugins.theme.
original:batch/com.refinedwiki.confluence.plugins.theme.original:batch.css HTTP/1.1" 200 7780 "http://www.mos-
eisley.dk/display/ATLASSIAN/JIRA+as+CMDB" "Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like
Gecko) Chrome/55.0.2883.87 Safari/537.36"
```

But what do we get from this logging?

What		Comment
Timestamp	✓	

Remote IP	✓	
Username	✗	Apache has no user context
Spacename	✗	
Pagename	✗	Apache has no app context. An URL is logged, but this can due to special chars in the Page Title be something like http://www.mos-eisley.dk/viewpage.action?id=1000 or a Tiny Link to a Confluence page, hence not the Pagename
URL	✓	
Return HTTP Code	✓	
Responseti me	✓	
UserAgent	✓	

Tomcat Valve Logging

As in the link: https://confluence.atlassian.com/confkb/how-to-enable-user-access-logging-182943.html

The configuration is very tricky and somewhat limited ragarding to filter mappings in log

Log sample:

```
2017-02-06 18:56:36,633 INFO [http-nio-8090-exec-24] [atlassian.confluence.util.AccessLogFilter] doFilter - GET
http://www.mos-eisley.dk/display/khvg145/2010/07/12/Papirer+fra+Vendia+er+ankommet 518258-23894 180 0:0:0:0:0:0:
0:1
2017-02-06 18:56:40,654 INFO [http-nio-8090-exec-18] [atlassian.confluence.util.AccessLogFilter] doFilter - GET
http://www.mos-eisley.dk/download/attachments/10027086/drengen%20i%20kufferten-christopher-1b-2011.wmv 476473-
548 164 0:0:0:0:0:0:0:1
2017-02-06 18:56:53,740 INFO [http-nio-8090-exec-24] [atlassian.confluence.util.AccessLogFilter] doFilter - GET
http://www.mos-eisley.dk/media/FamilieBilleder/Kaeledyr/Mikkel/thumbs/800pxHigh/DSC01632.JPG 354076 13 0:0:0:0:
0:0:0:1
2017-02-06 19:32:13,527 INFO [http-nio-8090-exec-23] [atlassian.confluence.util.AccessLogFilter] doFilter - GET
http://www.mos-eisley.dk/pages/viewpage.action 1077585-29559 362 0:0:0:0:0:0:0:1
```

What		Comment
Timestamp	✓	
Remote IP	✗	Not if there is a proxy in front, then the IP = 0:0:0:0:0:0:0:1
Username	✓	
Spacename	?	
Pagename	✗	An URL is logged, but this can due to special chars in the Page Title be something like http://www.mos-eisley.dk/viewpage.action?id=1000 or a Tiny Link to a Confluence page, hence not the Pagename. http://www.mos-eisley.dk/viewpage.action?id=1000 is just logged as http://www.mos-eisley.dk/pages/viewpage.action
URL	✓	
Return HTTP Code	✗	This only logs what is "200 OK" requests, not "404 Page not found" and other.
Responseti me	✓	
UserAgent	✗	

View Tracker

The Plugin: https://marketplace.atlassian.com/plugins/ch.bitvoodoo.confluence.plugins.viewtracker/server/overview

I have not tested this.

Confluence Event Logging

Is possible to use Apatavists Scriptrunner for Confluence to create an Event Handler that logs Page Access (View, Update, Delte etc etc) and Blogs and so on.

A small sample can be found on https://scriptrunner.adaptavist.com/latest/confluence/ConfluenceEventHandlers.html#_collecting_stats

My own working sample is on Logging PageEvents to Splunk

What		Comment
Timestamp	✓	
Remote IP	✗	
Username	✓	Extracted by the Script and hashed
Spacename	✓	Extracted by the Script
Pagename	✓	Extracted by the Script
URL	✓	
Return HTTP Code	✗	This only logs actual Page Events, so no return code is available.
Responsetime	✗	
UserAgent	✗	

Google Analytics (GA)

The overall best is Google Analytics, it gets most of the "good" stuff - the downside can be that using Google Analytics is not allowed in all organisations, as confidential data can be transmitted over the internet into Google Analytics.

Also, GA is Javascript based, so a lot of traffic will never show up on the GA Dashboard - so it for sure cant be used for any load-/overall logging issues.

You can use Google Analytics natively be inserting a small script in the "Custom HTML" on the admin pages:

Google Analytics script

```
</script>
<!--GOOGLE-ANALYTICS-PLUGIN-START-->
<!-- Updated: 2015-08-18 20:20:29 -->
<script type="text/javascript">
    // Disable any trackers that Atlassian may have added
    for (i=0; i < 50; i++) {  // normally it's 14, but let's use a hammer to kill any new ones too.
        window['ga-disable-UA-20272869-'+i] = true;
    }
    var _gaq = _gaq || [];
    _gaq.push(['af._setAccount', 'UA-XXXXXXXX-1']);

    AJS.toInit(function(){
        // set custom variables
        setCustomVarSpaceKey();
        setCustomVarUserKey();
        if (typeof AJS.params.pageId === 'string') {
            _gaq.push(
                    ['af._setCustomVar', 3, 'confluence-page-id', AJS.params.pageId, 3 ],
                    ['af._setCustomVar', 4, 'confluence-content-type', AJS.params.contentType, 3 ],
                    ['af._set', 'title', AJS.params.pageTitle]
            );
        }
        // track page view
        _gaq.push(['af._trackPageview']);

        AJS.$("a[href*='/download/attachments/']").click(function(event) {
            event.preventDefault();
            var pageIdAttachmentName = this.pathname.substring(this.pathname.indexOf('/download/attachments/')
+ 23);
            var parts = pageIdAttachmentName.split('/');
            // set custom variables
            setCustomVarSpaceKey();
            setCustomVarUserKey();
            _gaq.push(
                    ['af._setCustomVar', 3, 'confluence-page-id', parts[0], 3 ],
                    ['af._setCustomVar', 4, 'confluence-content-type', 'attachment', 3 ],
                    ["af._set", "title", decodeURIComponent(parts[1])],
                    ['af._setReferrerOverride', location.href]
            );
            var attachmentUrl = AJS.$(this).attr('href');

            // track page views
            _gaq.push(['af._trackPageview', attachmentUrl]);

            setTimeout(function() { document.location.href = attachmentUrl; }, 500);
        });

        (function() {
            var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
            ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-
analytics.com/ga.js';
            var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
        })();

    });

    function setCustomVarSpaceKey() {
        if (typeof AJS.params.spaceKey === 'string') {
            _gaq.push(['af._setCustomVar', 1, 'confluence-space-key', AJS.params.spaceKey, 3 ]);
        }
    }
    function setCustomVarUserKey() {
        if (typeof AJS.params.userKey === 'string') {
            _gaq.push(['af._setCustomVar', 2, 'confluence-user-key', AJS.params.userKey, 1 ]);
        }
    }

</script>
<!--GOOGLE-ANALYTICS-PLUGIN-END-->
```

No valid Data Center license found
Please go to **Atlassian Marketplace** to purchase or evaluate Refined Toolkit for Confluence Data Center.
Please read this **document** to get more information about the newly released Data Center version.

Or include the above with a plugin to embed Google Analytics pages in Confluence. Look at AppFusion and at Tracking Atlassian Confluence usage with Google Analytics

What		Comment
Timestamp	✓	
Remote IP	✓	
Username	✓	Extracted by the Script and hashed
Spacename	✓	Extracted by the Script
Pagename	✓	
URL	✓	
Return HTTP Code	✓	
Responsetime	✓	
UserAgent	✓	

Client Side Javascript

Make Your own Google Analytics clone; where a browser Javascript posts to a datasource.

With the Google Analytics scripts, make some changes to post to Your own data-backend; as the client side Google Analytics script can capture most data; and typically bots and crawler does not run the Javascript.