Server Logs

After Excel Fails
@ohgm

```
< Good Morning #BrightonSEO >
```

Prepare for walls of text.

About Me

- Former Senior Technical Consultant @ builtvisible.
- Now Freelance Technical SEO Consultant.
- @ohgm on Twitter.
- ohgm.co.uk for my webzone.

What I’d like to do today

1. Talk about access logs.
2. Show you some *command line tools*.
3. Show you some ways to apply these tools to common scenarios.
4. Sit back down.

This talk is on the first significant difficulty spike in server log analysis – *having too much information*.

Assumptions.

Assumptions

1. Your client is retaining their logs.
2. You don’t have access to your client’s server.

What is an Access.log?

```
ohgm.co.uk 108.162.246.241 - - [07/Apr/2016:01:08:07 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)" 108.162.246.241ohgm.co.uk 173.245.55.107 - - [07/Apr/2016:01:12:00 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5;
ET /feed/ HTTP/1.1" 200 136953 "-" "rogerbot/1.0 (http://www.moz.com/dp/rogerbot, rogerbot-crawler@moz.com)" 173.245.50.116ohgm.co.uk 141.101.105.124 - - [07/Apr/2016:01:39:35 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.105.12
insides/ HTTP/1.1" 200 6048 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.105.147ohgm.co.uk 173.245.49.104 - - [07/Apr/2016:01:39:45 +0100] "GET /?p=674 HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 173.245.49.1
; Linux x86_64; rv:43.0) Gecko/20100101 Firefox/43.0" 162.158.147.53www.ohgm.co.uk 173.245.52.188 - - [07/Apr/2016:01:46:27 +0100] "GET / HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; NetcraftSurveyAgent/1.0; +info@netcraft.com)" 173.245.52.188ohgm.co.uk 173.245.52.101 - - [07/Apr/2016:01:46:34 +0100] "GET / HTTP/
rove-performance/?share=twitter HTTP/1.1" 302 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.104.245ohgm.co.uk 173.245.49.24 - - [07/Apr/2016:02:09:42 +0100] "GET /faster-google-penalty-removal/?replytocom=161835 HTTP/1.1" 200 8939 "-" "Mozilla/5.0 (compatible; MJ12
edly.com/fetcher.html; like FeedFetcher-Google)" 199.27.133.72ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:15 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:16 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 T
ordpress-from-android/ HTTP/1.1" 200 6970 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:22 +0100] "GET /ahrefs-coms-link-profile/ HTTP/1.1" 200 7291 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:22 +0100] "GET /gorilla-warfare/ HTTP/1.1" 200 5056 "-" "-" 141.101.99.
8125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 141.101.99.109 - - [07/Apr/2016:02:23:27 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.99.109ohgm.co.uk 141.101.99.12 - - [07/Apr/2016:02:23
.uk/bot.php?+)" 162.158.93.119ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:26 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 162.158.93.119 - - [07/Apr/2016:02:41:27 +0100] "GET /on-not-blogging-for-ages/ HTTP/1.1" 200 7879 "-" "Mozilla/5.0 (compatible; MJ
s-obeyed/ HTTP/1.1" 200 7808 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:30 +0100] "GET /ecommerce-linkbuilding/ HTTP/1.1" 200 9311 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:28 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.10
-" "-" 141.101.99.116ohgm.co.uk 162.158.93.119 - - [07/Apr/2016:02:41:29 +0100] "GET /?p=309 HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 162.158.93.119ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:24 +0100] "POST /wp-cron.php?doing_wp_cron=1459993284.691
.uk/bot.php?+)" 162.158.95.143ohgm.co.uk 108.162.221.178 - - [07/Apr/2016:02:44:59 +0100] "GET /about/ HTTP/1.1" 200 5216 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 108.162.221.178ohgm.co.uk 162.158.95.143 - - [07/Apr/2016:02:44:59 +0100] "GET /2-tips-to-make-your-life-marginally-better/?sh
ke-your-life-marginally-better/?share=reddit HTTP/1.1" 302 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 162.158.95.143ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:45:18 +0100] "POST /wp-cron.php?doing_wp_cron=1459993518.2541680335998535156250 HTTP/1.1" 200 20 "-" "WordPr
wp_cron=1459993766.6555008888244628906250 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 162.158.255.188 - - [07/Apr/2016:02:58:57 +0100] "GET /know-when-a-canonical-is-obeyed/ HTTP/1.1" 200 7808 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 162.158.255.188o
7 +0100] "GET /preventing-tiered-link-spam/ HTTP/1.1" 200 8639 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:59:07 +0100] "GET /diminishing-returns-relauthor/ HTTP/1.1" 200 8556 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:59:06 +0100] "GET /page-sitemap.xml HTTP/1.1" 20
7/Apr/2016:03:05:59 +0100] "GET /about/ HTTP/1.1" 200 5216 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 108.162.221.178ohgm.co.uk 108.162.221.178 - - [07/Apr/2016:03:17:53 +0100] "GET /about/ HTTP/1.1" 200 5216 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 108.162.221.17
4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:03:19:44 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:03:19:39 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 Total Cache/0.9.4.1" 141
5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 162.158.51.191www.ohgm.co.uk 108.162.223.239 - - [07/Apr/2016:03:26:14 +0100] "GET / HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)" 108.162.223.239www.ohgm.co.uk 108.162.223.239 - - [07/Apr/2016:03:26:
7.0; rv:11.0) like Gecko" 108.162.212.237ohgm.co.uk 108.162.219.171 - - [07/Apr/2016:03:30:07 +0100] "GET /feed/ HTTP/1.1" 200 136953 "-" "Flamingo_SearchEngine (+http://www.flamingosearch.com/bot)" 108.162.219.171ohgm.co.uk 108.162.246.235 - - [07/Apr/2016:03:43:21 +0100] "GET /feed HTTP/1.1" 304 - "http://ohgm.co
din-stalking-your-own-employees-so-you-can-have-awkward-conversations-about-their-profile-updates/ HTTP/1.1" 200 7544 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:03:43:27 +0100] "GET / HTTP/1.1" 200 6812 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:03:43:31 +0100] "GET /fast
ks-penalties/ HTTP/1.1" 200 8665 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:03:43:31 +0100] "GET /filter-server-logs-to-googlebot/ HTTP/1.1" 200 10596 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:03:43:22 +0100] "POST /wp-cron.php?doing_wp_cron=1459997002.698383092880249023
rackback/ HTTP/1.1" 200 85 "http://ohgm.co.uk/about/" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 162.158.26.124ohgm.co.uk 141.101.105.134 - - [07/Apr/2016:03:54:39 +0100] "GET /know-when-a-canonical-is-obeyed/feed/ HTTP/1.1" 200 2048 "-" "Mozilla/5.0 (compatible; AhrefsBot/5.1; +http://ahrefs.c
1" 200 8867 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:01:18 +0100] "GET /installing-applications-to-sd-card-windows-10/ HTTP/1.1" 200 7442 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:01:18 +0100] "GET /bulk-inspect-http-response-headers/ HTTP/1.1" 200 8209 "-" "-" 1
h-file-aliasing/ HTTP/1.1" 200 8584 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:01:18 +0100] "GET /wmt-crawl-representative-url-transfer-link-equity/ HTTP/1.1" 200 7976 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:01:18 +0100] "GET /broken-link-destruction-for-better-r
:06 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:20:07 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:20:09 +0100] "GET /polls/ HTTP/
99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:20:09 +0100] "GET /speeding-default-wordpress/ HTTP/1.1" 200 10483 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:20:09 +0100] "GET /pdf-to-html-and-seo/ HTTP/1.1" 200 10153 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:2
folder-depth/amp/ HTTP/1.1" 200 3127 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.177ohgm.co.uk 108.162.216.171 - - [07/Apr/2016:04:37:52 +0100] "GET
.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:38:00 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:38:01 +0100] "GET /intentions-and-permissibility/ HTTP/1.1" 200 37869 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [
41.101.99.116 - - [07/Apr/2016:04:38:01 +0100] "GET /not-falling-stillborn-from-the-press/ HTTP/1.1" 200 7570 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:38:01 +0100] "GET /gorilla-warfare/ HTTP/1.1" 200 5056 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:04:37:54 +0100] "P
etcher-Google)" 199.27.133.72ohgm.co.uk 173.245.52.101 - - [07/Apr/2016:04:55:14 +0100] "GET /robots.txt HTTP/1.1" 304 - "https://jad.subarctic.org/status/ohgm.co.uk" "awooo" 173.245.52.101ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:04:08 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.
116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:04:13 +0100] "GET /identifying-widget-template-and-embed-spam/ HTTP/1.1" 200 7391 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:04:06 +0100] "POST /wp-cron.php?doing_wp_cron=1460001846.8121600151062011718750 HTTP/1.1" 200 20 "-" "WordPress/4.4.2
" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:04:13 +0100] "GET /link-analysis-filter-non-indexed-domains/ HTTP/1.1" 200 9398 "-" "-" 141.101.99.116ohgm.co.uk 108.162.246.234 - - [07/Apr/2016:05:04:04 +0100] "GET / HTTP/1.1" 200 6812 "http://ohgm.co.uk/" "Mozilla/5.0 (Windows NT 5.1; rv:31.0) Ge
"Mozilla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.139ohgm.co.uk 199.27.133.72 - - [07/Apr/2016:05:22:15 +0100] "GET /feed/ HTTP/1.1" 304 - "-" "Feedly/1.0 (+h
ap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:22:21 +0100] "GET /about/ HTTP/1.1" 200 5215 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:22:16 +0100] "POST /wp-cron.php?doing_wp_cron=1460002936.4769539833068847656250 HTTP/1.1"
222.176 - - [07/Apr/2016:05:24:42 +0100] "GET /?p=1201 HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 162.158.222.176ohgm.co.uk 162.158.222.176 - - [07/Apr/2016:05:24:45 +0100] "GET /speeding-default-wordpress/ HTTP/1.1" 200 10483 "-" "Mozilla/5.0 (compatible;
43 - - [07/Apr/2016:05:33:25 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "python-requests/2.2.1 CPython/2.7.3 Linux/3.2.0-56-generic" 108.162.221.143ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:33:29 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 173.245.50.1
-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:33:33 +0100] "GET / HTTP/1.1" 200 6812 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:33:27 +0100] "POST /wp-cron.php?doing_wp_cron=1460003607.7350389957427978515625 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.
gs-subdomains-cctlds/ HTTP/1.1" 200 7684 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:33:33 +0100] "GET /fun-unnatural-outbound-links-penalties/ HTTP/1.1" 200 8665 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:33:33 +0100] "GET /filter-server-logs-to-googlebot/ HTTP/1.1"
ke FeedFetcher-Google)" 199.27.133.72ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:56:24 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:56:26 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 Total Cache/0.9.4.1" 141.1
/page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:56:28 +0100] "GET /preserve-link-equity-with-file-aliasing/ HTTP/1.1" 200 8584 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:05:56:28 +0100] "GET /could-rotating-gifs-impro
ankings/ HTTP/1.1" 200 10310 "-" "-" 141.101.99.116ohgm.co.uk 108.162.222.228 - - [07/Apr/2016:06:05:57 +0100] "GET /polls/ HTTP/1.1" 200 6872 "-" "Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)" 108.162.222.228ohgm.co.uk 108.162.221.178 - - [07/Apr/2016:06:09:48 +0100] "GET /ab
+http://go.mail.ru/help/robots)" 162.158.92.78ohgm.co.uk 173.245.49.109 - - [07/Apr/2016:06:22:31 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 173.245.49.109ohgm.co.uk 173.245.49.124 - - [07/Apr/2016:06:22:34 +0100] "GET /?p=309 HTTP/1
ap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:22:45 +0100] "GET /sift-grep-on-steroids/ HTTP/1.1" 200 7938 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:22:45 +0100] "GET /speeding-default-wordpress-part-2-images/ HTTP/1.1" 20
ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:22:45 +0100] "GET /pdf-to-html-and-seo/ HTTP/1.1" 200 10153 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:22:45 +0100] "GET /speeding-default-wordpress/ HTTP/1.1" 200 10483 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:22:45 +
r/2016:06:30:31 +0100] "POST /wp-cron.php?doing_wp_cron=1460007031.7777969837188720703125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 108.162.221.178 - - [07/Apr/2016:06:30:28 +0100] "GET /about/ HTTP/1.1" 200 5216 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like G
-so-you-can-have-awkward-conversations-about-their-profile-updates/recaptcha/api/challenge?k=6LfOYgoTAAAAAInWDVTLSc8Yibqp-c9DaLimzNGM HTTP/1.1" 404 3938 "http://ohgm.co.uk/recaptcha/api/challenge?k=6LfOYgoTAAAAAInWDVTLSc8Yibqp-c9DaLimzNGM" "Mozilla/5.0 (Windows NT 5.2; WOW64; rv:38.0) Gecko/20100101 Firefox/38.0" 1
.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:44:36 +0100] "GET /cadbury-eggs-with-alternate-color-insides/ HTTP/1.1" 200 6048 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:44:36 +0100] "GET /posting-into-wordpress-from-android/ HTTP/1.1" 200 6970 "-" "-" 141.101.99.116ohgm.co.uk 14
nk-profile/ HTTP/1.1" 200 7291 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:06:44:36 +0100] "GET /using-multilinks-for-link-prospecting/ HTTP/1.1" 200 8145 "-" "-" 141.101.99.116ohgm.co.uk 199.27.128.188 - - [07/Apr/2016:06:44:43 +0100] "GET /automate-linkedin-stalking-your-own-employees-so-you-
- [07/Apr/2016:06:58:12 +0100] "POST /wp-cron.php?doing_wp_cron=1460008692.7255558967590332031250 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 108.162.221.178 - - [07/Apr/2016:07:09:17 +0100] "GET /about/ HTTP/1.1" 200 5216 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0
016:07:09:29 +0100] "GET /truly-exceptional-content/ HTTP/1.1" 200 7455 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:09:29 +0100] "GET /identifying-widget-template-and-embed-spam/ HTTP/1.1" 200 7391 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:09:29 +0100] "GET /how-acc
ins/ HTTP/1.1" 200 9398 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:09:26 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:09:29 +0100] "GET /identify-urls-sitemap-arent-indexed/ HTTP/1.1" 200 8976 "-" "-
0 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 173.245.49.130 - - [07/Apr/2016:07:16:25 +0100] "POST /about/trackback/ HTTP/1.1" 200 85 "http://ohgm.co.uk/about/" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 173.245.49.130ohgm.co.uk 199.27.133.72 - - [07/Apr/2016:07:28:31
9.116 - - [07/Apr/2016:07:28:37 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:28:38 +0100] "GET /preventing-tiered-link-spam/ HTTP/1.1" 200 8639 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:28:38 +0100]
W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:37:55 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:37:57 +0100] "GET /faster-google-penalty-removal/ HTTP/1.1" 200 8899 "-" "-" 141.101.99.
" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:37:56 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:37:57 +0100] "GET /block-googlebot-crawl-folder-depth/ HTTP/1.1" 200 9004 "-" "-" 141.101.99.116ohgm.co.uk
1.1" 200 7131 "-" "Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)" 108.162.222.228ohgm.co.uk 199.27.133.129 - - [07/Apr/2016:07:42:52 +0100] "GET /feed/ HTTP/1.1" 304 - "-" "Digg Feed Fetcher 1.0 (Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_1) AppleWebKit/534.48.3 (KHTML, like G
uk 173.245.49.109 - - [07/Apr/2016:07:46:11 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 173.245.49.109ohgm.co.uk 173.245.49.24 - - [07/Apr/2016:07:46:12 +0100] "GET /ecommerce-linkbuilding/?share=linkedin HTTP/1.1" 302 20 "-" "Mozilla
bot.php?+)" 173.245.49.24ohgm.co.uk 173.245.49.24 - - [07/Apr/2016:07:46:36 +0100] "GET /learning-to-type-faster/?share=email HTTP/1.1" 302 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 173.245.49.24ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:46:37 +0100] "POST /wp-cron.
tible; Yahoo! Slurp; http://help.yahoo.com/help/us/ysearch/slurp)" 108.162.246.243ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:57:50 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:57:51 +0100] "GET /post-sitemap.xml HTTP/1.1
-rankings/ HTTP/1.1" 200 10310 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:57:48 +0100] "POST /wp-cron.php?doing_wp_cron=1460012268.2747070789337158203125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:07:57:55 +0100] "GET /
+0100] "GET /preserve-link-equity-with-file-aliasing/ HTTP/1.1" 200 8584 "-" "-" 141.101.99.116ohgm.co.uk 173.245.52.166 - - [07/Apr/2016:07:57:46 +0100] "GET /pdf-to-html-and-seo/?replytocom=168286 HTTP/1.1" 200 10199 "-" "Mozilla/5.0 (compatible; linkdexbot/2.0; +http://www.linkdex.com/bots/)" 173.245.52.166ohgm.
2.0; +http://www.linkdex.com/bots/)" 173.245.52.166ohgm.co.uk 173.245.52.166 - - [07/Apr/2016:08:02:40 +0100] "GET /seo-2/feed/ HTTP/1.1" 200 136971 "-" "Mozilla/5.0 (compatible; linkdexbot/2.0; +http://www.linkdex.com/bots/)" 173.245.52.166ohgm.co.uk 108.162.219.103 - - [07/Apr/2016:08:03:05 +0100] "GET /pdf-to-ht
com/bots)" 108.162.238.139ohgm.co.uk 141.101.104.247 - - [07/Apr/2016:08:04:51 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)" 141.101.104.247ohgm.co.uk 141.101.104.247 - - [07/Apr/2016:08:04:55 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; Ya
01.99.116 - - [07/Apr/2016:08:14:03 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:14:04 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:0
wordpress-part-2-images/ HTTP/1.1" 200 12650 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:14:06 +0100] "GET /codeacademy-code-year-by-christmas/ HTTP/1.1" 200 8271 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:14:00 +0100] "POST /wp-cron.php?doing_wp_cron=1460013240.8727
la/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)" 108.162.222.228ohgm.co.uk 108.162.222.228 - - [07/Apr/2016:08:26:00 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)" 108.162.222.228ohgm.co.uk 199.27.133.72 - - [
1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:29:40 +0100] "GET /using-twitter-to-find-guest-post-opportunities/ HTTP/1.1" 200 10664 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:29:40 +0100] "GET /cadbury-eggs-with-alternate-col
the-press/ HTTP/1.1" 200 7570 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:29:40 +0100] "GET /posting-into-wordpress-from-android/ HTTP/1.1" 200 6970 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:29:40 +0100] "GET /backlink-analysis-just-got-harder/ HTTP/1.1" 200 8195 "-
om/)" 108.162.219.175ohgm.co.uk 108.162.219.175 - - [07/Apr/2016:08:32:43 +0100] "GET /link-audits-rank-cracker/?share=linkedin HTTP/1.1" 302 20 "-" "ltx71 - (http://ltx71.com/)" 108.162.219.175ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:32:45 +0100] "POST /wp-cron.php?doing_wp_cron=1460014365.9527349472045898437
2016:08:34:43 +0100] "GET /wp-json/ HTTP/1.1" 200 318 "-" "ltx71 - (http://ltx71.com/)" 108.162.219.175ohgm.co.uk 173.245.49.135 - - [07/Apr/2016:08:34:44 +0100] "GET /wp-content/plugins/jetpack/css/jetpack.css?ver=3.9.4 HTTP/1.1" 200 10030 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW6
7 - - [07/Apr/2016:08:34:45 +0100] "GET /wp-content/plugins/jetpack/modules/sharedaddy/sharing.js?ver=3.9.4 HTTP/1.1" 200 4046 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safari/537.36" 173.245.49.137ohgm.co.uk 173.245.49.
1.png HTTP/1.1" 200 564231 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safari/537.36" 108.162.229.208ohgm.co.uk 173.245.49.94 - - [07/Apr/2016:08:34:45 +0100] "GET /wp-content/uploads/2015/10/srcsetnexus-1024x350.png HTTP/
ecko) Chrome/49.0.2623.110 Safari/537.36" 173.245.49.144ohgm.co.uk 173.245.49.84 - - [07/Apr/2016:08:34:46 +0100] "GET /srcset-on-wordpress/?relatedposts=1 HTTP/1.1" 200 814 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safa
200 7391 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:36 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:37 +0100] "GET /page-anchors-content-marketing/ HTTP/1.1" 200 8134 "-" "-" 141.101.99.116ohgm
141.101.99.116 - - [07/Apr/2016:08:47:37 +0100] "GET /ultimate-guide-seo-2014/ HTTP/1.1" 200 6640 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:31 +0100] "POST /wp-cron.php?doing_wp_cron=1460015251.2599670886993408203125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.1
y/2.2w; +https://www.qwant.com/)/*" 173.245.49.109ohgm.co.uk 173.245.49.112 - - [07/Apr/2016:08:56:25 +0100] "GET /wp-json/oembed/1.0/embed?url=http://ohgm.co.uk/speeding-default-wordpress-part-2-images/&format=xml HTTP/1.1" 200 1240 "-" "Mozilla/5.0 (compatible; Qwantify/2.2w; +https://www.qwant.com/)/*" 173.245.4
```

```
ohgm.co.uk 108.162.246.241 - - [07/Apr/2016:01:08:07 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)" 108.162.246.241ohgm.co.uk 173.245.55.107 - - [07/Apr/2016:01:12:00 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5;
ET /feed/ HTTP/1.1" 200 136953 "-" "rogerbot/1.0 (http://www.moz.com/dp/rogerbot, rogerbot-crawler@moz.com)" 173.245.50.116ohgm.co.uk 141.101.105.124 - - [07/Apr/2016:01:39:35 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.105.12
insides/ HTTP/1.1" 200 6048 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.105.147ohgm.co.uk 173.245.49.104 - - [07/Apr/2016:01:39:45 +0100] "GET /?p=674 HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 173.245.49.1
; Linux x86_64; rv:43.0) Gecko/20100101 Firefox/43.0" 162.158.147.53www.ohgm.co.uk 173.245.52.188 - - [07/Apr/2016:01:46:27 +0100] "GET / HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; NetcraftSurveyAgent/1.0; +info@netcraft.com)" 173.245.52.188ohgm.co.uk 173.245.52.101 - - [07/Apr/2016:01:46:34 +0100] "GET / HTTP/
rove-performance/?share=twitter HTTP/1.1" 302 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.104.245ohgm.co.uk 173.245.49.24 - - [07/Apr/2016:02:09:42 +0100] "GET /faster-google-penalty-removal/?replytocom=161835 HTTP/1.1" 200 8939 "-" "Mozilla/5.0 (compatible; MJ12
edly.com/fetcher.html; like FeedFetcher-Google)" 199.27.133.72ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:15 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:16 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 T
ordpress-from-android/ HTTP/1.1" 200 6970 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:22 +0100] "GET /ahrefs-coms-link-profile/ HTTP/1.1" 200 7291 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:22 +0100] "GET /gorilla-warfare/ HTTP/1.1" 200 5056 "-" "-" 141.101.99.
```

```
ohgm.co.uk 162.158.93.95 - - [11/Apr/2016:10:14:20 +0100] "GET /wmt-crawl-representative-url-transfer-link-equity/ HTTP/1.1" 200 7976 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 162.158.93.95

ohgm.co.uk 108.162.219.171 - - [11/Apr/2016:10:15:07 +0100] "GET /feed/ HTTP/1.1" 200 136953 "-" "Flamingo_SearchEngine (+http://www.flamingosearch.com/bot)" 108.162.219.171

ohgm.co.uk 108.162.219.176 - - [11/Apr/2016:10:22:54 +0100] "GET /wayback-machine-seo HTTP/1.1" 200 9079 "http://www.traackr.com/" "Traackr.com" 108.162.219.176

ohgm.co.uk 173.245.55.114 - - [11/Apr/2016:10:23:35 +0100] "GET /author/ohgm/ HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.114

ohgm.co.uk 173.245.55.123 - - [11/Apr/2016:10:23:42 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.123
```

```
2016:08:34:43 +0100] "GET /wp-json/ HTTP/1.1" 200 318 "-" "ltx71 - (http://ltx71.com/)" 108.162.219.175ohgm.co.uk 173.245.49.135 - - [07/Apr/2016:08:34:44 +0100] "GET /wp-content/plugins/jetpack/css/jetpack.css?ver=3.9.4 HTTP/1.1" 200 10030 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW6
7 - - [07/Apr/2016:08:34:45 +0100] "GET /wp-content/plugins/jetpack/modules/sharedaddy/sharing.js?ver=3.9.4 HTTP/1.1" 200 4046 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safari/537.36" 173.245.49.137ohgm.co.uk 173.245.49.
1.png HTTP/1.1" 200 564231 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safari/537.36" 108.162.229.208ohgm.co.uk 173.245.49.94 - - [07/Apr/2016:08:34:45 +0100] "GET /wp-content/uploads/2015/10/srcsetnexus-1024x350.png HTTP/
ecko) Chrome/49.0.2623.110 Safari/537.36" 173.245.49.144ohgm.co.uk 173.245.49.84 - - [07/Apr/2016:08:34:46 +0100] "GET /srcset-on-wordpress/?relatedposts=1 HTTP/1.1" 200 814 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safa
200 7391 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:36 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:37 +0100] "GET /page-anchors-content-marketing/ HTTP/1.1" 200 8134 "-" "-" 141.101.99.116ohgm
141.101.99.116 - - [07/Apr/2016:08:47:37 +0100] "GET /ultimate-guide-seo-2014/ HTTP/1.1" 200 6640 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:31 +0100] "POST /wp-cron.php?doing_wp_cron=1460015251.2599670886993408203125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.1
y/2.2w; +https://www.qwant.com/)/*" 173.245.49.109ohgm.co.uk 173.245.49.112 - - [07/Apr/2016:08:56:25 +0100] "GET /wp-json/oembed/1.0/embed?url=http://ohgm.co.uk/speeding-default-wordpress-part-2-images/&format=xml HTTP/1.1" 200 1240 "-" "Mozilla/5.0 (compatible; Qwantify/2.2w; +https://www.qwant.com/)/*" 173.245.4
```

ohgm.co.uk[1] 173.245.55.123[2] - - [11/Apr/2016:10:23:42 +0100[3]] "GET[4] /please[5] HTTP/1.1[6]" 200[7] 6812[8] "-[9]" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)[10]" 173.245.55.123[11]

1. The host responding to the request.
2. The IP that serviced the request.
3. The date and time of the request.
4. The HTTP method: GET, POST, PUT, HEAD, or DELETE.
5. The resource requested.
6. The HTTP Version {HTTP/1.0|HTTP/1.1|HTTP/2}
7. The server response.
8. The download size.
9. The referring URL.
10. The reported User-Agent.
11. The IP that made the request.

Configurations vary substantially.

Why SEOs Like Them.

There is a lack of overlap between server logs and crawl simulation tools.

Access logs show what’s being accessed rather than what’s simply accessible.

We find correlation between crawl efficiency improvements and organic performance. Access logs are one of the best tools for identifying crawl waste.

Why ‘Excel Fails’?

Microsoft Excel currently supports ***1,048,576*** rows of data.

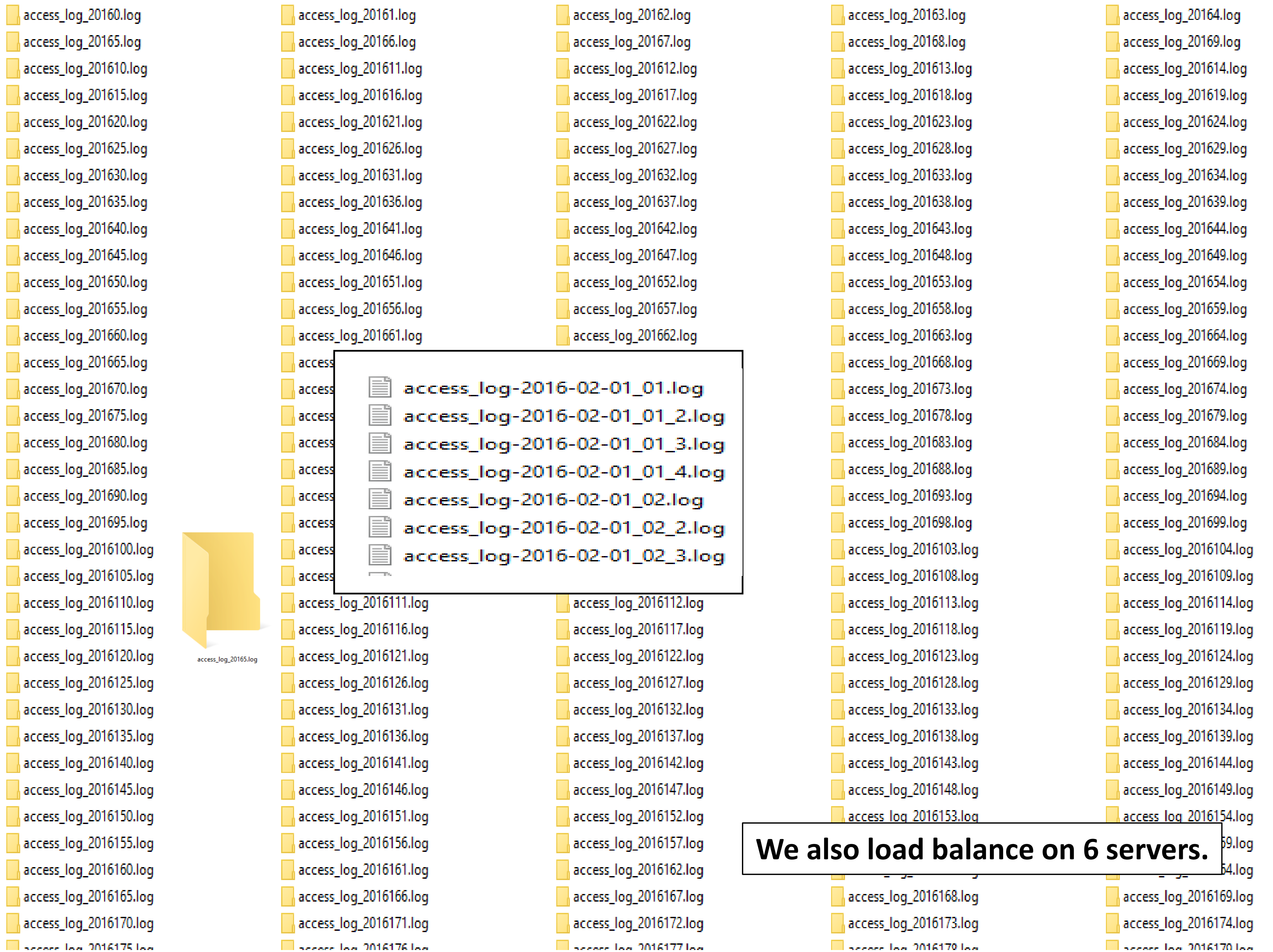
There are no plans to increase this.

Agency Scenario

Your manager has sold a Server Log Analysis project, requesting 1 month of access logs from the client, a UK high street retailer.

You receive 15 *access_log.gz* files, totalling 17.6GB. **Excel *won't* open** any of them. You don't know it yet, but they are unfiltered.

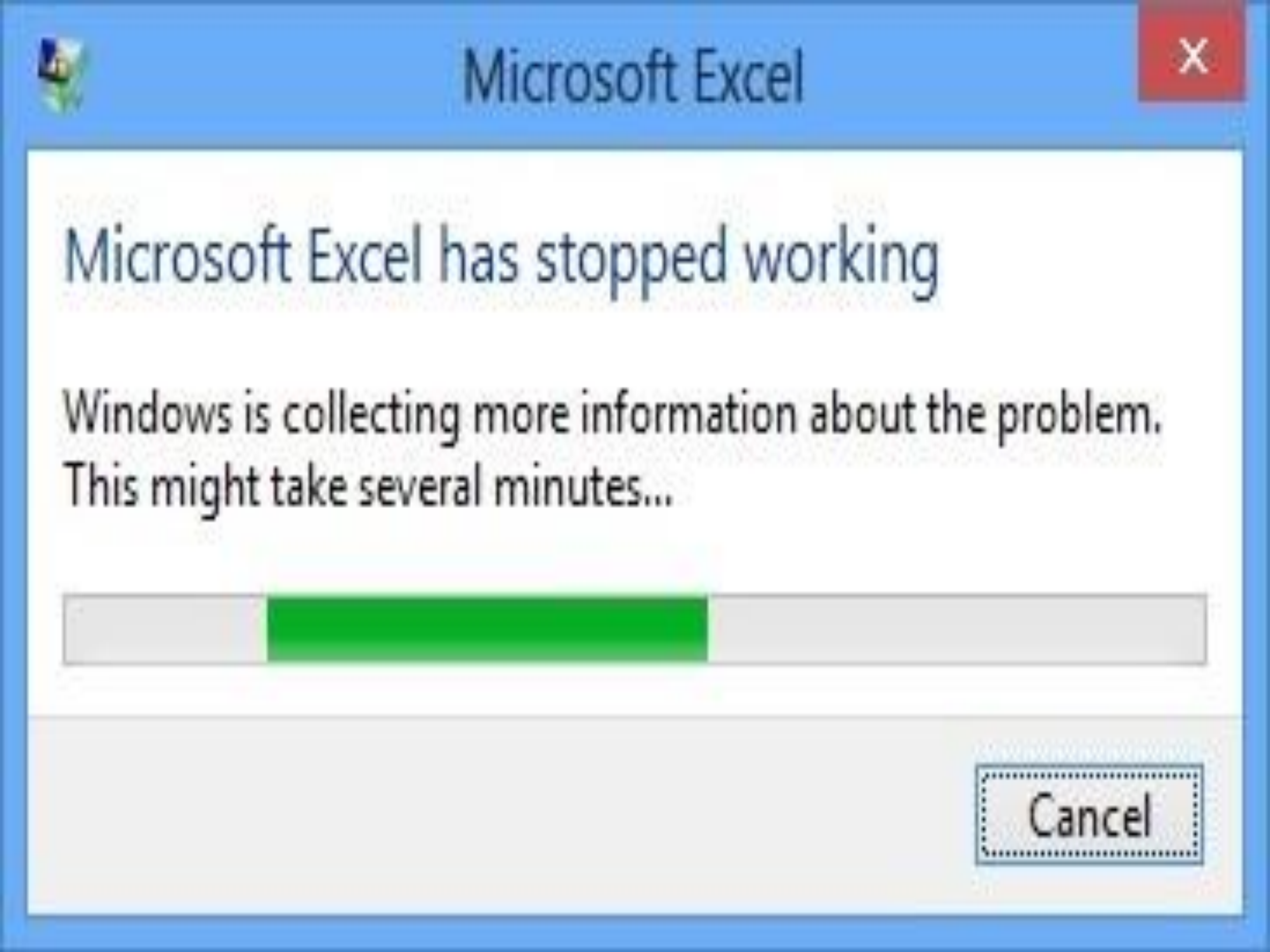
Good Luck.



```
access_log_20160.log    access_log_20161.log    access_log_20162.log    access_log_20163.log    access_log_20164.log
access_log_20165.log    access_log_20166.log    access_log_20167.log    access_log_20168.log    access_log_20169.log
access_log_201610.log   access_log_201611.log   access_log_201612.log   access_log_201613.log   access_log_201614.log
access_log_201615.log   access_log_201616.log   access_log_201617.log   access_log_201618.log   access_log_201619.log
access_log_201620.log   access_log_201621.log   access_log_201622.log   access_log_201623.log   access_log_201624.log
access_log_201625.log   access_log_201626.log   access_log_201627.log   access_log_201628.log   access_log_201629.log
access_log_201630.log   access_log_201631.log   access_log_201632.log   access_log_201633.log   access_log_201634.log
access_log_201635.log   access_log_201636.log   access_log_201637.log   access_log_201638.log   access_log_201639.log
access_log_201640.log   access_log_201641.log   access_log_201642.log   access_log_201643.log   access_log_201644.log
access_log_201645.log   access_log_201646.log   access_log_201647.log   access_log_201648.log   access_log_201649.log
access_log_201650.log   access_log_201651.log   access_log_201652.log   access_log_201653.log   access_log_201654.log
access_log_201655.log   access_log_201656.log   access_log_201657.log   access_log_201658.log   access_log_201659.log
access_log_201660.log   access_log_201661.log   access_log_201662.log   access_log_201663.log   access_log_201664.log
access_log_201665.log   access                                          access_log_201668.log   access_log_201669.log
access_log_201670.log   access                                          access_log_201673.log   access_log_201674.log
access_log_201675.log   access                                          access_log_201678.log   access_log_201679.log
access_log_201680.log   access                                          access_log_201683.log   access_log_201684.log
access_log_201685.log   access                                          access_log_201688.log   access_log_201689.log
access_log_201690.log   access                                          access_log_201693.log   access_log_201694.log
access_log_201695.log   access                                          access_log_201698.log   access_log_201699.log
access_log_2016100.log  access                                          access_log_2016103.log  access_log_2016104.log
access_log_2016105.log  access                                          access_log_2016108.log  access_log_2016109.log
access_log_2016110.log  access_log_2016111.log  access_log_2016112.log  access_log_2016113.log  access_log_2016114.log
access_log_2016115.log  access_log_2016116.log  access_log_2016117.log  access_log_2016118.log  access_log_2016119.log
access_log_2016120.log  access_log_2016121.log  access_log_2016122.log  access_log_2016123.log  access_log_2016124.log
access_log_2016125.log  access_log_2016126.log  access_log_2016127.log  access_log_2016128.log  access_log_2016129.log
access_log_2016130.log  access_log_2016131.log  access_log_2016132.log  access_log_2016133.log  access_log_2016134.log
access_log_2016135.log  access_log_2016136.log  access_log_2016137.log  access_log_2016138.log  access_log_2016139.log
access_log_2016140.log  access_log_2016141.log  access_log_2016142.log  access_log_2016143.log  access_log_2016144.log
access_log_2016145.log  access_log_2016146.log  access_log_2016147.log  access_log_2016148.log  access_log_2016149.log
access_log_2016150.log  access_log_2016151.log  access_log_2016152.log  access_log_2016153.log  access_log_2016154.log
access_log_2016155.log  access_log_2016156.log  access_log_2016157.log                                 59.log
access_log_2016160.log  access_log_2016161.log  access_log_2016162.log                                 54.log
access_log_2016165.log  access_log_2016166.log  access_log_2016167.log  access_log_2016168.log  access_log_2016169.log
access_log_2016170.log  access_log_2016171.log  access_log_2016172.log  access_log_2016173.log  access_log_2016174.log
access_log_2016175.log  access_log_2016176.log  access_log_2016177.log  access_log_2016178.log  access_log_2016179.log
```

```
access_log-2016-02-01_01.log
access_log-2016-02-01_01_2.log
access_log-2016-02-01_01_3.log
access_log-2016-02-01_01_4.log
access_log-2016-02-01_02.log
access_log-2016-02-01_02_2.log
access_log-2016-02-01_02_3.log
```

We also load balance on 6 servers.



Microsoft Excel has stopped working

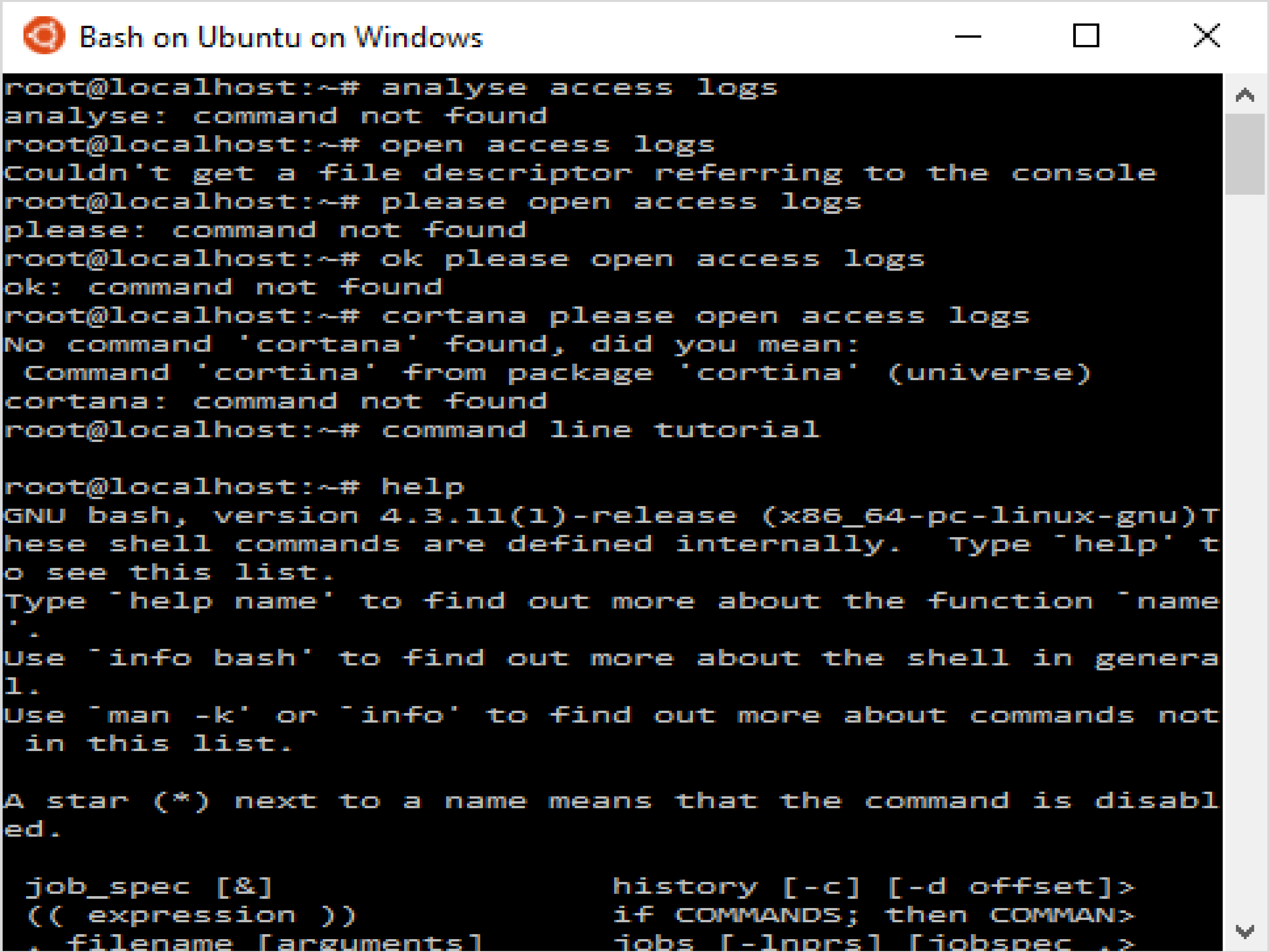
Windows is collecting more information about the problem.
This might take several minutes...

“Just use a sample.”

NO.

“How do I even get a sample?”

Command Line Tools



```
root@localhost:~# analyse access logs
analyse: command not found
root@localhost:~# open access logs
Couldn't get a file descriptor referring to the console
root@localhost:~# please open access logs
please: command not found
root@localhost:~# ok please open access logs
ok: command not found
root@localhost:~# cortana please open access logs
No command 'cortana' found, did you mean:
 Command 'cortina' from package 'cortina' (universe)
cortana: command not found
root@localhost:~# command line tutorial

root@localhost:~# help
GNU bash, version 4.3.11(1)-release (x86_64-pc-linux-gnu)T
hese shell commands are defined internally.  Type `help' t
o see this list.
Type `help name' to find out more about the function `name
'.
Use `info bash' to find out more about the shell in genera
l.
Use `man -k' or `info' to find out more about commands not
 in this list.

A star (*) next to a name means that the command is disabl
ed.

 job_spec [&]                 history [-c] [-d offset]>
 (( expression ))             if COMMANDS; then COMMAN>
 . filename [arguments]       jobs [-lnprs] [jobspec ..>
```

Advantages of Command Line Tools.

- They’re fast.
- They’re not ***in the cloud***.
- The main limit is you, not a development queue.

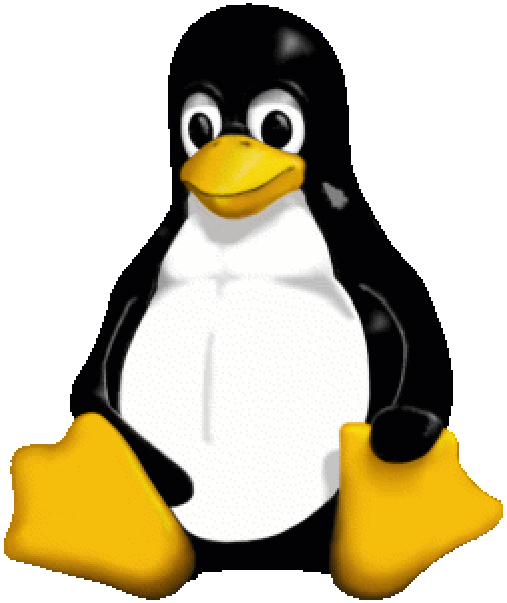
Disadvantages of Command Line Tools.

- They’re scary at first.
- You can delete your computer.
- Don’t delete your computer.

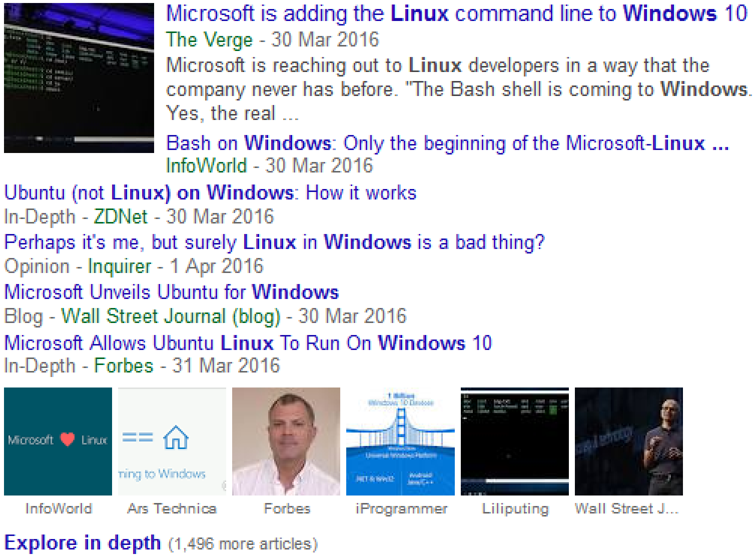
Installation

If you’re on Mac, you’re ready.

If you’re on Linux, you’re ready.



If you’re on Windows, you probably aren’t ready*.



Microsoft is adding the **Linux** command line to **Windows** 10
The Verge - 30 Mar 2016
Microsoft is reaching out to **Linux** developers in a way that the company never has before. "The Bash shell is coming to **Windows**. Yes, the real ...

Bash on **Windows**: Only the beginning of the Microsoft-**Linux ...**
InfoWorld - 30 Mar 2016

Ubuntu (not **Linux**) **on Windows**: How it works
In-Depth - ZDNet - 30 Mar 2016

Perhaps it's me, but surely **Linux** in **Windows** is a bad thing?
Opinion - Inquirer - 1 Apr 2016

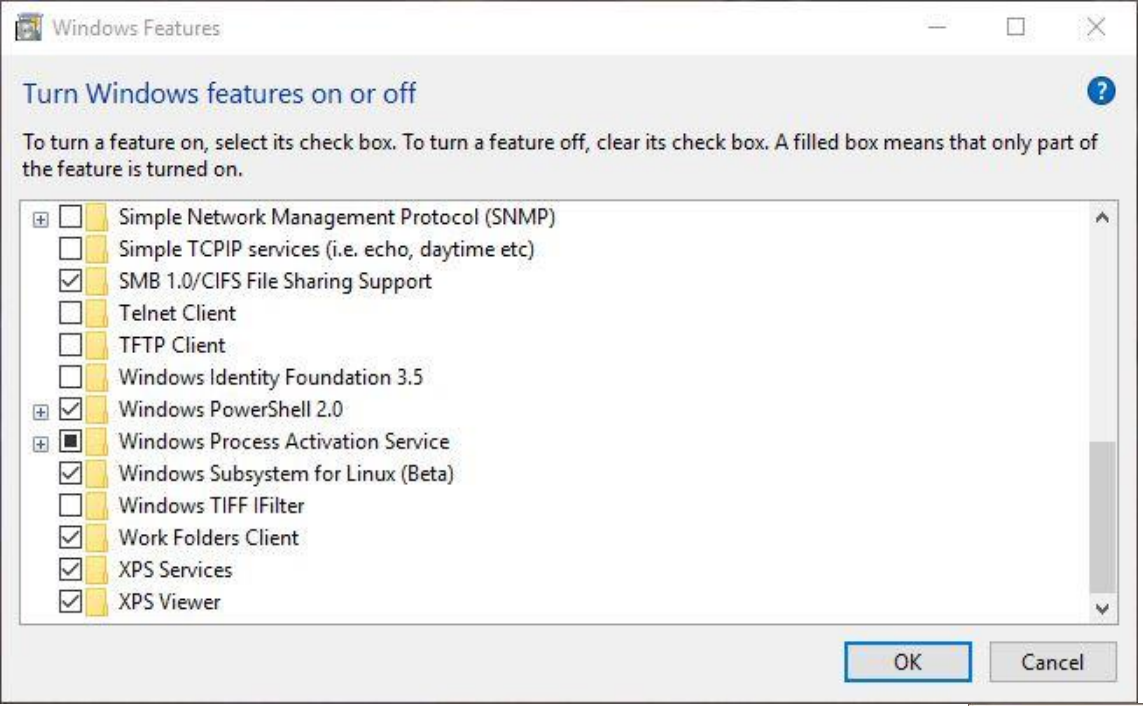
Microsoft Unveils Ubuntu for **Windows**
Blog - Wall Street Journal (blog) - 30 Mar 2016

Microsoft Allows Ubuntu **Linux** To Run On **Windows** 10
In-Depth - Forbes - 31 Mar 2016

InfoWorld · Ars Technica · Forbes · iProgrammer · Liliputing · Wall Street J...

Explore in depth (1,496 more articles)

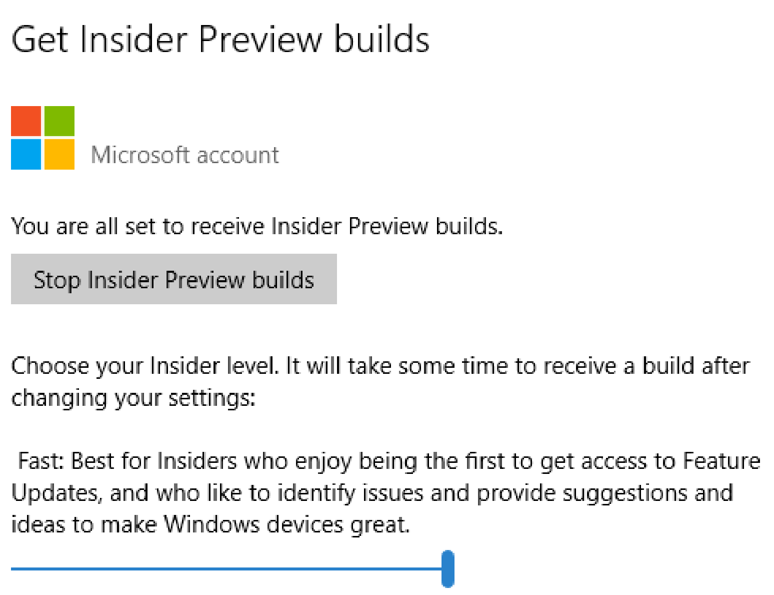
*Unless ‘Ubuntu on Windows’ becomes part of the non-developer release.



Turn Windows features on or off

To turn a feature on, select its check box. To turn a feature off, clear its check box. A filled box means that only part of the feature is turned on.

- ⊞ □ Simple Network Management Protocol (SNMP)
- □ Simple TCPIP services (i.e. echo, daytime etc)
- ☑ SMB 1.0/CIFS File Sharing Support
- □ Telnet Client
- □ TFTP Client
- □ Windows Identity Foundation 3.5
- ⊞ ☑ Windows PowerShell 2.0
- ⊞ ■ Windows Process Activation Service
- ☑ Windows Subsystem for Linux (Beta)
- □ Windows TIFF IFilter
- ☑ Work Folders Client
- ☑ XPS Services
- ☑ XPS Viewer



Get Insider Preview builds

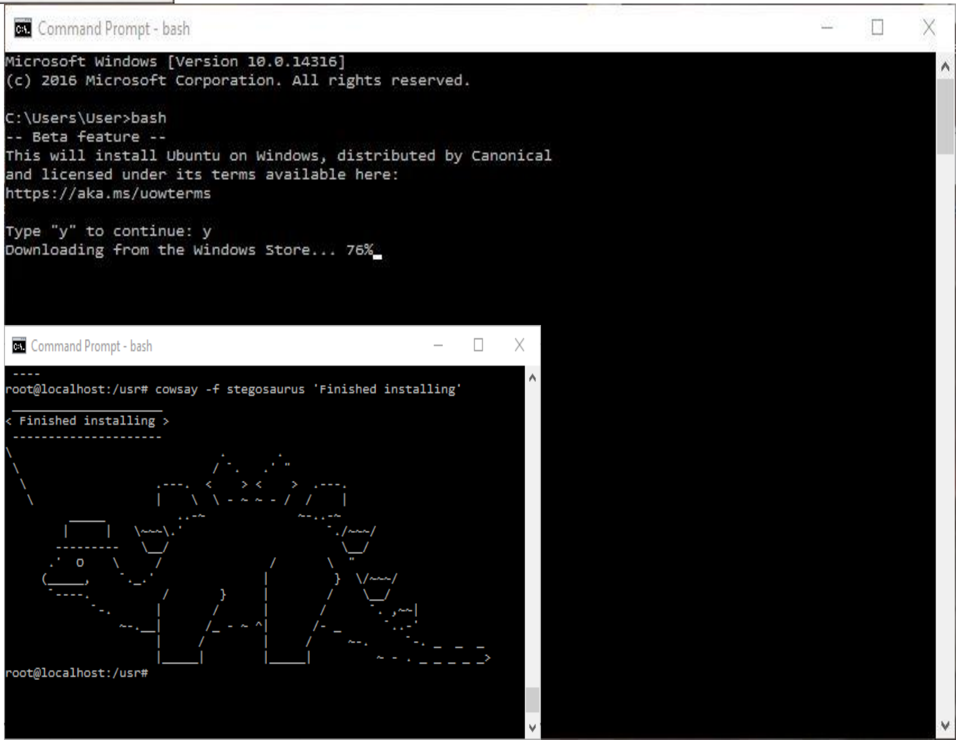
Microsoft account

You are all set to receive Insider Preview builds.

Choose your Insider level. It will take some time to receive a build after changing your settings:

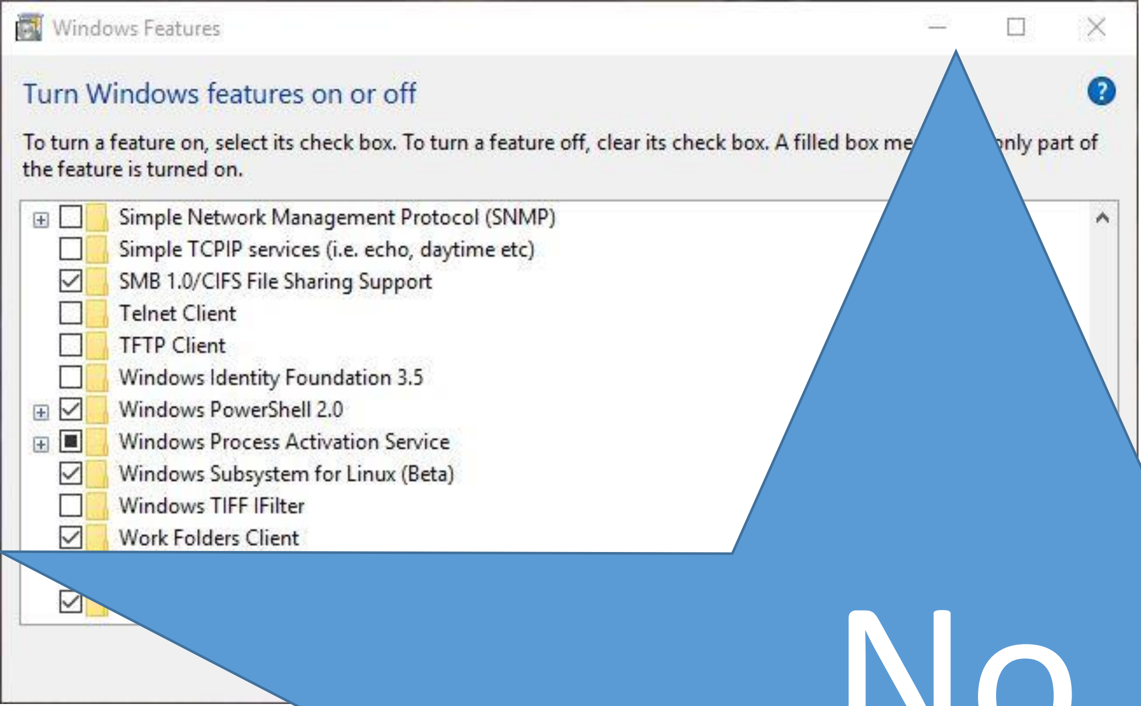
Fast: Best for Insiders who enjoy being the first to get access to Feature Updates, and who like to identify issues and provide suggestions and ideas to make Windows devices great.

1. **Windows Update > Update Settings > Advanced > Get Insider Preview Builds**.
2. Install **Build 14316 or greater**.
3. Enable ‘**Windows Subsystem for Linux (Beta)**’.
4. Open **cmd** and type ‘**bash**’.
5. Type ‘**y**’ and hit enter at the prompt.



```
Microsoft Windows [Version 10.0.14316]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\User>bash
-- Beta feature --
This will install Ubuntu on Windows, distributed by Canonical
and licensed under its terms available here:
https://aka.ms/uowterms

Type "y" to continue: y
Downloading from the Windows Store... 76%
```

```
 ----
root@localhost:/usr# cowsay -f stegosaurus 'Finished installing'
 _____________________
< Finished installing >
 ---------------------
\                             .       .
 \                           / `.   .' "
  \                  .---.  <    > <    >  .---.
   \                 |    \  \ - ~ ~ - /  /    |
         _____          ..-~             ~-..-~
        |     |   \~~~\.'                    `./~~~/
       ---------   \__/                        \__/
      .'  O    \     /               /       \  "
     (_____,    `._.'               |         }  \/~~~/
      `----.          /       }     |        /    \__/
            `-.      |       /      |       /      `. ,~~|
                ~-.__|      /_ - ~ ^|      /- _      `..-'
                     |     /        |     /     ~-.     `-. _  _  _
                     |_____|        |_____|         ~ - . _ _ _ _ _>
root@localhost:/usr#
```

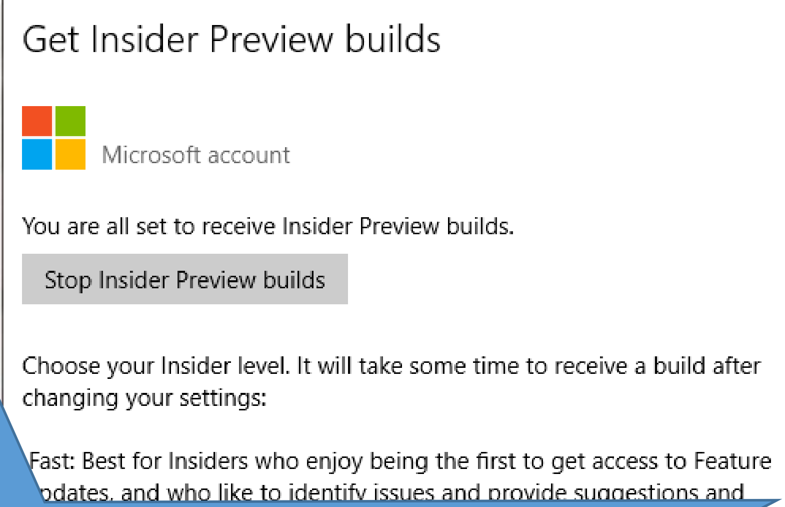


Windows Features

Turn Windows features on or off

To turn a feature on, select its check box. To turn a feature off, clear its check box. A filled box me… …nly part of the feature is turned on.

- ⊞ □ Simple Network Management Protocol (SNMP)
- □ Simple TCPIP services (i.e. echo, daytime etc)
- ☑ SMB 1.0/CIFS File Sharing Support
- □ Telnet Client
- □ TFTP Client
- □ Windows Identity Foundation 3.5
- ⊞ ☑ Windows PowerShell 2.0
- ⊞ ■ Windows Process Activation Service
- ☑ Windows Subsystem for Linux (Beta)
- □ Windows TIFF IFilter
- ☑ Work Folders Client
- ☑



Get Insider Preview builds

Microsoft account

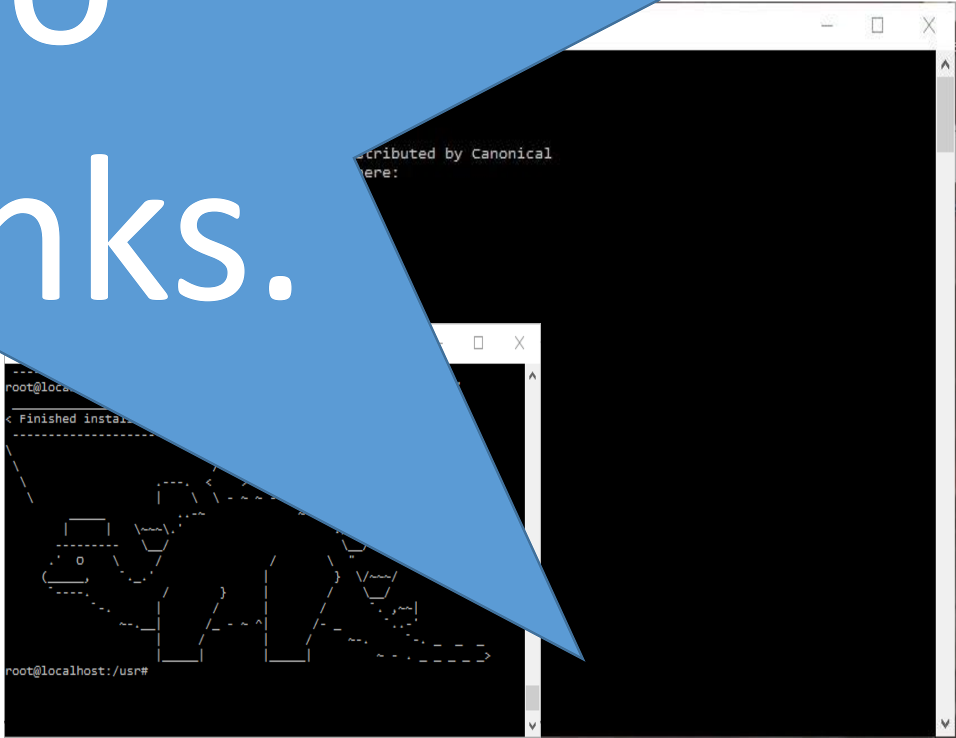
You are all set to receive Insider Preview builds.

Choose your Insider level. It will take some time to receive a build after changing your settings:

Fast: Best for Insiders who enjoy being the first to get access to Feature …pdates, and who like to identify issues and provide suggestions and

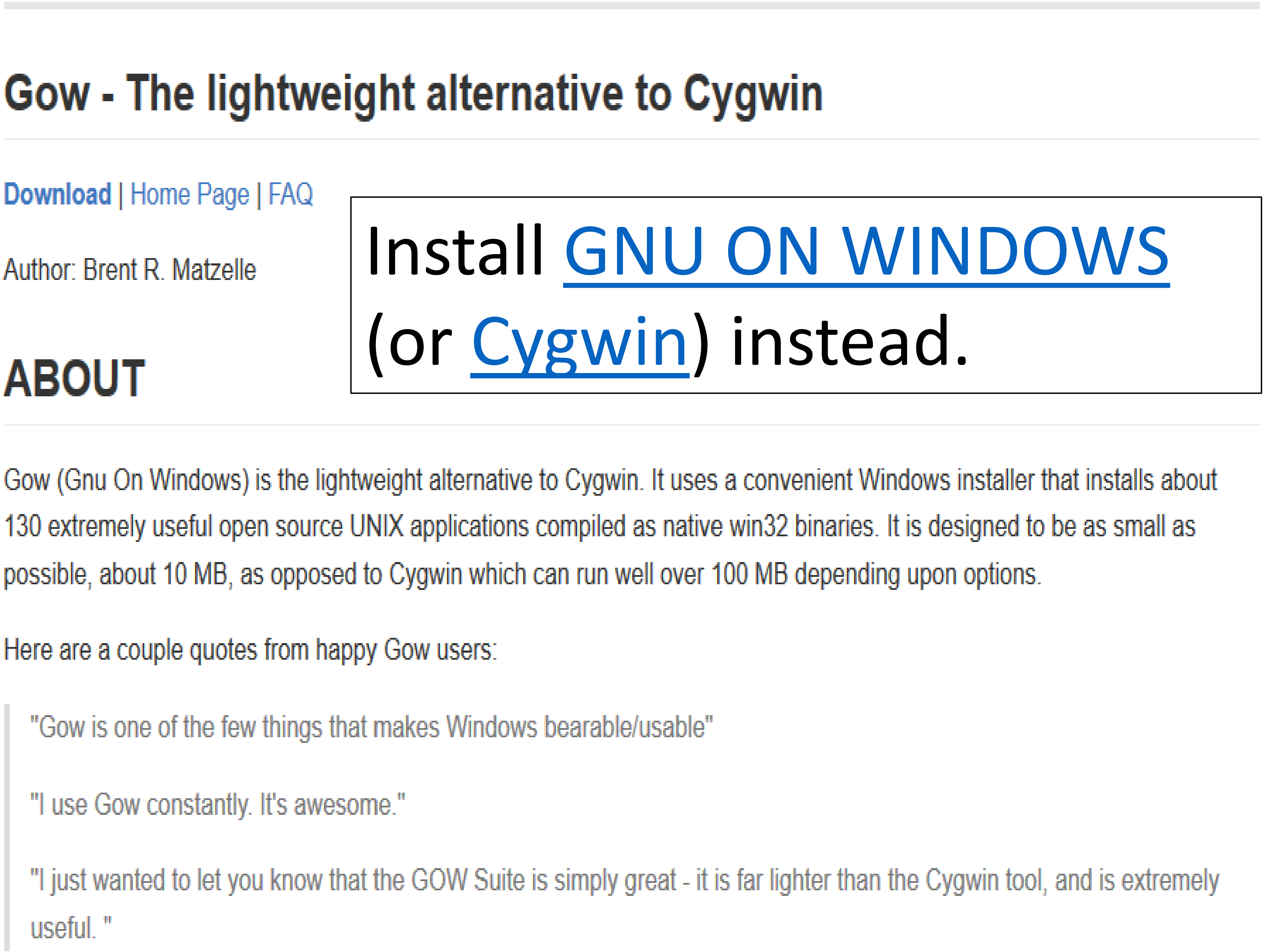
No thanks.

1. **Windows Update > Up… Advanced > Get Insid… Builds**.
2. Install **Build 14316**
3. Enable ‘**Windows** … **Linux (Beta)**’.
4. Open **cmd** and … …ash’.
5. Type ‘y’ and h… …nter at the prompt.



```
…tributed by Canonical
…ere:
```

```
root@loca…
< Finished insta…
root@localhost:/usr#
```



Gow - The lightweight alternative to Cygwin

Download | Home Page | FAQ

Author: Brent R. Matzelle

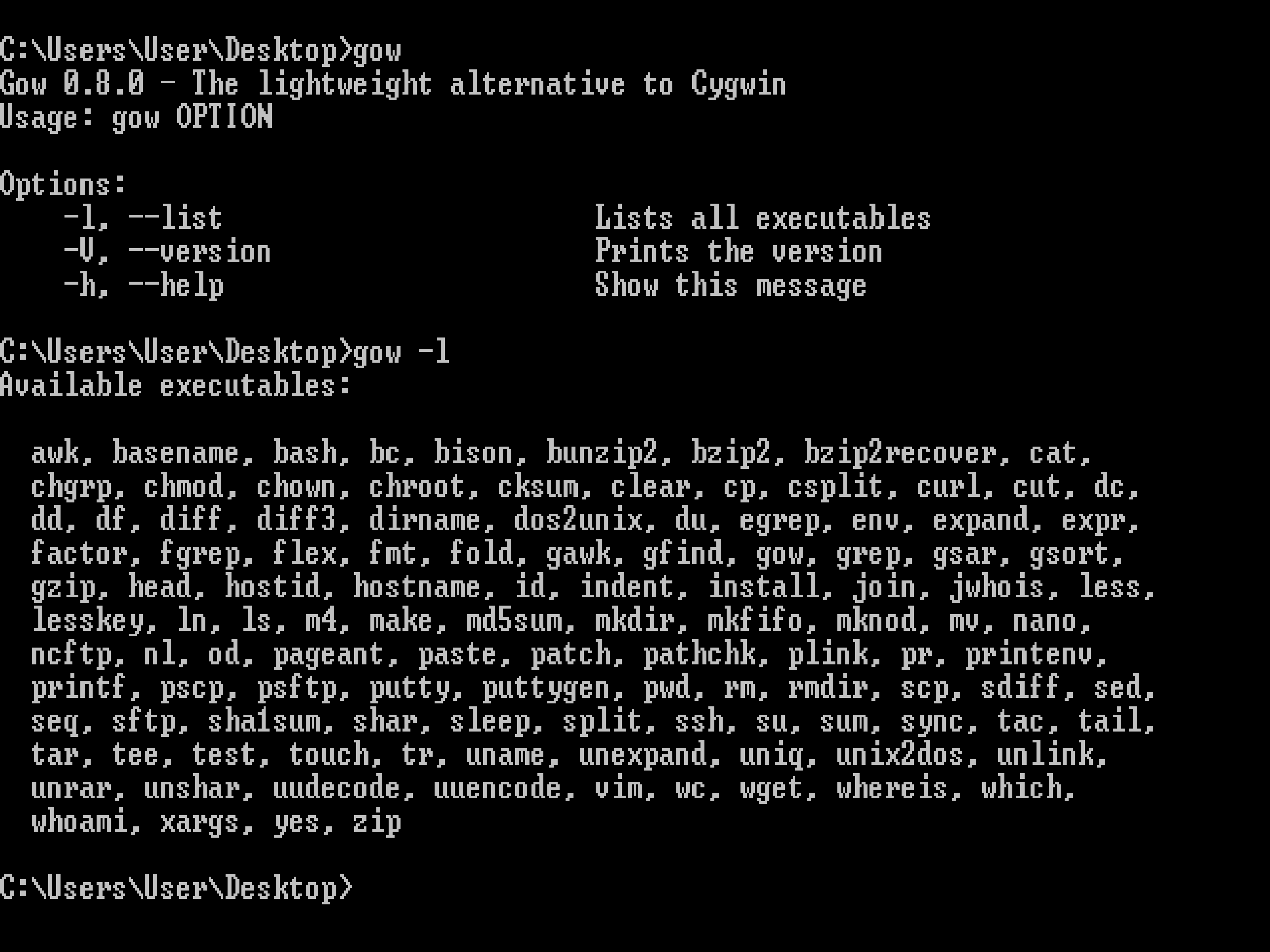
Install GNU ON WINDOWS (or Cygwin) instead.

ABOUT

Gow (Gnu On Windows) is the lightweight alternative to Cygwin. It uses a convenient Windows installer that installs about 130 extremely useful open source UNIX applications compiled as native win32 binaries. It is designed to be as small as possible, about 10 MB, as opposed to Cygwin which can run well over 100 MB depending upon options.

Here are a couple quotes from happy Gow users:

> "Gow is one of the few things that makes Windows bearable/usable"
>
> "I use Gow constantly. It's awesome."
>
> "I just wanted to let you know that the GOW Suite is simply great - it is far lighter than the Cygwin tool, and is extremely useful. "

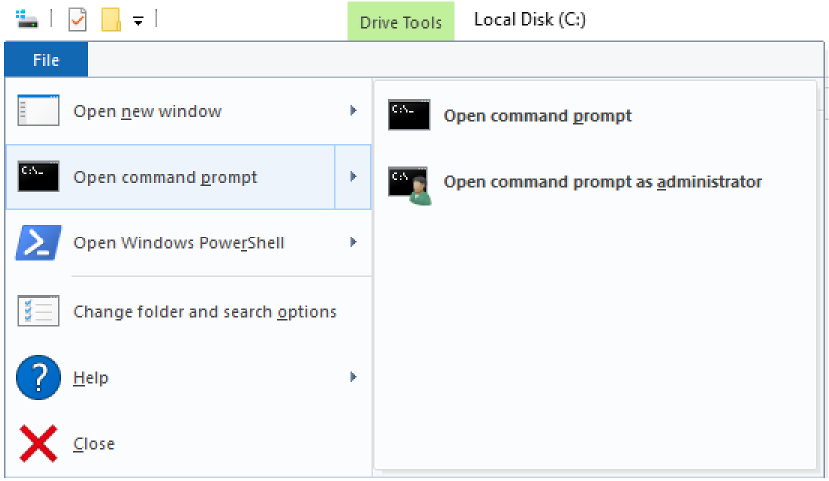


```
C:\Users\User\Desktop>gow
Gow 0.8.0 - The lightweight alternative to Cygwin
Usage: gow OPTION

Options:
    -l, --list                       Lists all executables
    -V, --version                    Prints the version
    -h, --help                       Show this message

C:\Users\User\Desktop>gow -l
Available executables:

  awk, basename, bash, bc, bison, bunzip2, bzip2, bzip2recover, cat,
  chgrp, chmod, chown, chroot, cksum, clear, cp, csplit, curl, cut, dc,
  dd, df, diff, diff3, dirname, dos2unix, du, egrep, env, expand, expr,
  factor, fgrep, flex, fmt, fold, gawk, gfind, gow, grep, gsar, gsort,
  gzip, head, hostid, hostname, id, indent, install, join, jwhois, less,
  lesskey, ln, ls, m4, make, md5sum, mkdir, mkfifo, mknod, mv, nano,
  ncftp, nl, od, pageant, paste, patch, pathchk, plink, pr, printenv,
  printf, pscp, psftp, putty, puttygen, pwd, rm, rmdir, scp, sdiff, sed,
  seq, sftp, sha1sum, shar, sleep, split, ssh, su, sum, sync, tac, tail,
  tar, tee, test, touch, tr, uname, unexpand, uniq, unix2dos, unlink,
  unrar, unshar, uudecode, uuencode, vim, wc, wget, whereis, which,
  whoami, xargs, yes, zip

C:\Users\User\Desktop>
```

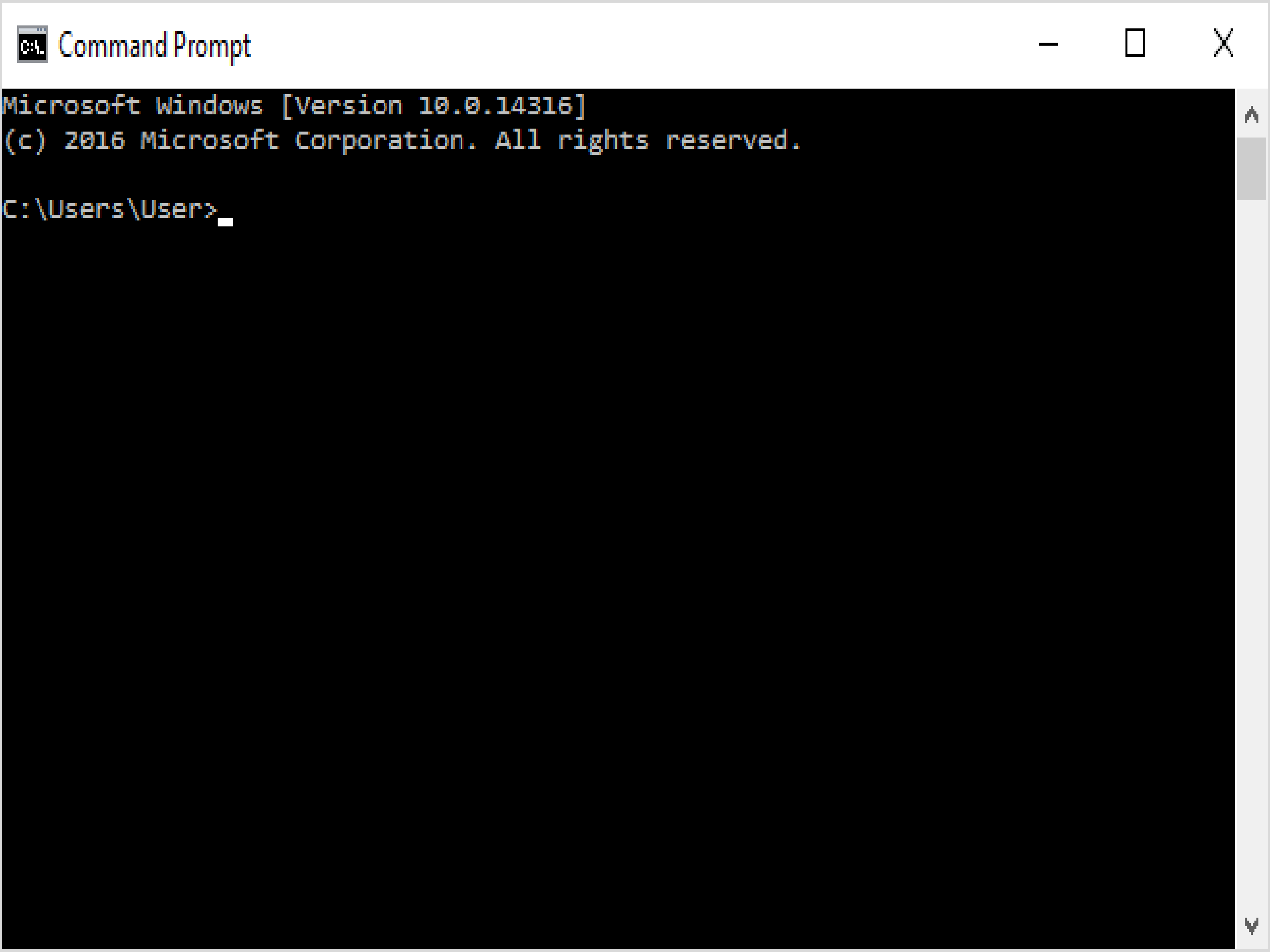
Installation Done

Getting Started

- Navigate to the folder containing the downloaded files.
- Open your chosen terminal (cmd, terminal, or bash).



CTRL+SHIFT+Rclick inside a folder is an alternate method.



```
Microsoft Windows [Version 10.0.14316]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\User>
```

```
~$ type-commands-here
```

Then hit enter.

```
~$ echo hello.

hello.
```

The Title of The Talk Was a Lie and We’re Going to Try to Use Excel Anyway. Sorry.

```
 ________________________
< I'm Sorry #BrightonSEO >
 ------------------------
```

Server Logs

Until Excel Fails
@ohgm

Sorry about the walls of text.

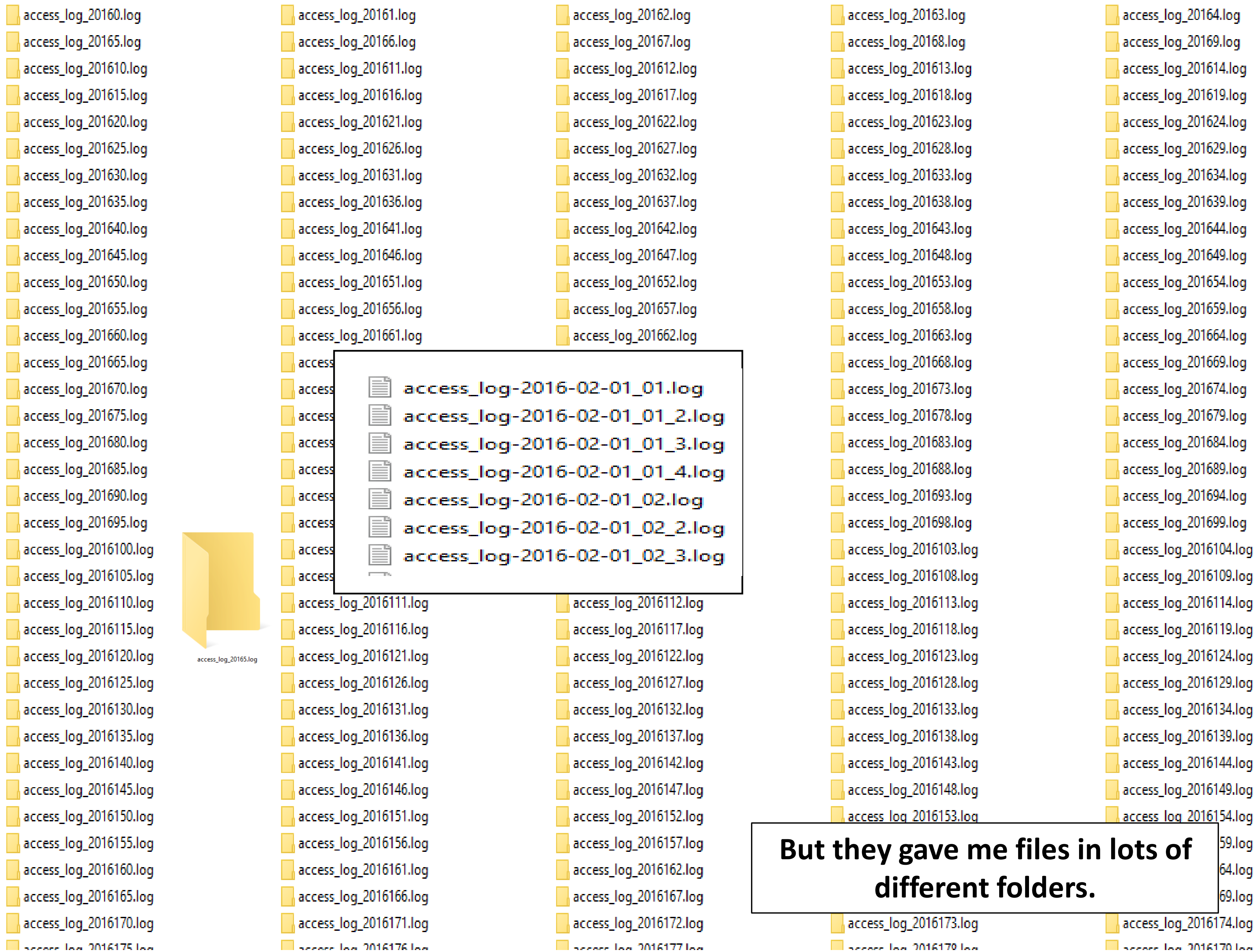
Combining Files.

Combine Multiple Log Files

We navigate to a folder containing all our server logs, open the terminal, and type:

```
~$ cat *.log >> combined.log
```

*"Take **every** **.log file** in the folder and **append** each to combined.log"*



access_log_20160.log access_log_20161.log access_log_20162.log access_log_20163.log access_log_20164.log
access_log_20165.log access_log_20166.log access_log_20167.log access_log_20168.log access_log_20169.log
access_log_201610.log access_log_201611.log access_log_201612.log access_log_201613.log access_log_201614.log
access_log_201615.log access_log_201616.log access_log_201617.log access_log_201618.log access_log_201619.log
access_log_201620.log access_log_201621.log access_log_201622.log access_log_201623.log access_log_201624.log
access_log_201625.log access_log_201626.log access_log_201627.log access_log_201628.log access_log_201629.log
access_log_201630.log access_log_201631.log access_log_201632.log access_log_201633.log access_log_201634.log
access_log_201635.log access_log_201636.log access_log_201637.log access_log_201638.log access_log_201639.log
access_log_201640.log access_log_201641.log access_log_201642.log access_log_201643.log access_log_201644.log
access_log_201645.log access_log_201646.log access_log_201647.log access_log_201648.log access_log_201649.log
access_log_201650.log access_log_201651.log access_log_201652.log access_log_201653.log access_log_201654.log
access_log_201655.log access_log_201656.log access_log_201657.log access_log_201658.log access_log_201659.log
access_log_201660.log access_log_201661.log access_log_201662.log access_log_201663.log access_log_201664.log
access_log_201665.log access access_log_201668.log access_log_201669.log
access_log_201670.log access access_log_201673.log access_log_201674.log
access_log_201675.log access access_log_201678.log access_log_201679.log
access_log_201680.log access access_log_201683.log access_log_201684.log
access_log_201685.log access access_log_201688.log access_log_201689.log
access_log_201690.log access access_log_201693.log access_log_201694.log
access_log_201695.log access access_log_201698.log access_log_201699.log
access_log_2016100.log access access_log_2016103.log access_log_2016104.log
access_log_2016105.log access access_log_2016108.log access_log_2016109.log
access_log_2016110.log access_log_2016111.log access_log_2016112.log access_log_2016113.log access_log_2016114.log
access_log_2016115.log access_log_2016116.log access_log_2016117.log access_log_2016118.log access_log_2016119.log
access_log_2016120.log access_log_2016121.log access_log_2016122.log access_log_2016123.log access_log_2016124.log
access_log_2016125.log access_log_2016126.log access_log_2016127.log access_log_2016128.log access_log_2016129.log
access_log_2016130.log access_log_2016131.log access_log_2016132.log access_log_2016133.log access_log_2016134.log
access_log_2016135.log access_log_2016136.log access_log_2016137.log access_log_2016138.log access_log_2016139.log
access_log_2016140.log access_log_2016141.log access_log_2016142.log access_log_2016143.log access_log_2016144.log
access_log_2016145.log access_log_2016146.log access_log_2016147.log access_log_2016148.log access_log_2016149.log
access_log_2016150.log access_log_2016151.log access_log_2016152.log access_log_2016153.log access_log_2016154.log
access_log_2016155.log access_log_2016156.log access_log_2016157.log 59.log
access_log_2016160.log access_log_2016161.log access_log_2016162.log 64.log
access_log_2016165.log access_log_2016166.log access_log_2016167.log 69.log
access_log_2016170.log access_log_2016171.log access_log_2016172.log access_log_2016173.log access_log_2016174.log
access_log_2016175.log access_log_2016176.log access_log_2016177.log access_log_2016178.log access_log_2016179.log

access_log_20165.log

access_log-2016-02-01_01.log
access_log-2016-02-01_01_2.log
access_log-2016-02-01_01_3.log
access_log-2016-02-01_01_4.log
access_log-2016-02-01_02.log
access_log-2016-02-01_02_2.log
access_log-2016-02-01_02_3.log

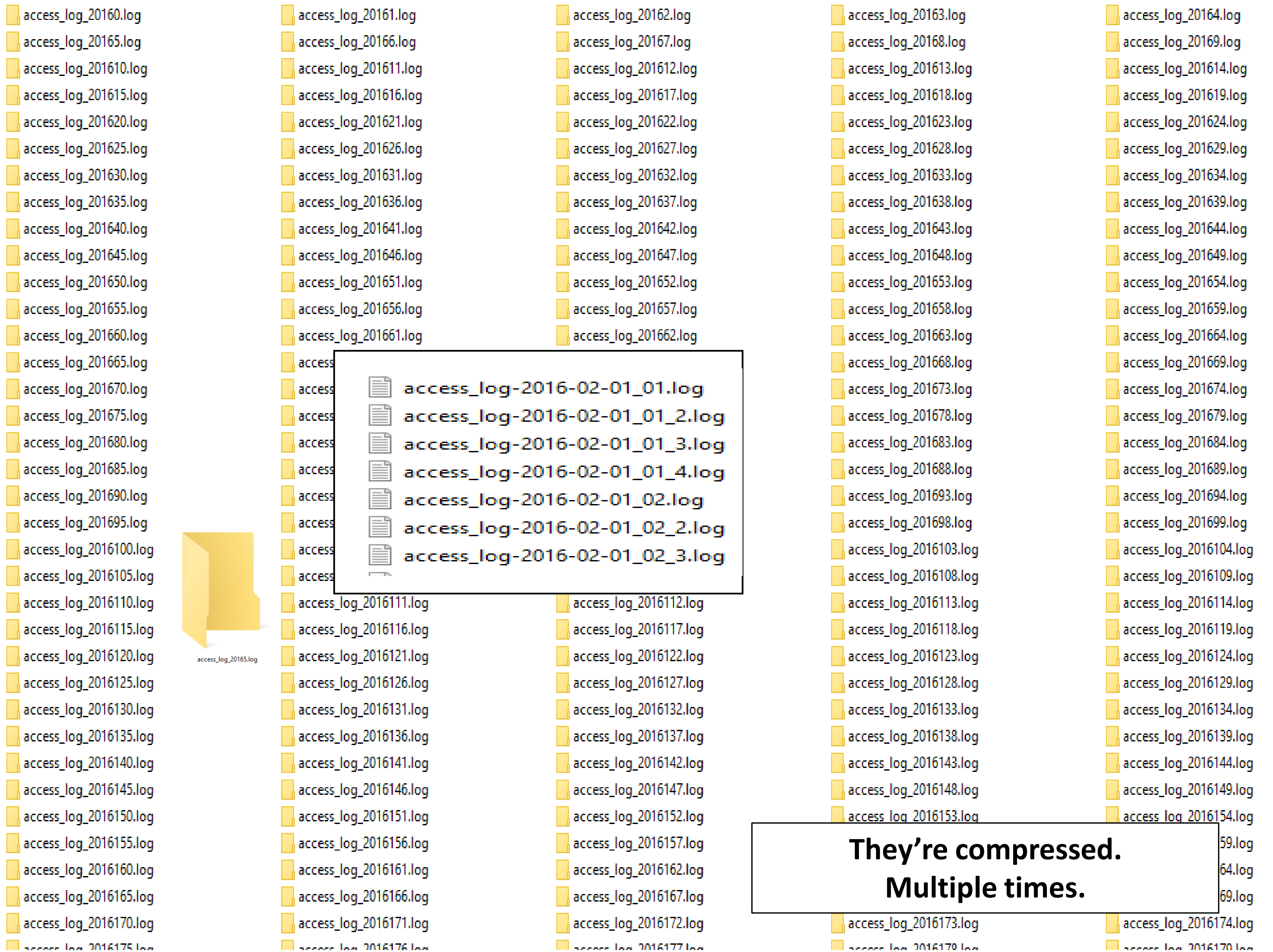
But they gave me files in lots of different folders.

Combine Multiple Files in Multiple Folders

~$ find . -name '*.log' -exec cat {} >> combined.log ;

*"**Search** the **current folder**, and all subfolders for filenames ending with '**.log**'. Append the contents of these files to a new file called combined.log."*

gfind in GOW



```
access_log_20160.log     access_log_20161.log     access_log_20162.log     access_log_20163.log     access_log_20164.log
access_log_20165.log     access_log_20166.log     access_log_20167.log     access_log_20168.log     access_log_20169.log
access_log_201610.log    access_log_201611.log    access_log_201612.log    access_log_201613.log    access_log_201614.log
access_log_201615.log    access_log_201616.log    access_log_201617.log    access_log_201618.log    access_log_201619.log
access_log_201620.log    access_log_201621.log    access_log_201622.log    access_log_201623.log    access_log_201624.log
access_log_201625.log    access_log_201626.log    access_log_201627.log    access_log_201628.log    access_log_201629.log
access_log_201630.log    access_log_201631.log    access_log_201632.log    access_log_201633.log    access_log_201634.log
access_log_201635.log    access_log_201636.log    access_log_201637.log    access_log_201638.log    access_log_201639.log
access_log_201640.log    access_log_201641.log    access_log_201642.log    access_log_201643.log    access_log_201644.log
access_log_201645.log    access_log_201646.log    access_log_201647.log    access_log_201648.log    access_log_201649.log
access_log_201650.log    access_log_201651.log    access_log_201652.log    access_log_201653.log    access_log_201654.log
access_log_201655.log    access_log_201656.log    access_log_201657.log    access_log_201658.log    access_log_201659.log
access_log_201660.log    access_log_201661.log    access_log_201662.log    access_log_201663.log    access_log_201664.log
access_log_201665.log    access                                            access_log_201668.log    access_log_201669.log
access_log_201670.log    access                                            access_log_201673.log    access_log_201674.log
access_log_201675.log    access                                            access_log_201678.log    access_log_201679.log
access_log_201680.log    access                                            access_log_201683.log    access_log_201684.log
access_log_201685.log    access                                            access_log_201688.log    access_log_201689.log
access_log_201690.log    access                                            access_log_201693.log    access_log_201694.log
access_log_201695.log    access                                            access_log_201698.log    access_log_201699.log
access_log_2016100.log   access                                            access_log_2016103.log   access_log_2016104.log
access_log_2016105.log   access                                            access_log_2016108.log   access_log_2016109.log
access_log_2016110.log   access_log_2016111.log   access_log_2016112.log   access_log_2016113.log   access_log_2016114.log
access_log_2016115.log   access_log_2016116.log   access_log_2016117.log   access_log_2016118.log   access_log_2016119.log
access_log_2016120.log   access_log_2016121.log   access_log_2016122.log   access_log_2016123.log   access_log_2016124.log
access_log_2016125.log   access_log_2016126.log   access_log_2016127.log   access_log_2016128.log   access_log_2016129.log
access_log_2016130.log   access_log_2016131.log   access_log_2016132.log   access_log_2016133.log   access_log_2016134.log
access_log_2016135.log   access_log_2016136.log   access_log_2016137.log   access_log_2016138.log   access_log_2016139.log
access_log_2016140.log   access_log_2016141.log   access_log_2016142.log   access_log_2016143.log   access_log_2016144.log
access_log_2016145.log   access_log_2016146.log   access_log_2016147.log   access_log_2016148.log   access_log_2016149.log
access_log_2016150.log   access_log_2016151.log   access_log_2016152.log   access_log_2016153.log   access_log_2016154.log
access_log_2016155.log   access_log_2016156.log   access_log_2016157.log                            59.log
access_log_2016160.log   access_log_2016161.log   access_log_2016162.log                            64.log
access_log_2016165.log   access_log_2016166.log   access_log_2016167.log                            69.log
access_log_2016170.log   access_log_2016171.log   access_log_2016172.log   access_log_2016173.log   access_log_2016174.log
access_log_2016175.log   access_log_2016176.log   access_log_2016177.log   access_log_2016178.log   access_log_2016179.log
```

```
access_log-2016-02-01_01.log
access_log-2016-02-01_01_2.log
access_log-2016-02-01_01_3.log
access_log-2016-02-01_01_4.log
access_log-2016-02-01_02.log
access_log-2016-02-01_02_2.log
access_log-2016-02-01_02_3.log
```

They're compressed.
Multiple times.

Combine Multiple Files in Multiple Folders Some of Which are compressed

```
~$ find . -name *.gz -exec gzip -dkr {} +
&& find . -name ‘*.log’ -exec cat {} >> combined.log ;
```

“Find all the files with the .gz extension beneath the current folder.

***Recursively Decompress** all files. **Keep** the originals.*

***Once finished**, find all the .log files, append them to a new combined.log file.”*

Preview Huge Files with less

less *streams* the contents of a file to the terminal without loading the whole file into memory.

```
$~ less combined.log
```

You can use less to review access logs without crippling your machine.

HELP.

R.T.F.M

READ THE *FRIENDLY* MANUAL

RTFM

If at any time you get stuck:

```
~$ toolname --help
```

or

```
~$ man toolname
```

or

Google what you are trying to do.

The **--help** (often **-h**) flag will usually give you what you need to know. ‘**man**’ (manual) tends to be much more in depth. Both are read from the command line.

We now have one large file.

UA Filtering: Googlebot

Our combined access logs are in a single file:

combined.log – 16.4GB

Too large to open in Excel.
Too large to open in Notepad.
*Examining it with **less**?*
It’s too full of filthy human data.

We need to cut it down to Googlebot.

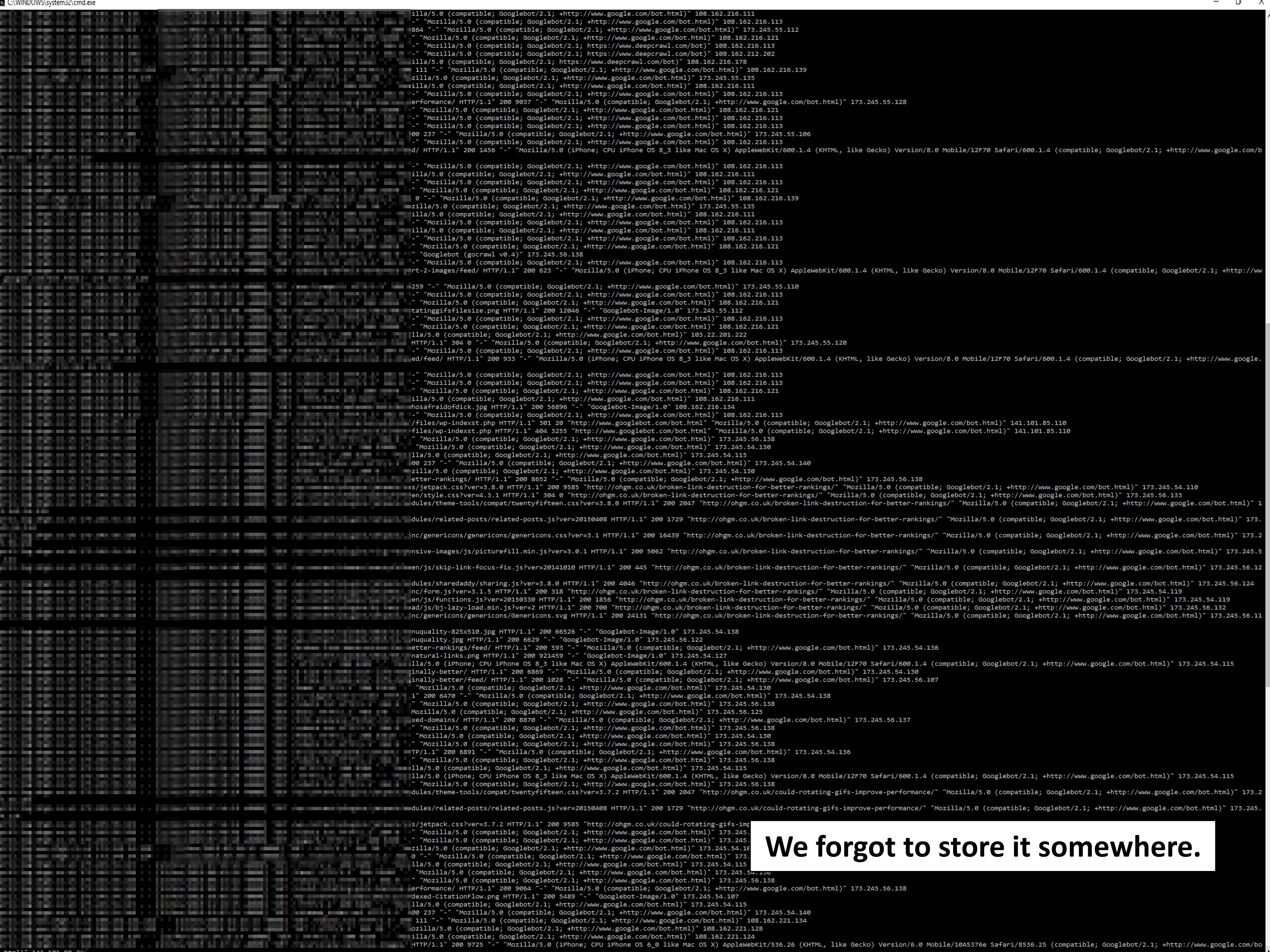
grep

grep is a tool that extracts lines from text files based on a regular expression. Using **grep** is pretty simple:

```
~$ grep [options] [pattern] [file]
…
~$ grep ‘Googlebot’ combined.log
```

“Give me all the lines containing Googlebot in combined.log”

Press Enter.



[redacted]

```
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.111
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
864 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.112
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
-" "Mozilla/5.0 (compatible; Googlebot/2.1; https://www.deepcrawl.com/bot)" 108.162.216.113
-" "Mozilla/5.0 (compatible; Googlebot/2.1; https://www.deepcrawl.com/bot)" 108.162.212.202
illa/5.0 (compatible; Googlebot/2.1; https://www.deepcrawl.com/bot)" 108.162.216.178
 111 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.139
zilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.135
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.111
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
erformance/ HTTP/1.1" 200 9037 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.128
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
00 237 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.106
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
d/ HTTP/1.1" 200 1456 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://www.google.com/b

-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.111
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
 0 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.139
zilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.135
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.111
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.111
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
" "Googlebot (gocrawl v0.4)" 173.245.56.138
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
rt-2-images/feed/ HTTP/1.1" 200 623 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://ww

259 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.110
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
tatinggifsfilesize.png HTTP/1.1" 200 12046 "-" "Googlebot-Image/1.0" 173.245.55.112
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
lla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 103.22.201.222
HTTP/1.1" 304 0 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.55.120
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
ed/feed/ HTTP/1.1" 200 933 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://www.google.

-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.121
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.111
hosafraidofdick.jpg HTTP/1.1" 200 56896 "-" "Googlebot-Image/1.0" 108.162.216.134
-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.216.113
/files/wp-indexst.php HTTP/1.1" 301 20 "http://www.googlebot.com/bot.html" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 141.101.85.110
files/wp-indexst.php HTTP/1.1" 404 3255 "http://www.googlebot.com/bot.html" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 141.101.85.110
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
 "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.130
lla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.115
00 237 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.140
zilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.130
etter-rankings/ HTTP/1.1" 200 8652 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
s/jetpack.css?ver=3.8.0 HTTP/1.1" 200 9585 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.110
en/style.css?ver=4.3.1 HTTP/1.1" 304 0 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.133
dules/theme-tools/compat/twentyfifteen.css?ver=3.8.0 HTTP/1.1" 200 2047 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 1

dules/related-posts/related-posts.js?ver=20150408 HTTP/1.1" 200 1729 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.

nc/genericons/genericons/genericons.css?ver=3.1 HTTP/1.1" 200 16439 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.2

nsive-images/js/picturefill.min.js?ver=3.0.1 HTTP/1.1" 200 5062 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.5

en/js/skip-link-focus-fix.js?ver=20141010 HTTP/1.1" 200 445 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.12

dules/sharedaddy/sharing.js?ver=3.8.0 HTTP/1.1" 200 4046 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.124
nc/form.js?ver=3.1.5 HTTP/1.1" 200 318 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.119
en/js/functions.js?ver=20150330 HTTP/1.1" 200 1856 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.119
ad/js/bj-lazy-load.min.js?ver=2 HTTP/1.1" 200 700 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.132
nc/genericons/genericons/Genericons.svg HTTP/1.1" 200 24131 "http://ohgm.co.uk/broken-link-destruction-for-better-rankings/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.11

nuquality-825x510.jpg HTTP/1.1" 200 66526 "-" "Googlebot-Image/1.0" 173.245.54.138
nuquality.jpg HTTP/1.1" 200 6629 "-" "Googlebot-Image/1.0" 173.245.56.122
etter-rankings/feed/ HTTP/1.1" 200 593 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.136
natural-links.png HTTP/1.1" 200 921459 "-" "Googlebot-Image/1.0" 173.245.54.127
lla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.115
inally-better/ HTTP/1.1" 200 6869 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.130
inally-better/feed/ HTTP/1.1" 200 1028 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.107
 "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.130
1" 200 6470 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.138
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.125
xed-domains/ HTTP/1.1" 200 8870 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.137
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
 "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.130
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
TP/1.1" 200 6891 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.136
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
lla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.115
lla/5.0 (iPhone; CPU iPhone OS 8_3 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12F70 Safari/600.1.4 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.115
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
dules/theme-tools/compat/twentyfifteen.css?ver=3.7.2 HTTP/1.1" 200 2047 "http://ohgm.co.uk/could-rotating-gifs-improve-performance/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.2

dules/related-posts/related-posts.js?ver=20150408 HTTP/1.1" 200 1729 "http://ohgm.co.uk/could-rotating-gifs-improve-performance/" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.

s/jetpack.css?ver=3.7.2 HTTP/1.1" 200 9585 "http://ohgm.co.uk/could-rotating-gifs-imp
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.
zilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.1
0 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.
lla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.115
 "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.130
" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
erformance/ HTTP/1.1" 200 9064 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.56.138
dexed-CitationFlow.png HTTP/1.1" 200 5489 "-" "Googlebot-Image/1.0" 173.245.54.107
lla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.115
00 237 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 173.245.54.140
 111 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.221.134
ozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.221.128
illa/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 108.162.221.124
HTTP/1.1" 200 9725 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 6_0 like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10A5376e Safari/8536.25 (compatible; Googlebot/2.1; +http://www.google.com/bo
```

We forgot to store it somewhere.

Filtering Scenario: Googlebot

So we’ll store this output to a new file using ‘>>’:

```
~$ grep ‘Googlebot’ combined.log >> googlebot.log
```

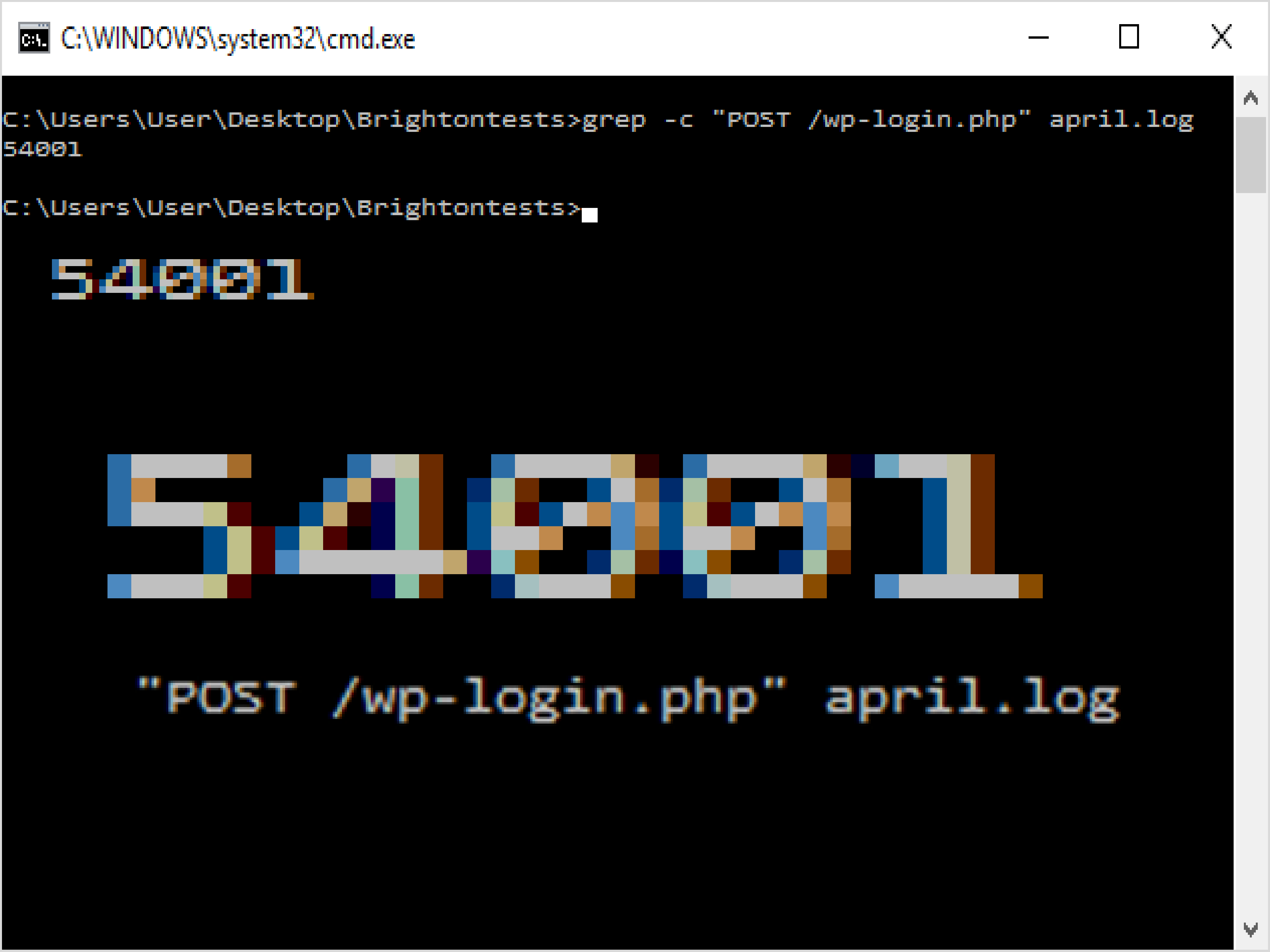
*“**Append** all lines in **combined.log** that contain **Googlebot** into a new file, **googlebot.log**”*

grep

Like other tools, **grep** has a number of optional *argument flags*. The **count** flag '**-c**' can provide a useful summary for direct questions:

```
~$ grep [options] [pattern] [file]
…
~$ grep -c “POST /wp-login” april.log
```

*“Show me the **count** of **login attempts** in **April on ohgm.co.uk**”*



```
C:\Users\User\Desktop\Brightontests>grep -c "POST /wp-login.php" april.log
54001

C:\Users\User\Desktop\Brightontests>
```

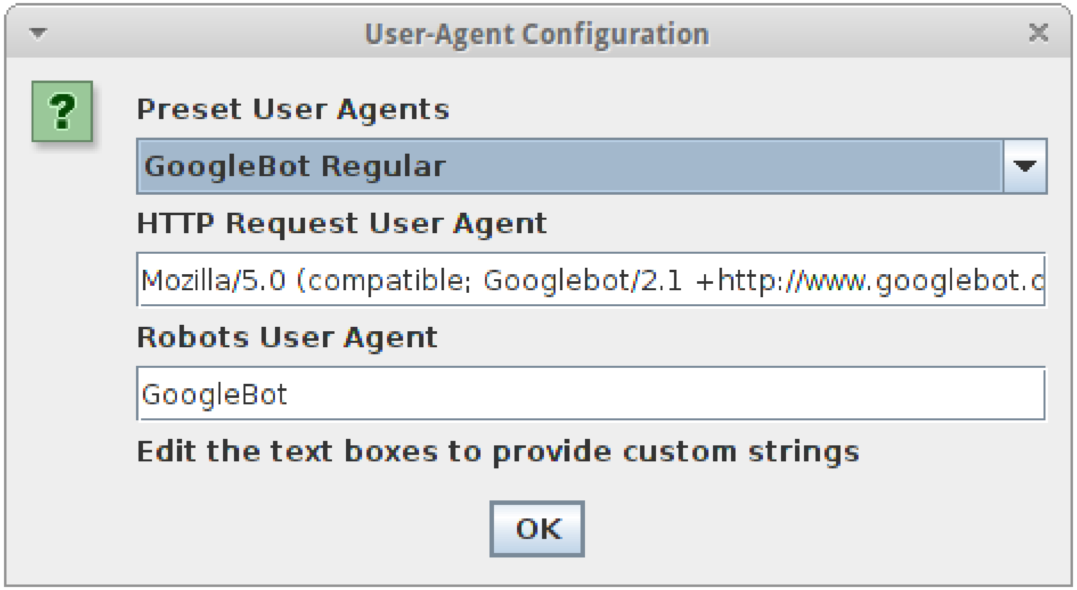
54001

54001

"POST /wp-login.php" april.log

Filtering Scenario – Googlebot

You can't just verify Googlebot by name.



User-Agent Configuration

Preset User Agents

GoogleBot Regular

HTTP Request User Agent

Mozilla/5.0 (compatible; Googlebot/2.1 +http://www.googlebot.c

Robots User Agent

GoogleBot

Edit the text boxes to provide custom strings

Apparently some people aren't honest on the internet.

IP Filtering

Filtering Scenario – IP Ranges

Start	End
64.233.160.0	64.233.191.255
66.102.0.0	66.102.15.255
66.249.64.0	66.249.95.255
72.14.192.0	72.14.255.255
74.125.0.0	74.125.255.255
209.85.128.0	209.85.255.255
216.239.32.0	216.239.63.255

If we were masochistic, we could write a regular expression to capture these all…

Filtering Scenario – IP Ranges

The **-E** flag lets **grep** use **Extended Regular Expressions**.

```
~$ grep -E "((\b(64)\.233\.(1([6-8][0-9]|9[0-1]))))|(\b(66)\.102\.([0-9]|1[0-5]))|(\b(66)\.249\.(6[4-9]|[7-8][0-9]|9[0-5]))|(\b(72)\.14\.(1(9[2-9])|2([0-4][0-9]|5[0-5])))|(\b(74)\.125\.(25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?))|(209\.85\.(1(2[8-9]|[3-9][0-9])|2([0-4][0-9]|5[0-5])))|(216\.239\.(25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)))\.(25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)" GbotUA.log > GbotIP.log
```

This shouldn’t work, but it does*.

***WOMM**

Filtering Scenario – Impostors

The **-v** flag **inverts** the **grep** query to find impostors:

```
~$ grep -vE "((\b(64)\.233\.(1([6-8][0-9]|9[0-1])))|(\b(66)\.102\.([0-9]|1[0-5]))|(\b(66)\.249\.(6[4-9]|[7-8][0-9]|9[0-5]))|(\b(72)\.14\.(1(9[2-9])|2([0-4][0-9]|5[0-5])))|(\b(74)\.125\.(25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?))|(209\.85\.(1(2[8-9]|[3-9][0-9])|2([0-4][0-9]|5[0-5])))|(216\.239\.(25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)))\.(25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)" GbotUA.log > Impostors.log
```

“Give me every request that claims to be Googlebot, but ***doesn’t*** *come from this IP range. Put them in an impostors file.”*

Filtering Scenario – Verifying Googlebot

- **Disclaimer**: don’t blindly use awful regex (check with Regexr) or IP ranges, especially if you’re analysing logs for a site using IP detection for international SEO purposes. Read more about Googlebot’s Geo-distributed Crawling here first.
- Use the correct reverse DNS > forward DNS lookup when it’s important to be right. This can be automated.

Filtering Scenario – IP Ranges

Anyone cloaking today will have a good list.

You might find them at the bar.

“I Just Want A Sample.”

The file is still too big.

Fine.

I Want A Sample

The **sort** and **split** utilities do what you'd expect:

```
~$ sort -R combined.log | split -l 1048576
```

*"**Randomly sort** the lines in the combined.log. **split** the output of this command into multiple files (up to) 1048576 **lines** long."*

A **pipe** '**|**' takes the output of one command as the input of another.

shuf is easier, but not default OSX/GOW.

“I Just Want it in Excel.”

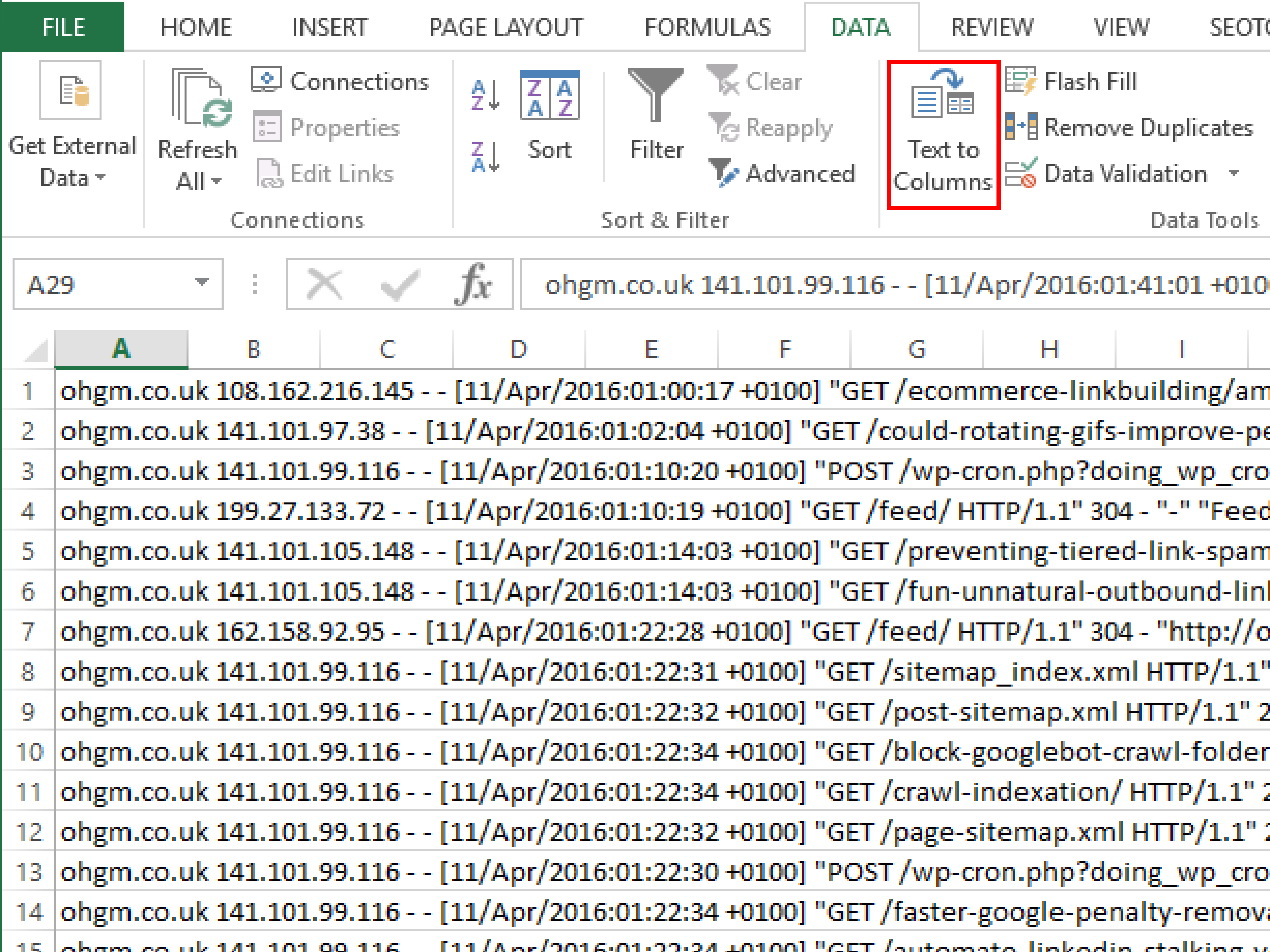
Fine.

I Just Want it in Excel.

Use **wc** to check it has fewer than 1,048,576 rows.

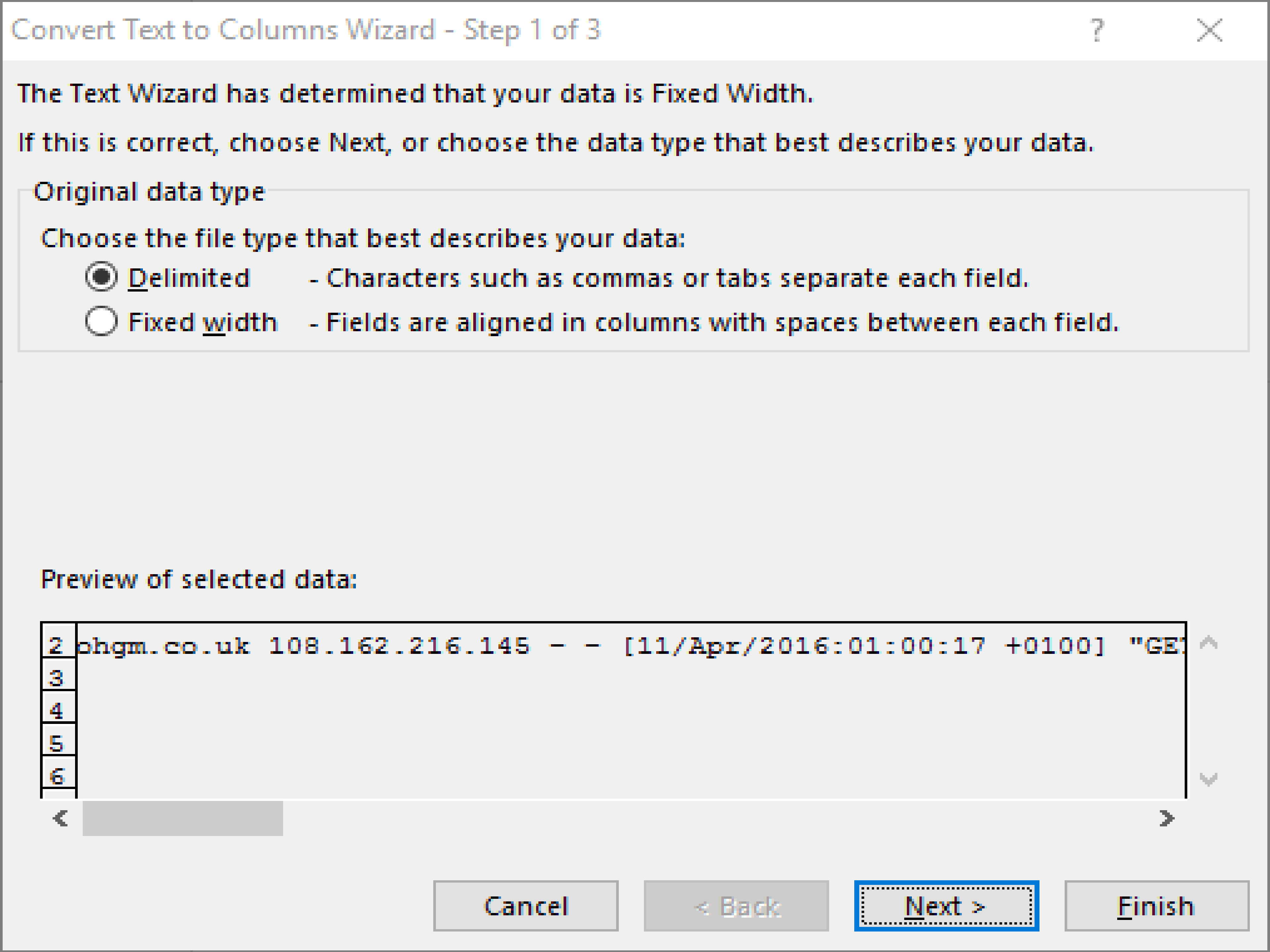
```
~$ wc -l sample.log
```

*“**Count** the number of **lines** in sample.log.”*



```
ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:41:01 +010
```

```
1  ohgm.co.uk 108.162.216.145 - - [11/Apr/2016:01:00:17 +0100] "GET /ecommerce-linkbuilding/am
2  ohgm.co.uk 141.101.97.38 - - [11/Apr/2016:01:02:04 +0100] "GET /could-rotating-gifs-improve-pe
3  ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:10:20 +0100] "POST /wp-cron.php?doing_wp_cro
4  ohgm.co.uk 199.27.133.72 - - [11/Apr/2016:01:10:19 +0100] "GET /feed/ HTTP/1.1" 304 - "-" "Feed
5  ohgm.co.uk 141.101.105.148 - - [11/Apr/2016:01:14:03 +0100] "GET /preventing-tiered-link-spam
6  ohgm.co.uk 141.101.105.148 - - [11/Apr/2016:01:14:03 +0100] "GET /fun-unnatural-outbound-linl
7  ohgm.co.uk 162.158.92.95 - - [11/Apr/2016:01:22:28 +0100] "GET /feed/ HTTP/1.1" 304 - "http://o
8  ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:31 +0100] "GET /sitemap_index.xml HTTP/1.1"
9  ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:32 +0100] "GET /post-sitemap.xml HTTP/1.1" 2
10 ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:34 +0100] "GET /block-googlebot-crawl-folder
11 ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:34 +0100] "GET /crawl-indexation/ HTTP/1.1" 2
12 ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:32 +0100] "GET /page-sitemap.xml HTTP/1.1" 2
13 ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:30 +0100] "POST /wp-cron.php?doing_wp_cro
14 ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:34 +0100] "GET /faster-google-penalty-remova
15 ohgm.co.uk 141.101.99.116 - - [11/Apr/2016:01:22:34 +0100] "GET /automate-linkedin-stalking-
```



Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Fixed Width.

If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

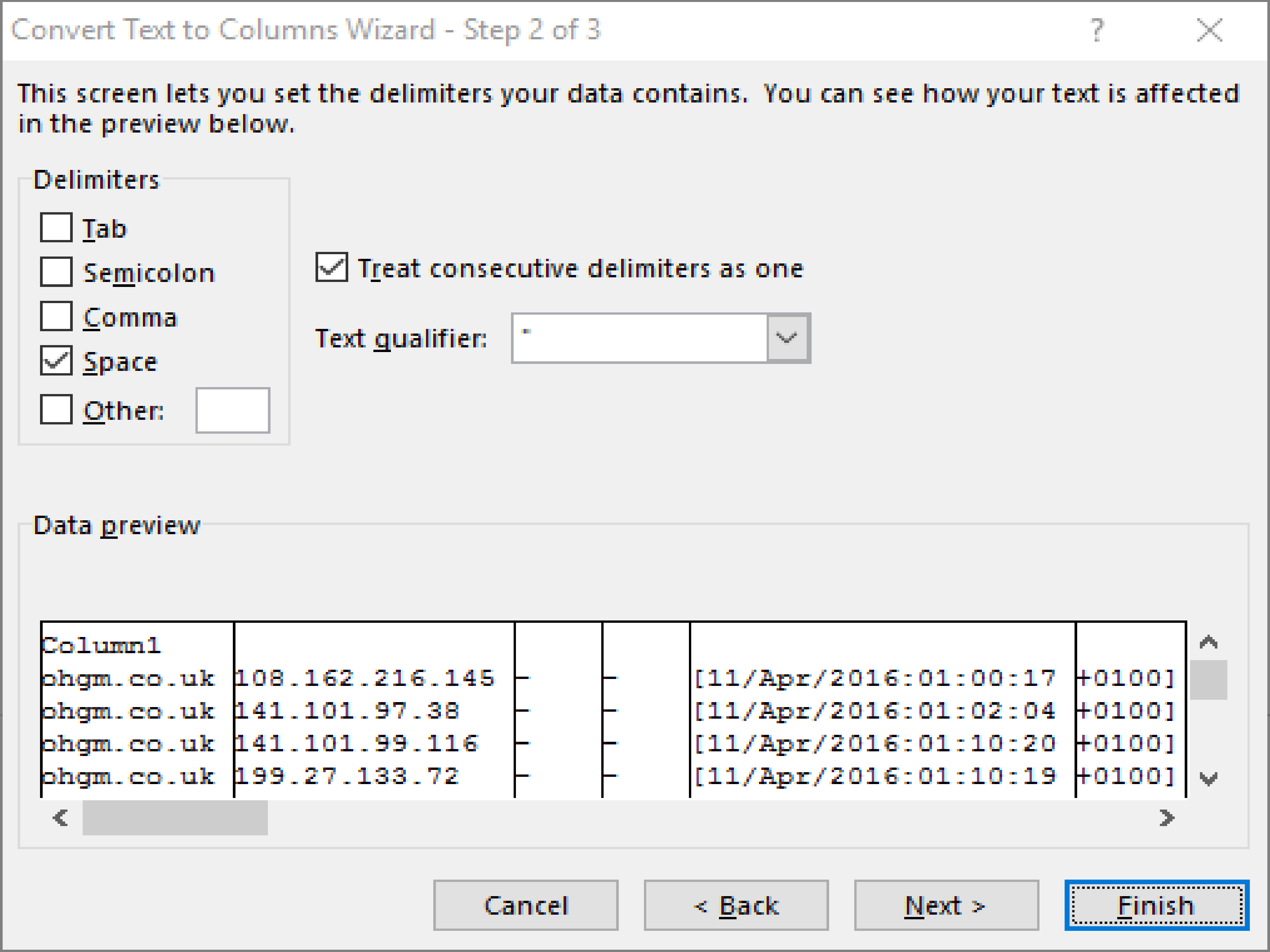
Choose the file type that best describes your data:

● Delimited - Characters such as commas or tabs separate each field.

○ Fixed width - Fields are aligned in columns with spaces between each field.

Preview of selected data:

```
2 ohgm.co.uk 108.162.216.145 - - [11/Apr/2016:01:00:17 +0100] "GE
3
4
5
6
```



This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

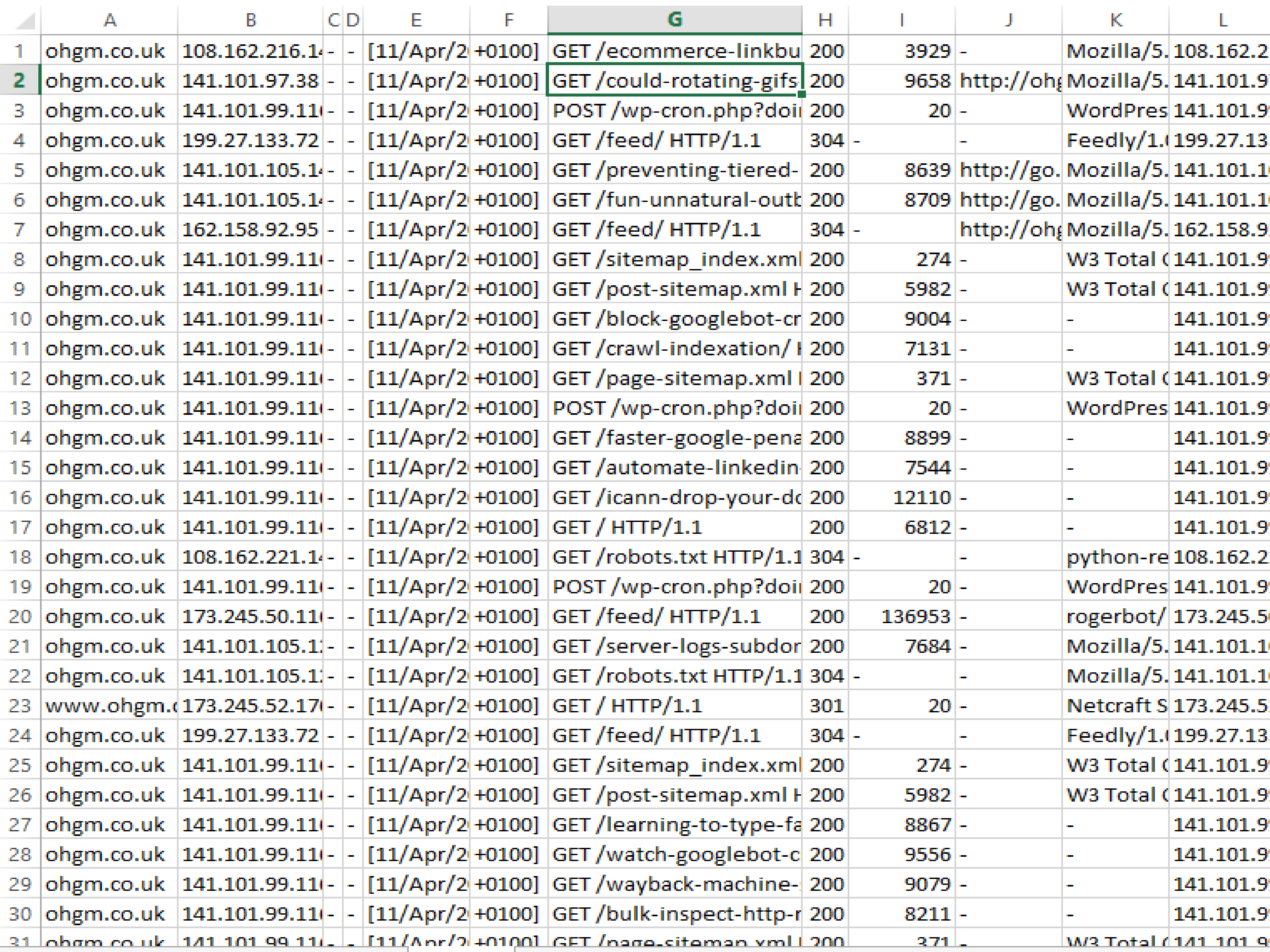
☐ Tab
☐ Semicolon
☐ Comma
☑ Space
☐ Other:

☑ Treat consecutive delimiters as one

Text qualifier: "

Data preview

Column1					
phgm.co.uk	108.162.216.145	-	-	[11/Apr/2016:01:00:17	+0100]
phgm.co.uk	141.101.97.38	-	-	[11/Apr/2016:01:02:04	+0100]
phgm.co.uk	141.101.99.116	-	-	[11/Apr/2016:01:10:20	+0100]
phgm.co.uk	199.27.133.72	-	-	[11/Apr/2016:01:10:19	+0100]



	A	B	C	D	E	F	G	H	I	J	K	L
1	ohgm.co.uk	108.162.216.1	-	-	[11/Apr/2	+0100]	GET /ecommerce-linkbu	200	3929	-	Mozilla/5.	108.162.2
2	ohgm.co.uk	141.101.97.38	-	-	[11/Apr/2	+0100]	GET /could-rotating-gifs.	200	9658	http://ohg	Mozilla/5.	141.101.9
3	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	POST /wp-cron.php?doi	200	20	-	WordPres	141.101.9
4	ohgm.co.uk	199.27.133.72	-	-	[11/Apr/2	+0100]	GET /feed/ HTTP/1.1	304	-	-	Feedly/1.	199.27.13
5	ohgm.co.uk	141.101.105.1	-	-	[11/Apr/2	+0100]	GET /preventing-tiered-	200	8639	http://go.	Mozilla/5.	141.101.1
6	ohgm.co.uk	141.101.105.1	-	-	[11/Apr/2	+0100]	GET /fun-unnatural-outb	200	8709	http://go.	Mozilla/5.	141.101.1
7	ohgm.co.uk	162.158.92.95	-	-	[11/Apr/2	+0100]	GET /feed/ HTTP/1.1	304	-	http://ohg	Mozilla/5.	162.158.9
8	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /sitemap_index.xml	200	274	-	W3 Total C	141.101.9
9	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /post-sitemap.xml H	200	5982	-	W3 Total C	141.101.9
10	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /block-googlebot-cr	200	9004	-	-	141.101.9
11	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /crawl-indexation/ H	200	7131	-	-	141.101.9
12	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /page-sitemap.xml H	200	371	-	W3 Total C	141.101.9
13	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	POST /wp-cron.php?doi	200	20	-	WordPres	141.101.9
14	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /faster-google-pena	200	8899	-	-	141.101.9
15	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /automate-linkedin-	200	7544	-	-	141.101.9
16	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /icann-drop-your-do	200	12110	-	-	141.101.9
17	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET / HTTP/1.1	200	6812	-	-	141.101.9
18	ohgm.co.uk	108.162.221.1	-	-	[11/Apr/2	+0100]	GET /robots.txt HTTP/1.1	304	-	-	python-re	108.162.2
19	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	POST /wp-cron.php?doi	200	20	-	WordPres	141.101.9
20	ohgm.co.uk	173.245.50.11	-	-	[11/Apr/2	+0100]	GET /feed/ HTTP/1.1	200	136953	-	rogerbot/	173.245.5
21	ohgm.co.uk	141.101.105.1	-	-	[11/Apr/2	+0100]	GET /server-logs-subdor	200	7684	-	Mozilla/5.	141.101.1
22	ohgm.co.uk	141.101.105.1	-	-	[11/Apr/2	+0100]	GET /robots.txt HTTP/1.1	304	-	-	Mozilla/5.	141.101.1
23	www.ohgm.c	173.245.52.17	-	-	[11/Apr/2	+0100]	GET / HTTP/1.1	301	20	-	Netcraft S	173.245.5
24	ohgm.co.uk	199.27.133.72	-	-	[11/Apr/2	+0100]	GET /feed/ HTTP/1.1	304	-	-	Feedly/1.	199.27.13
25	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /sitemap_index.xml	200	274	-	W3 Total C	141.101.9
26	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /post-sitemap.xml H	200	5982	-	W3 Total C	141.101.9
27	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /learning-to-type-fa	200	8867	-	-	141.101.9
28	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /watch-googlebot-c	200	9556	-	-	141.101.9
29	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /wayback-machine-	200	9079	-	-	141.101.9
30	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /bulk-inspect-http-r	200	8211	-	-	141.101.9
31	ohgm.co.uk	141.101.99.11	-	-	[11/Apr/2	+0100]	GET /page-sitemap.xml H	200	371	-	W3 Total C	141.101.9

The Title of The Talk Wasn’t a Lie And We Aren’t Going to Use Excel And Are Going to Answer Questions Just Using The Command Line I Hope That’s OK. Sorry.

Asking Useful Questions

Formulating Questions

Work a basic hypothesis. Decides what needs to be done if it is **true**, **false**, or **indeterminate** *before* you get the data.

"*Google is ignoring robots.txt*" may not be **action guiding**, whilst "*Googlebot is ignoring search console parameter restrictions*" just might be.

Formulating Questions

Some things just aren’t very useful to know.

Example Questions

How **deep** is Googlebot crawling?

Where is the **wasted crawl**? What proportion of requests are currently wasted?

Where is Googlebot **POST**ing?

What are the most popular **non-200/304** resources?

How many **unique resources** are being crawled?

Which is the more **popular** form of product page?

Which sitemap pages **aren’t** being crawled?

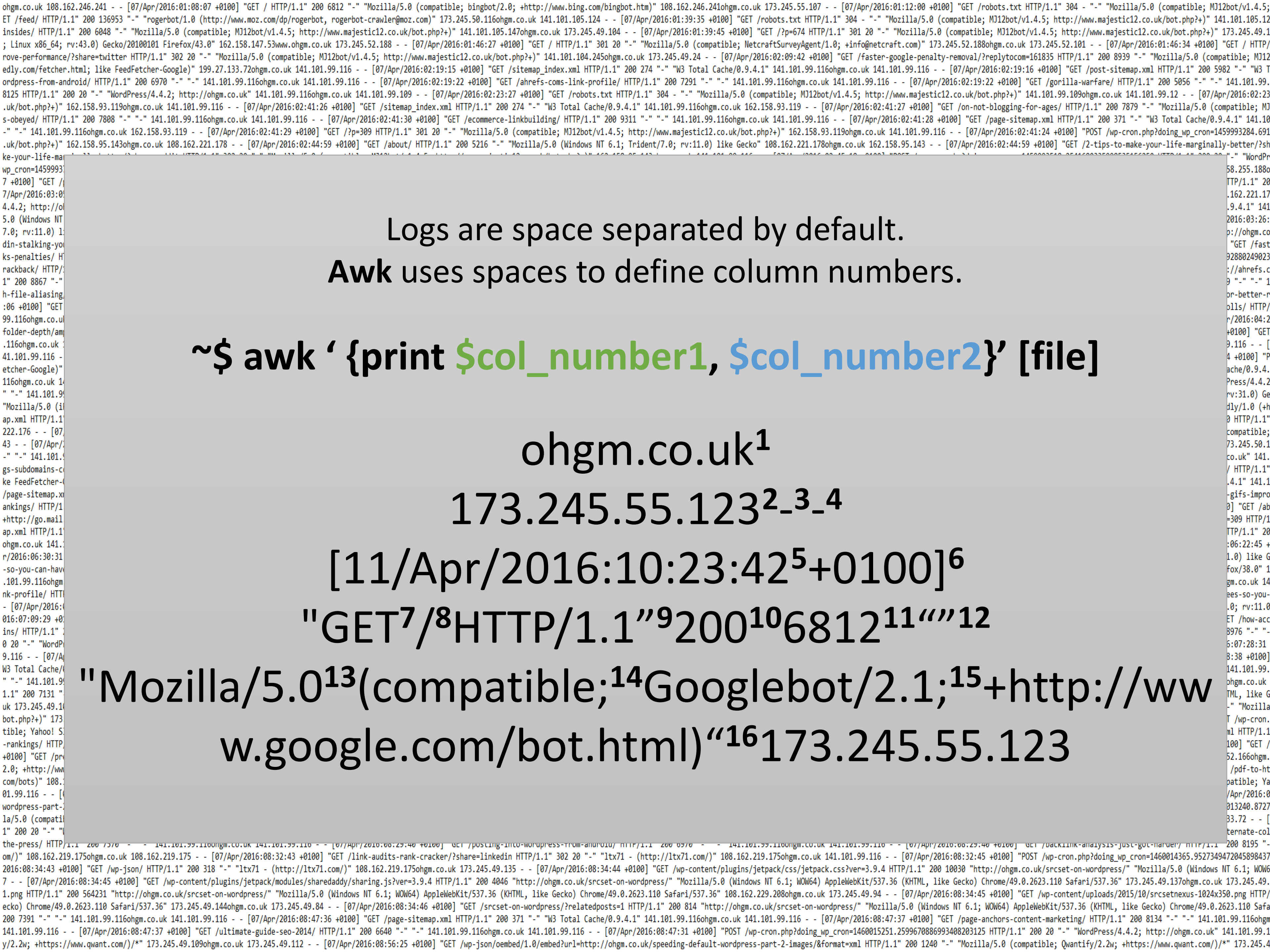
Always pivot with other data.

Getting Useful Answers

AWK

AWK is a programming language focused on text manipulation.

We are going to use it to print some columns from our log files. That’s it.



```
ohgm.co.uk 108.162.246.241 - - [07/Apr/2016:01:08:07 +0100] "GET / HTTP/1.1" 200 6812 "-" "Mozilla/5.0 (compatible; bingbot/2.0; +http://www.bing.com/bingbot.htm)" 108.162.246.241ohgm.co.uk 173.245.55.107 - - [07/Apr/2016:01:12:00 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5;
ET /feed/ HTTP/1.1" 200 136953 "-" "rogerbot/1.0 (http://www.moz.com/dp/rogerbot, rogerbot-crawler@moz.com)" 173.245.50.116ohgm.co.uk 141.101.105.124 - - [07/Apr/2016:01:39:35 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.105.12
insides/ HTTP/1.1" 200 6048 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.105.147ohgm.co.uk 173.245.49.104 - - [07/Apr/2016:01:39:45 +0100] "GET /?p=674 HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 173.245.49.1
; Linux x86_64; rv:43.0) Gecko/20100101 Firefox/43.0" 162.158.147.53www.ohgm.co.uk 173.245.52.188 - - [07/Apr/2016:01:46:27 +0100] "GET / HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; NetcraftSurveyAgent/1.0; +info@netcraft.com)" 173.245.52.188ohgm.co.uk 173.245.52.101 - - [07/Apr/2016:01:46:34 +0100] "GET / HTTP/
rove-performance/?share=twitter HTTP/1.1" 302 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.104.245ohgm.co.uk 173.245.49.24 - - [07/Apr/2016:02:09:42 +0100] "GET /faster-google-penalty-removal/?replytocom=161835 HTTP/1.1" 200 8939 "-" "Mozilla/5.0 (compatible; MJ12
edly.com/fetcher.html; like FeedFetcher-Google)" 199.27.133.72ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:15 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:16 +0100] "GET /post-sitemap.xml HTTP/1.1" 200 5982 "-" "W3 T
ordpress-from-android/ HTTP/1.1" 200 6970 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:22 +0100] "GET /ahrefs-coms-link-profile/ HTTP/1.1" 200 7291 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:19:22 +0100] "GET /gorilla-warfare/ HTTP/1.1" 200 5056 "-" "-" 141.101.99.
8125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.116ohgm.co.uk 141.101.99.109 - - [07/Apr/2016:02:23:27 +0100] "GET /robots.txt HTTP/1.1" 304 - "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 141.101.99.109ohgm.co.uk 141.101.99.12 - - [07/Apr/2016:02:23
.uk/bot.php?+)" 162.158.93.119ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:26 +0100] "GET /sitemap_index.xml HTTP/1.1" 200 274 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 162.158.93.119 - - [07/Apr/2016:02:41:27 +0100] "GET /on-not-blogging-for-ages/ HTTP/1.1" 200 7879 "-" "Mozilla/5.0 (compatible; MJ
s-obeyed/ HTTP/1.1" 200 7808 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:30 +0100] "GET /ecommerce-linkbuilding/ HTTP/1.1" 200 9311 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:28 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.10
-" "-" 141.101.99.116ohgm.co.uk 162.158.93.119 - - [07/Apr/2016:02:41:29 +0100] "GET /?p=309 HTTP/1.1" 301 20 "-" "Mozilla/5.0 (compatible; MJ12bot/v1.4.5; http://www.majestic12.co.uk/bot.php?+)" 162.158.93.119ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:02:41:24 +0100] "POST /wp-cron.php?doing_wp_cron=1459993284.691
.uk/bot.php?+)" 162.158.95.143ohgm.co.uk 108.162.221.178 - - [07/Apr/2016:02:44:59 +0100] "GET /about/ HTTP/1.1" 200 5216 "-" "Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko" 108.162.221.178ohgm.co.uk 162.158.95.143 - - [07/Apr/2016:02:44:59 +0100] "GET /2-tips-to-make-your-life-marginally-better/?sh
```

Logs are space separated by default.

Awk uses spaces to define column numbers.

~$ awk ‘ {print $col_number1, $col_number2}’ [file]

ohgm.co.uk^{1}

173.245.55.123^{2}-^{3}-^{4}

[11/Apr/2016:10:23:42^{5}+0100]^{6}

"GET^{7}/^{8}HTTP/1.1”^{9}200^{10}6812^{11}“”^{12}

"Mozilla/5.0^{13}(compatible;^{14}Googlebot/2.1;^{15}+http://www.google.com/bot.html)“^{16}173.245.55.123

```
the-press/ HTTP/1.1" 200 7370 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:29:40 +0100] "GET /posting-into-wordpress-from-android/ HTTP/1.1" 200 6970 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:29:40 +0100] "GET /backlink-analysis-just-got-harder/ HTTP/1.1" 200 8195 "-
om/)" 108.162.219.175ohgm.co.uk 108.162.219.175 - - [07/Apr/2016:08:32:43 +0100] "GET /link-audits-rank-cracker/?share=linkedin HTTP/1.1" 302 20 "-" "ltx71 - (http://ltx71.com/)" 108.162.219.175ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:32:45 +0100] "POST /wp-cron.php?doing_wp_cron=1460014365.9527349472045898437
2016:08:34:43 +0100] "GET /wp-json/ HTTP/1.1" 200 318 "-" "ltx71 - (http://ltx71.com/)" 108.162.219.175ohgm.co.uk 173.245.49.135 - - [07/Apr/2016:08:34:44 +0100] "GET /wp-content/plugins/jetpack/css/jetpack.css?ver=3.9.4 HTTP/1.1" 200 10030 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW6
7 - - [07/Apr/2016:08:34:45 +0100] "GET /wp-content/plugins/jetpack/modules/sharedaddy/sharing.js?ver=3.9.4 HTTP/1.1" 200 4046 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safari/537.36" 173.245.49.137ohgm.co.uk 173.245.49.
1.png HTTP/1.1" 200 564231 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safari/537.36" 108.162.229.208ohgm.co.uk 173.245.49.94 - - [07/Apr/2016:08:34:45 +0100] "GET /wp-content/uploads/2015/10/srcsetnexus-1024x350.png HTTP/
ecko) Chrome/49.0.2623.110 Safari/537.36" 173.245.49.144ohgm.co.uk 173.245.49.84 - - [07/Apr/2016:08:34:46 +0100] "GET /srcset-on-wordpress/?relatedposts=1 HTTP/1.1" 200 814 "http://ohgm.co.uk/srcset-on-wordpress/" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.110 Safa
200 7391 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:36 +0100] "GET /page-sitemap.xml HTTP/1.1" 200 371 "-" "W3 Total Cache/0.9.4.1" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:37 +0100] "GET /page-anchors-content-marketing/ HTTP/1.1" 200 8134 "-" "-" 141.101.99.116ohgm
141.101.99.116 - - [07/Apr/2016:08:47:37 +0100] "GET /ultimate-guide-seo-2014/ HTTP/1.1" 200 6640 "-" "-" 141.101.99.116ohgm.co.uk 141.101.99.116 - - [07/Apr/2016:08:47:31 +0100] "POST /wp-cron.php?doing_wp_cron=1460015251.2599670886993408203125 HTTP/1.1" 200 20 "-" "WordPress/4.4.2; http://ohgm.co.uk" 141.101.99.1
y/2.2w; +https://www.qwant.com/)/*" 173.245.49.109ohgm.co.uk 173.245.49.112 - - [07/Apr/2016:08:56:25 +0100] "GET /wp-json/oembed/1.0/embed?url=http://ohgm.co.uk/speeding-default-wordpress-part-2-images/&format=xml HTTP/1.1" 200 1240 "-" "Mozilla/5.0 (compatible; Qwantify/2.2w; +https://www.qwant.com/)/*" 173.245.4
```

AWK

```
~$ awk ‘{print $8, $10}’ Googlebot.log >> Gbot_responses.txt
```

*“Output the **requested resource** and **server response** of Googlebot.log to Gbot_responses.txt.”*

```
/                       200
/robots.txt             304
/robots.txt             500
/amazing-blog-post      200
/forgotten-blog-post    404
/forbidden-blog-post    403
/                       200
```

Tailor the command to the access log format in use.

uniq

uniq takes text as an input and returns **unique** lines.

uniq -c returns these lines prefixed with a count.

uniq -d returns only repeated lines.

uniq -u returns only non-repeated lines.

AWK

Like **grep**, **awk** also matches patterns, using **/pattern/**.

```
~$ awk ‘/bingbot/ {print $10}’ combined.log | uniq -c
```

*“Look for lines containing **bingbot** in the unfiltered logs and print their **server response codes**. **Deduplicate** these and **return a summary**.”*

```
216663 -     302
109232 -     200
18395 -      301
2568 -       404
2147 -       304
274 -        500
261 -        403
```

Example Use: Site Migrations

Ultimate Guide to Site Migrations

Get a big list of old URLs.

301 redirect them once to the right places.

Make sure they get crawled.

Site Migrations

"We want a list of all URLs requested by Googlebot in our pre-migration dataset, sorted by popularity (number of requests)."

e.g.

/	49587
/index.html	25169
/robots.txt	23334
/home	19417

Site Migrations

```
~$ awk ‘/Googlebot/ {print $7}’ combined.log |
uniq -c | sort -nr >> unique_requests.txt
```

“Take all access log requests, and filter to ***Googlebot****.*

Extract and output the ***requested resources****.*

Deduplicate these and ***return a summary****.*

Sort *these by number in descending order.”*

Site Migrations – Encouraging Crawl

“We want our migration to switch as quickly as possible.”

Get the list of redirects (URI stems) you want Google to crawl into a file.

grep can use this file as the match criteria (lines matching this OR this OR this).

Site Migrations – Encouraging Crawl

We want the URLs Google has not yet crawled.

```
~$ grep -f wishlist.txt postmigration.log | awk ‘/Googlebot/ {print $8}’ | uniq >> wishlist-hits.txt
```

*“Filter the post-migration log for **lines that match wishlist.txt**. Return the **resources requested** by Googlebot. **Deduplicate** and save.”*

Site Migrations – Encouraging Crawl

```
~$ cat wishlist-hits.txt wishlist.txt | uniq -u >> uncrawled.txt
```

*"Read the contents of **both files**. Save wishlist entries that **don't** appear in the access logs."*

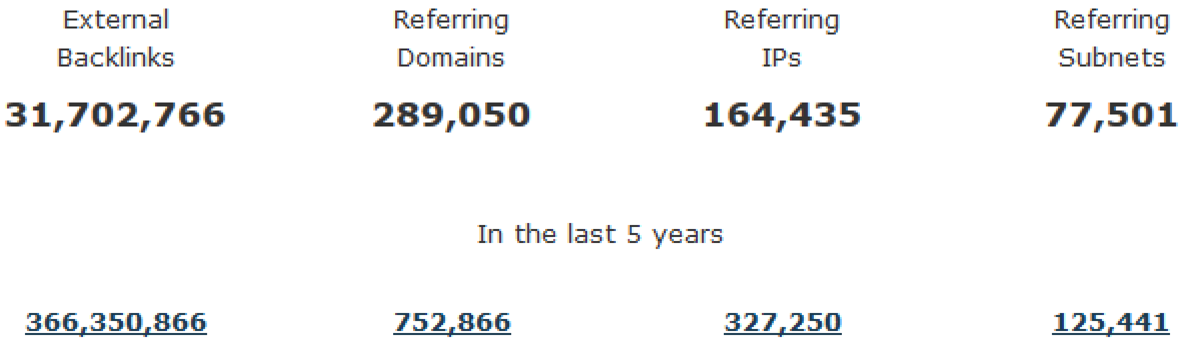
Tip: use an indexing service like linklicious to encourage crawling the uncrawled.

Taking This Further

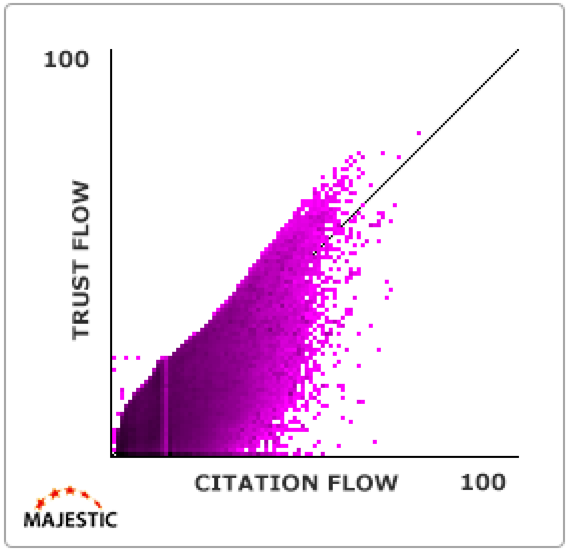
Keep Learning Unix Utilities. Learn SQL.

Also

These Techniques Apply to Other SEO Activities.



External Backlinks	Referring Domains	Referring IPs	Referring Subnets
31,702,766	289,050	164,435	77,501
In the last 5 years			
366,350,866	752,866	327,250	125,441



Enterprise Link Audits.
Enterprise Keyword Research.
Enterprise Spamming.

Thanks.

Oliver Mason
Technical SEO Consultant

Twitter: @ohgm
Email: ohgm@ohgm.co.uk

Resources

GOW:	https://github.com/bmatzelle/gow
Cygwin:	http://cygwin.com/install.html

Command Line Crash Course:

http://cli.learncodethehardway.org/book/

Shameless links to my own stuff:

http://ohgm.co.uk/filter-server-logs-to-googlebot/
http://ohgm.co.uk/watch-googlebot-crawling/
http://ohgm.co.uk/preserve-link-equity-with-file-aliasing/
http://ohgm.co.uk/wayback-machine-seo/

Tools Used in this Talk

grep

sort

split

shuf

find

uniq

awk

wc

cowsay