

RockyWeb /outgoing/Z/ Full NGINX Access-Log Analysis

Generated 2026-07-12 07:42:24 local runtime time

Executive summary

The pipeline parsed 361,642,469 RockyWeb custom nginx access-log records across 141 downloaded log files. The parsed records account for 2,381,501.08 GB of response-body bytes. The parsed server-local timestamp range is 2026-02-17 07:48:33 to 2026-07-08 09:57:34. The analysis summarizes downloads by data type, project, source IP/network, user-agent class, status code, and day/week/month.

Duplicate-line suppression was not used. Event identity is source file plus line number, and exact duplicate-looking rows are retained as observed server events.

Active-log overlap filtering excluded 13 parseable rows from the primary rollup based on timestamp coverage already present in rotated logs.

Country and operator attribution uses routing/registration evidence. Team Cymru maps IP addresses to BGP origin ASN, prefix, and country code; RDAP provides WHOIS-style registration data for Internet number resources. These fields identify network operators and egress countries, not necessarily the physical location or identity of a human end user.

Full-log analysis workflow

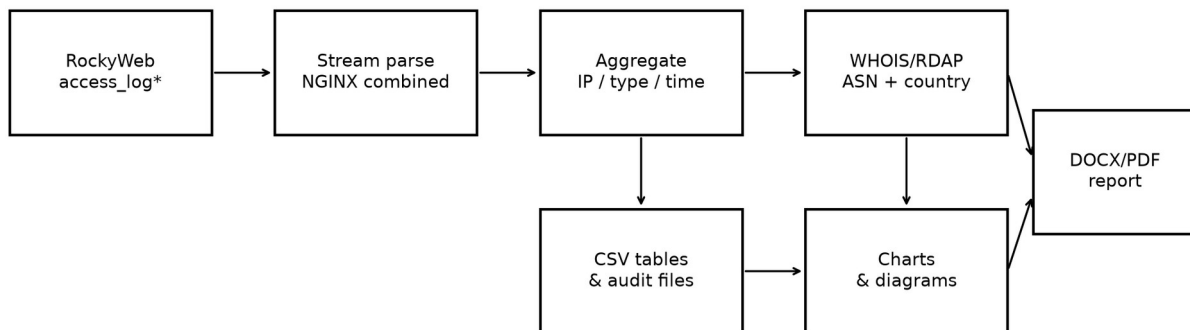


Figure 1. Full-log analysis workflow.

Methodology

The parser treats the files as a custom RockyWeb nginx HTTP access-log format: quoted remote address, remote user, local timestamp, quoted request, status, response-body bytes, then four quoted custom fields. The actual user-agent is the third quoted field after the byte count, not the second combined-log position. Each physical source_file + source_line_number is treated as an observed server event; exact duplicate-looking lines are retained

because range/parallel downloaders such as aria2 can legitimately produce visually identical 206 responses in a one-second-resolution log.

The output byte count is the logged response body byte field. It is a server log measurement of response payload bytes, not necessarily total wire bytes. Status codes, byte counts, user-agent strings, automated downloader indicators, and referrer hosts are retained in summary form.

Source analysis is performed at three levels: individual IP address, normalized source network (/24 for public IPv4 and /48 for public IPv6), and ASN/operator after WHOIS/RDAP enrichment. Private and non-public addresses are explicitly labeled rather than geolocated.

Download volume over time

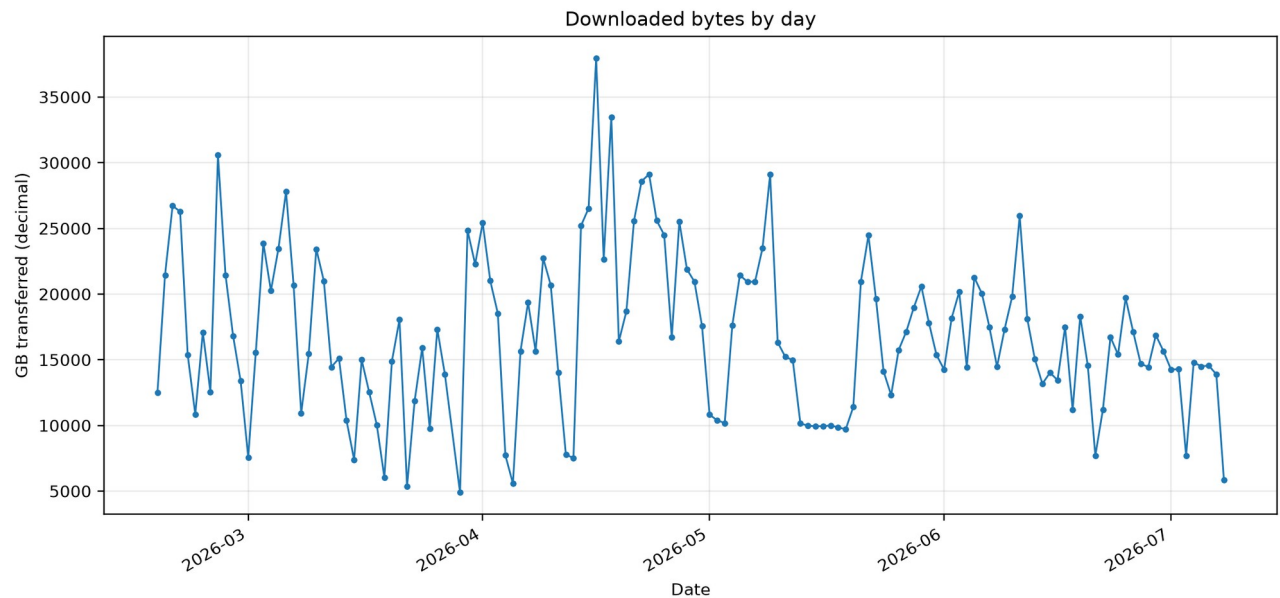


Figure 2. Parsed response-body bytes by day.

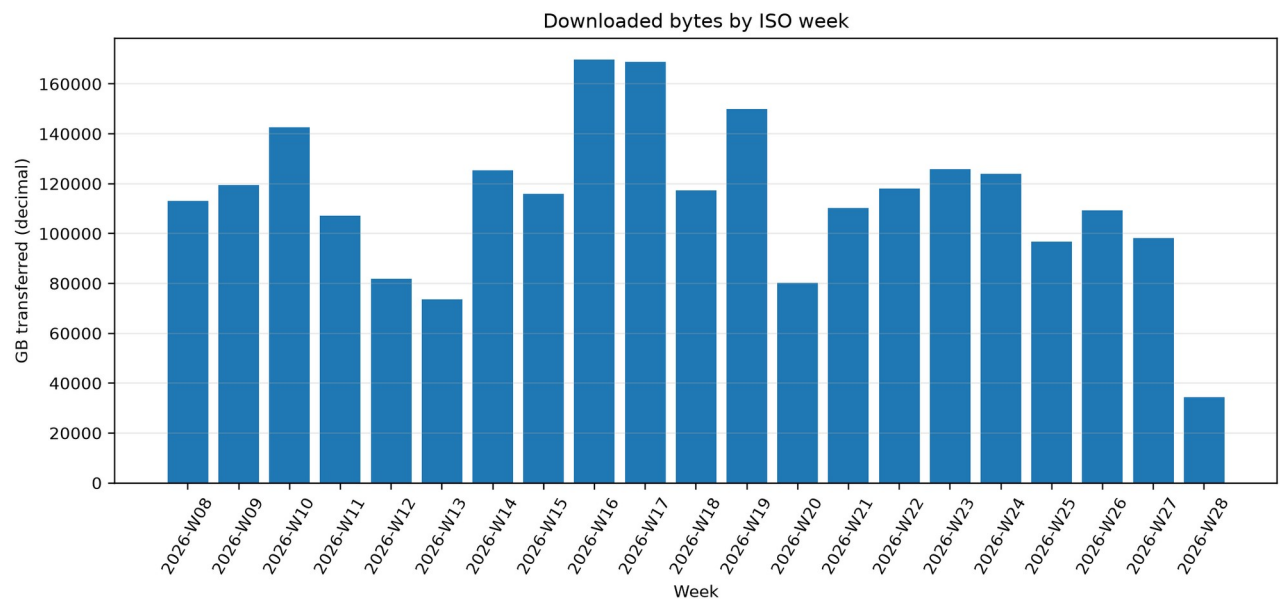


Figure 3. Parsed response-body bytes by ISO week.

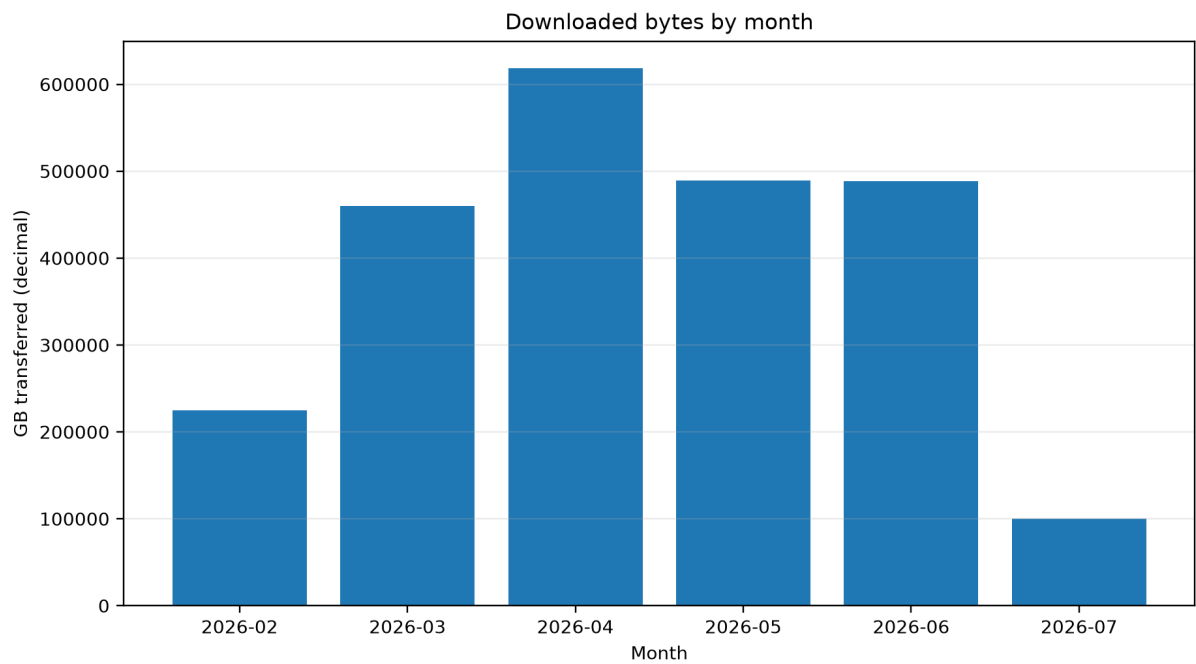


Figure 4. Parsed response-body bytes by month.

Highest-volume parsed days:

Date	Requests	GB	Unique IPs
2026-04-16	6,841,231	37,955.57	327,210
2026-04-18	6,198,037	33,480.58	153,940
2026-02-25	949,885	30,584.16	93,151
2026-04-23	1,393,862	29,128.66	56,071
2026-05-09	645,798	29,098.68	160,415
2026-04-22	1,352,229	28,583.37	14,150
2026-03-06	1,223,537	27,803.86	256,213
2026-02-19	1,167,100	26,723.45	27,872
2026-04-15	8,789,231	26,528.76	52,012
2026-02-20	1,267,578	26,279.80	32,889

What was downloaded

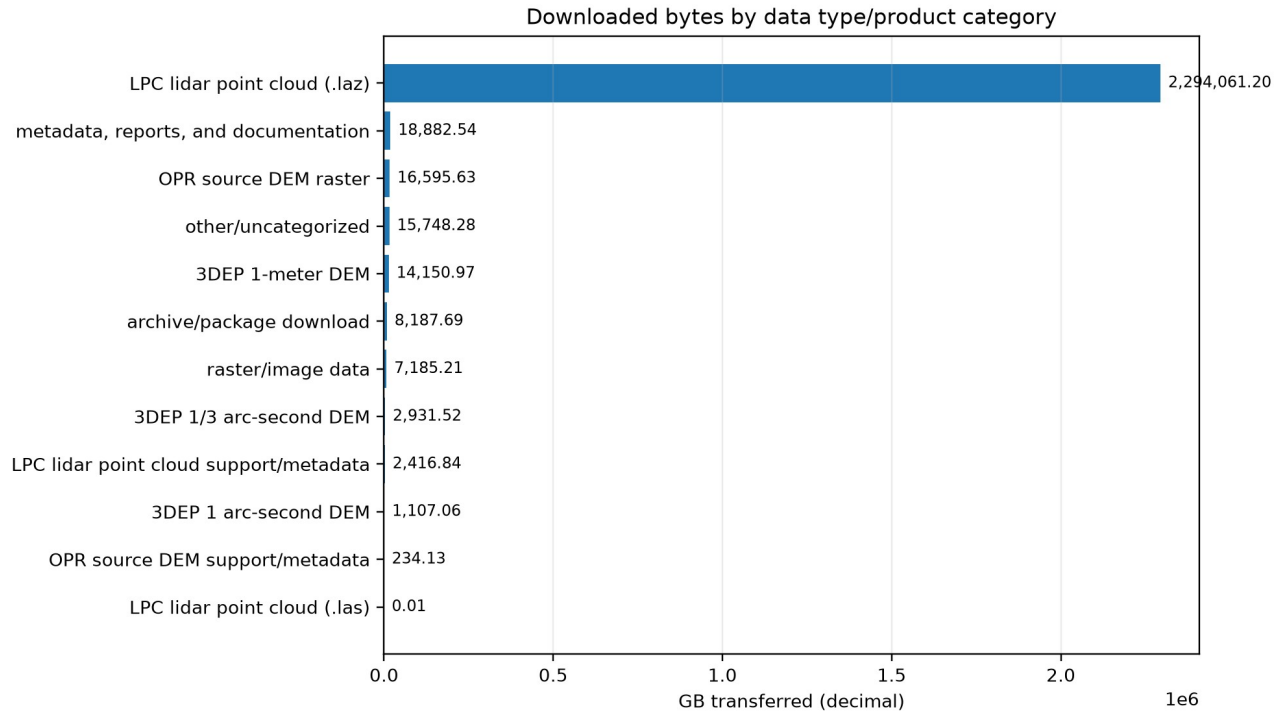


Figure 5. Response-body bytes by product/data category.

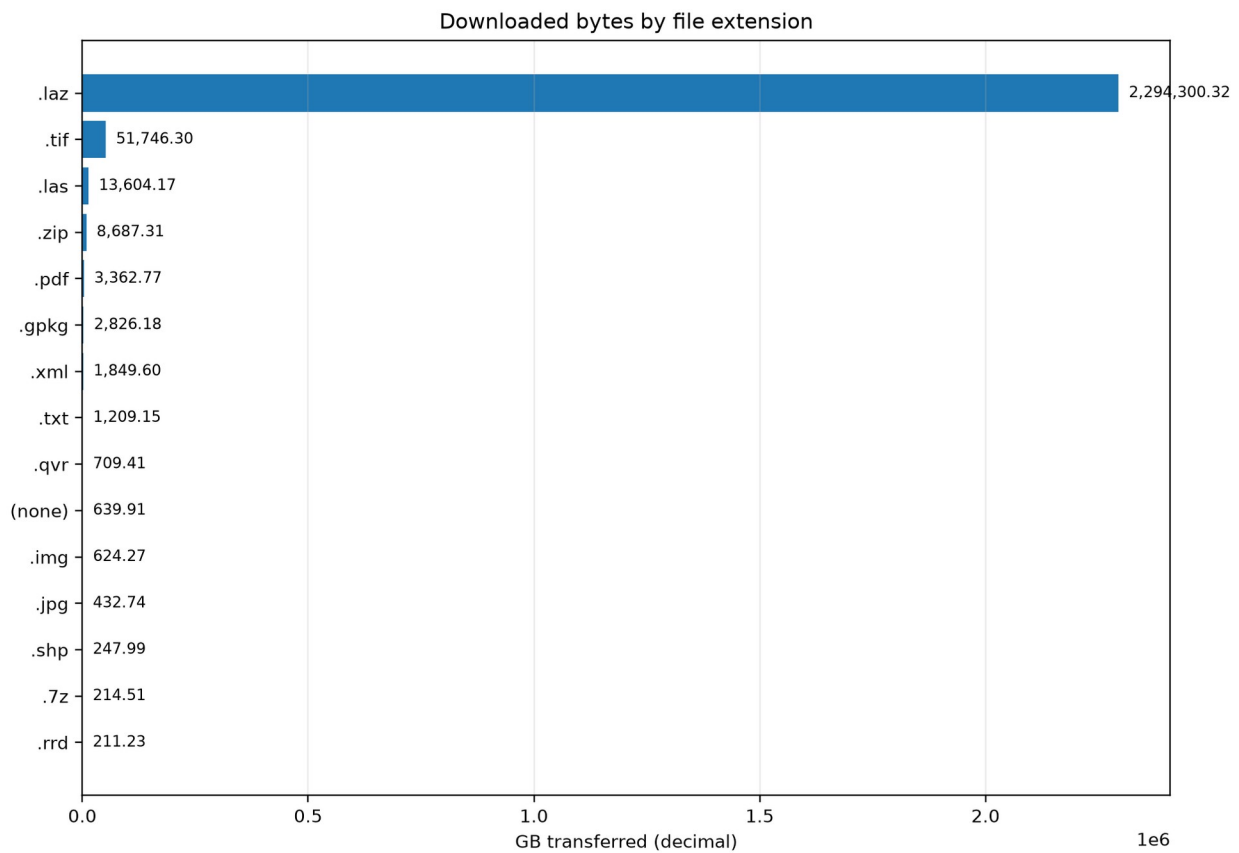


Figure 6. Response-body bytes by file extension.

Product/data category	Requests	GB
LPC lidar point cloud (.laz)	133,505,331	2,294,061.20
metadata, reports, and documentation	13,406,997	18,882.54
OPR source DEM raster	3,127,437	16,595.63
other/uncategorized	16,484,118	15,748.28
3DEP 1-meter DEM	2,620,076	14,150.97
archive/package download	419,212	8,187.69
raster/image data	547,400	7,185.21
3DEP 1/3 arc-second DEM	250,622	2,931.52
LPC lidar point cloud support/metadata	179,362,242	2,416.84
3DEP 1 arc-second DEM	406,875	1,107.06
OPR source DEM support/metadata	11,502,508	234.13
LPC lidar point cloud (.las)	9,651	0.01

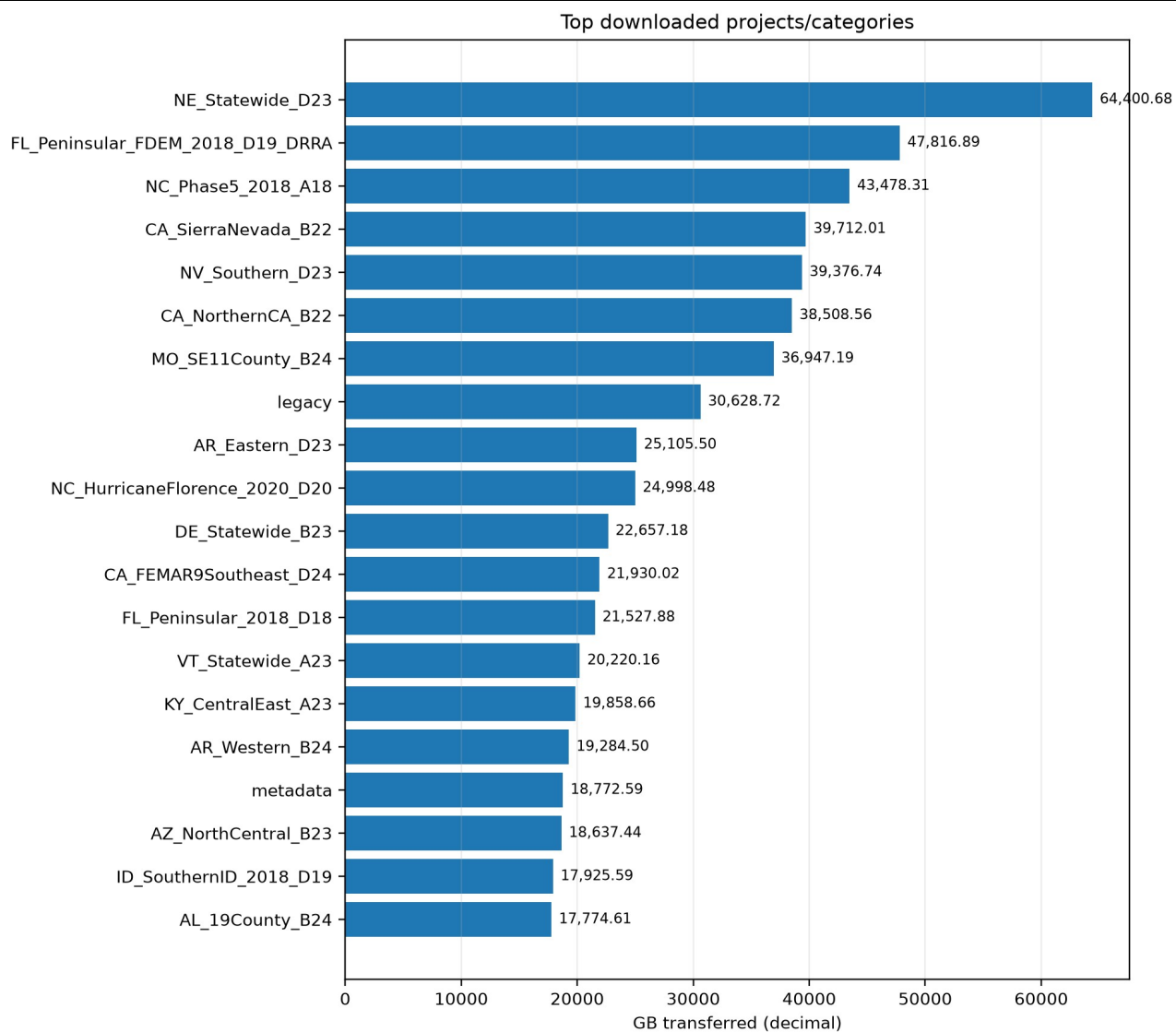


Figure 7. Top projects/categories by response-body bytes.

Who downloaded it and from where

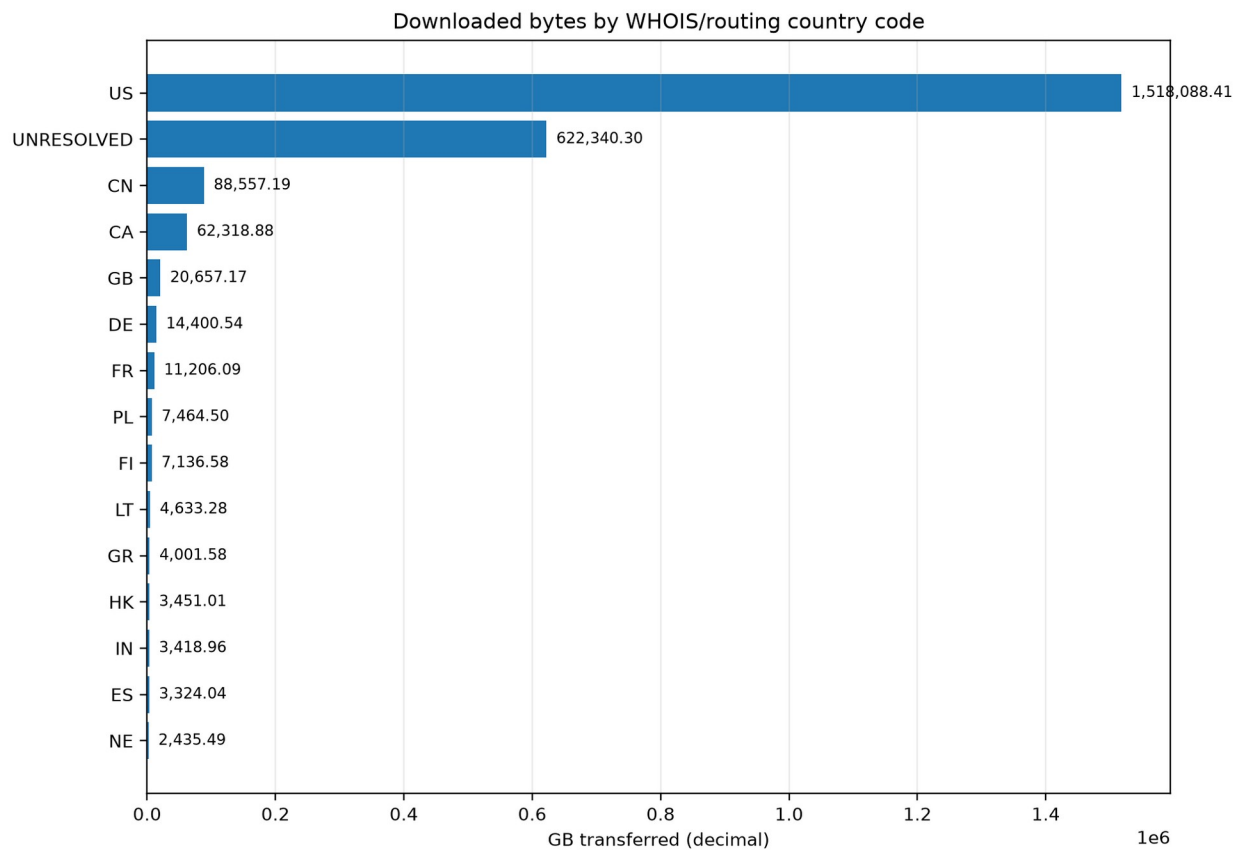


Figure 8. Response-body bytes by WHOIS/routing country code.

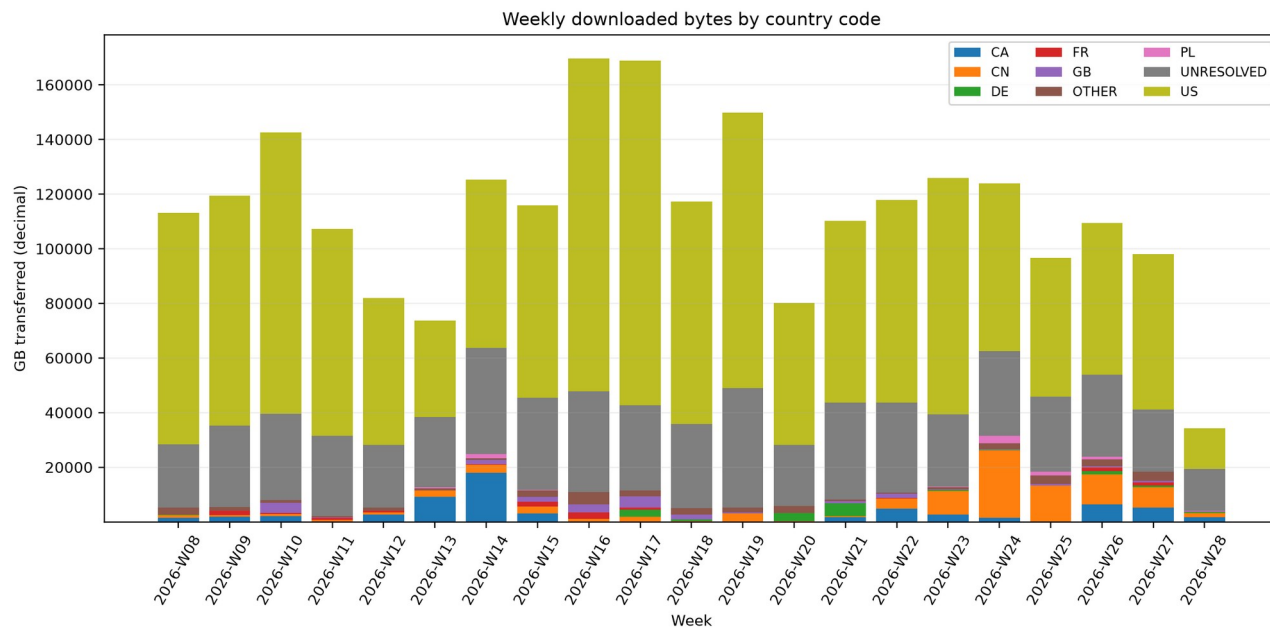


Figure 9. Weekly response-body bytes by top country codes.

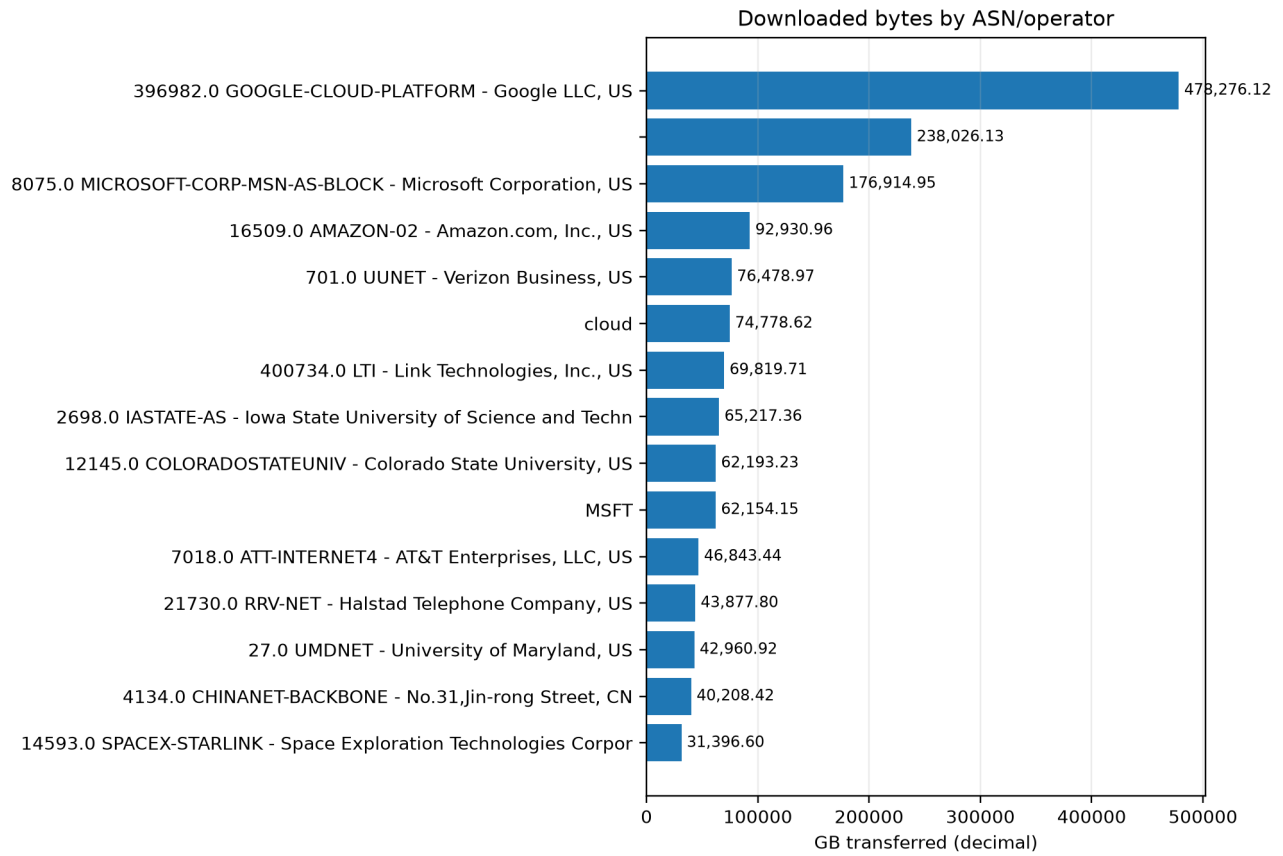


Figure 10. Response-body bytes by ASN/operator.

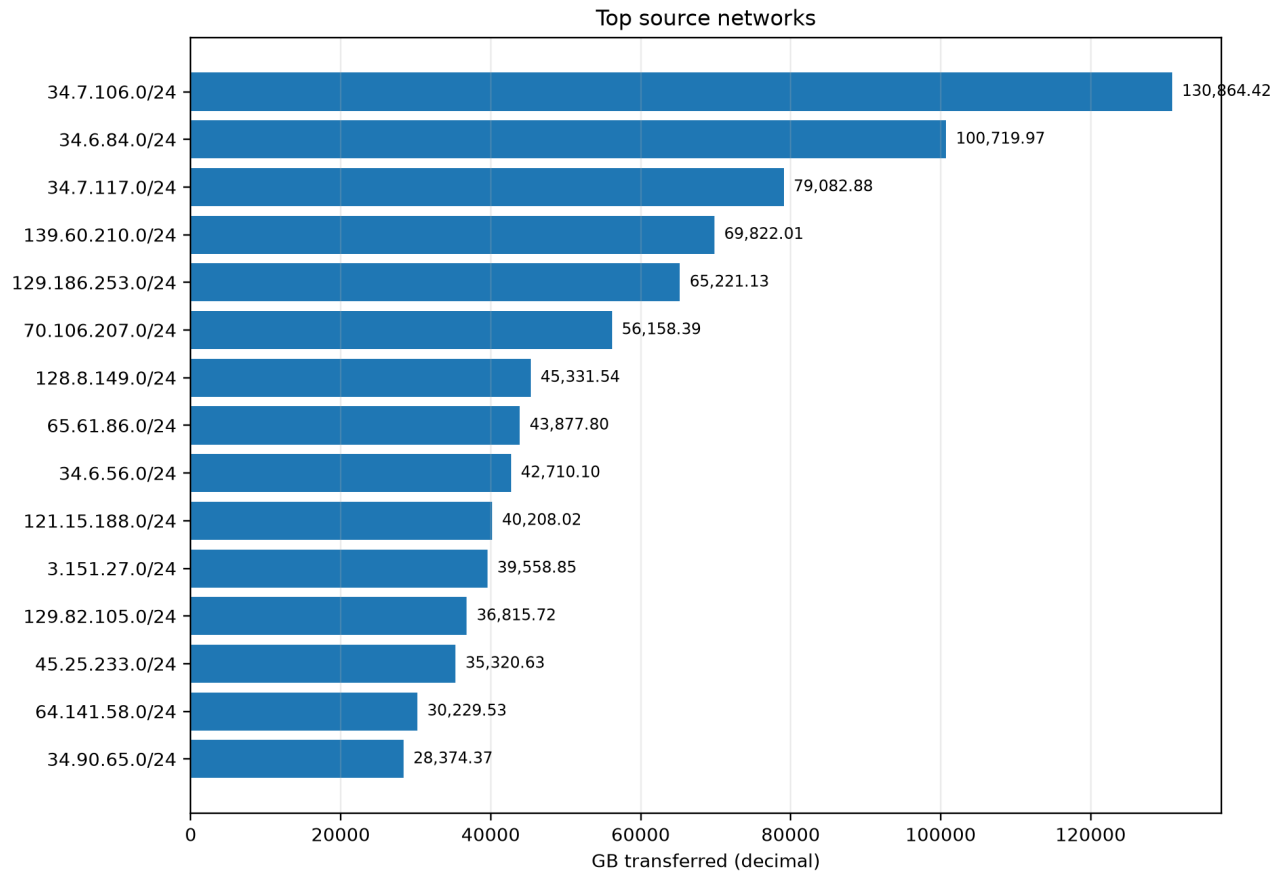


Figure 11. Top source networks by response-body bytes.

Country code	Requests	GB	Unique IPs
US	111,669,641	1,518,088.41	12,844
UNRESOLVED	201,812,062	622,340.30	8,160,939
CN	7,981,677	88,557.19	1,330
CA	1,722,093	62,318.88	838
GB	1,741,827	20,657.17	2,641
DE	1,991,916	14,400.54	1,126
FR	74,803	11,206.09	1,649
PL	312,312	7,464.50	345
FI	226,934	7,136.58	30
LT	242,186	4,633.28	248
GR	79,734	4,001.58	178
HK	452,943	3,451.01	402
IN	1,077,337	3,418.96	4,618
ES	46,509	3,324.04	1,272
NE	181,876	2,435.49	2
ASN	ASN/operator	CC	GB
396,982.00	GOOGLE-CLOUD-PLATFORM - Google LLC, US	US	478,276.12
nan	nan	UNRESOLVED	238,026.13
8,075.00	MICROSOFT-CORP-MSN-AS-BLOCK - Microsoft Corporation, US	US	176,914.95
16,509.00	AMAZON-02 - Amazon.com, Inc., US	US	92,930.96
701.00	UUNET - Verizon Business, US	US	76,478.97
nan	cloud	US	74,778.62
400,734.00	LTI - Link Technologies, Inc., US	US	69,819.71
2,698.00	IASTATE-AS - Iowa State University of Science and Technology, US	US	65,217.36
12,145.00	COLORADOSTATEUNIV - Colorado State University, US	US	62,193.23
nan	MSFT	UNRESOLVED	62,154.15
7,018.00	ATT-INTERNET4 - AT&T Enterprises, LLC, US	US	46,843.44
21,730.00	RRV-NET - Halstad Telephone Company, US	US	43,877.80
27.00	UMDNET - University of Maryland, US	US	42,960.92
4,134.00	CHINANET-BACKBONE - No.31,Jin-rong Street, CN	CN	40,208.42
14,593.00	SPACEX-STARLINK - Space Exploration Technologies Corporation, US	US	31,396.60

How the downloads were performed

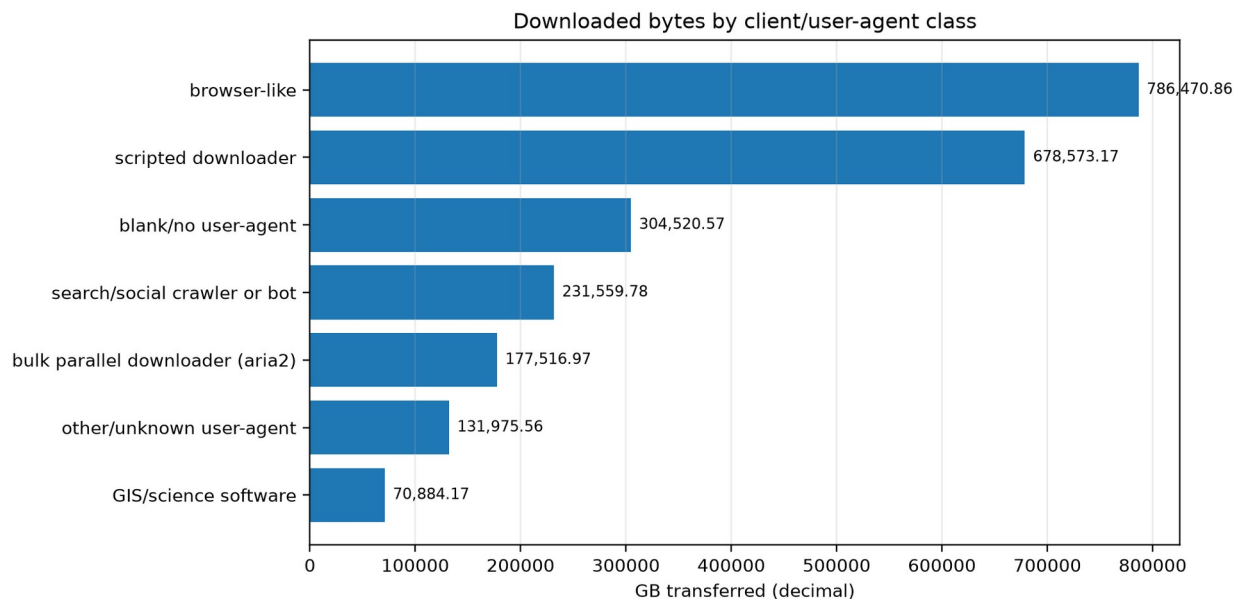


Figure 12. Response-body bytes by user-agent/client class.

Client class	Requests	GB
browser-like	51,405,975	786,470.86
scripted downloader	167,007,867	678,573.17
blank/no user-agent	19,141,645	304,520.57
search/social crawler or bot	47,708,635	231,559.78
bulk parallel downloader (aria2)	19,151,442	177,516.97
other/unknown user-agent	51,791,823	131,975.56
GIS/science software	5,435,082	70,884.17

HTTP status-code distribution:

Status	Requests	GB
200.00	269,535,268.00	2,222,736.12
206.00	84,471,236.00	158,762.51
404.00	6,714,238.00	2.15
403.00	354,397.00	0.18
416.00	240,162.00	0.05
405.00	80,139.00	0.04
301.00	154,265.00	0.02
400.00	74,003.00	0.01
414.00	455.00	0.00
501.00	111.00	0.00
412.00	8.00	0.00
413.00	8.00	0.00

Audit files and reproducibility

The output folder retains CSV tables for all report charts. The raw logs are intentionally not included in the compact report bundle; keep them locally if audit or reprocessing is required. The most important audit tables are `directory_manifest.csv`, `parse_summary.json`, `unmatched_examples.csv`, `source_ip_summary.csv`, `source_network_summary.csv`, `unique_public_ips.csv`, `cymru_ip_enrichment.csv`, and `rdap_network_enrichment.csv`.

Limitations

Log-derived byte counts depend on the configured nginx log format and the semantics of the logged byte variable. The byte field represented here is interpreted as response-body bytes, not total TCP/TLS transfer size. Confirm the server `log_format` directive if the site uses a custom byte variable.

WHOIS/RDAP and BGP-origin country data should be interpreted as registration/routing evidence. Cloud regions, VPN/proxy egress, institutional gateways, NAT, crawlers, and scripted download infrastructure can make the apparent country or organization differ from the final user or beneficiary of a download.

The active `access_log` may overlap with rotated daily logs depending on server log rotation. This parser does not suppress exact duplicate-looking rows. When `--include-active` is used with the default `--active-overlap-policy exclude-covered`, active-log rows at or before the latest rotated-log timestamp are excluded from the primary rollup; use `--active-overlap-policy include` to count all active-log rows.

References

- [1] U.S. Geological Survey RockyWeb directory listing, Index of `/outgoing/Z/`, <https://rockyweb.usgs.gov/outgoing/Z/>. Accessed 2026-07-08.
- [2] NGINX, Module `ngx_http_log_module`, official documentation for `access_log` and `log_format`, https://nginx.org/en/docs/http/ngx_http_log_module.html.
- [3] ARIN, Whois/Registration Data Access Protocol (RDAP), <https://www.arin.net/resources/registry/whois/rdap/>.
- [4] Team Cymru, IP to ASN Mapping Service, <https://www.team-cymru.com/ip-asn-mapping>.
- [5] Rekhter et al., RFC 1918: Address Allocation for Private Internets, Internet Engineering Task Force, <https://datatracker.ietf.org/doc/html/rfc1918>.
- [6] U.S. Geological Survey, Lidar Point Cloud - USGS National Map 3DEP Downloadable Data Collection, <https://data.usgs.gov/datacatalog/data/USGS%3Ab7e353d2-325f-4fc6-8d95-01254705638a>.
- [7] U.S. Geological Survey, Original Product Resolution Source Digital Elevation Models - USGS National Map 3DEP Downloadable Data Collection, <https://data.usgs.gov/datacatalog/data/USGS%3A1004ada7-1e9f-4fcf-adfb-677868b7317c>.