

Inside the Global Data Cloud: What Our Data Centers Really Store

Author: Andris Gailītis © AG

Technology: Local Ollama (qwen2.5-72B / qwen3.6-35B), n8n, SearXNG + Claude Code workflow management

What Actually Fills the World's Data Centers?

Percentage Breakdown

Category	Share Estimate	Key Sources
Video	30-45%	https://zipdo.co/big-data-statistics , https://www.optimixed.com/amount-of-data-created-daily-2026
Photos	10-15%	https://photutorial.com/photos-statistics , https://lapseoftheshutter.com/photography-statistics
Chats + Text	15-20%	https://expandedramblings.com/index.php/whatsapp-statistics , https://gitnux.org/mobile-messaging-statistics
AI	10-15%	https://www.gminsights.com/industry-analysis/ai-training-dataset-market , https://axis-intelligence.com/ai-data-center-statistics/
Enterprise + Backups	20-25%	https://gitnux.org/data-storage-industry-statistics , https://www.fortunebusinessinsights.com/data-storage-market-102991
Cold/Dark Data	10-15%	https://worldmetrics.org/dark-data-statistics , https://datastorage.com/articles/hot-warm-cold-archive-the-data-tiering-strategy-that-cuts-storage-bills-by-half
Other	5-10%	https://zipdo.co/big-data-statistics , https://www.demandsage.com/big-data-statistics

Key Facts

- **90% of the world's data has been created in the last two years.** (<https://www.demandsage.com/big-data-statistics>)
- **Video content constitutes about 82% of internet data traffic.** (<https://www.optimixed.com/amount-of-data-created-daily-2026>)
- **Approximately 1.95 trillion photos will be taken in 2026, equating to about 5.3 billion daily.** (<https://photutorial.com/photos-statistics>)
- **WhatsApp users send more than 100 billion messages per day.** (<https://expandedramblings.com/index.php/whatsapp-statistics>)
- **The global AI training dataset market is projected to grow at a CAGR of 20.5% between 2025 and 2034.** (<https://www.gminsights.com/industry-analysis/ai-training-dataset-market>)
- **Enterprise storage utilization averages 55%.** (<https://gitnux.org/data-storage-industry-statistics>)
- **90% of enterprise data is unstructured.** (<https://shelf.io/wp-content/uploads/2024/10/idc-unstructured-data-report-sept-2024.pdf>)
- **Dark data drives 30% of data center inefficiencies.** (<https://worldmetrics.org/dark-data-statistics>)
- **The Internet Archive holds approximately 624 billion archived items.** (<https://archive.org/>)
- **Modern games now exceed 100GB install sizes due to textures, live services, and parity demands.** (<https://www.techtimes.com/articles/313751/20260105/why-modern-games-are-getting-bigger-explaining-massive-install-sizes-storage-demands.htm>)

MYTHS vs DATA

- **Myth:** Most data in data centers is used frequently.
- **Data:** Approximately 41% of stored data has never been accessed. (<https://www.itbear.com/technews/report-reveals-shocking-data-wastage-70-80-of-data-on-some-servers-remains-unused/>)
- **Myth:** AI training datasets are the primary driver of data center growth.
- **Data:** Inference workloads are outpacing training and are projected to account for roughly two-thirds of all AI compute by 2026. (<https://agentmarketcap.ai/blog/2026/04/17/ai-inference-overtakes-training-two-thirds-2026-compute>)
- **Myth:** Cold storage is rarely used.
- **Data:** Cold storage is a significant tier, characterized by cheap bytes with no governance, and is essential for long-term data retention. (<https://technewskb.com/cold-storage-vs-archive-tiers/>)

Methodology Note

This analysis is based on 425 web-research signals from various reputable sources. The ranges provided reflect the disagreement among these sources to ensure a comprehensive and accurate representation of the data.

Appendix: signal base

#	Category	Signal	Source
1	total	Current projections estimate 200 zettabytes of digital information stored globally, with about half of that in data cent	https://cardinalnews.org/2025/08/22/you-could-never-count-the-amount-of-digital-information-stored-in-data-centers/
2	total	Global data soared to 120 zettabytes in 2023 and should near 181 zettabytes by 2025.	https://zipdo.co/data-storage-industry-statistics/
3	total	The world generated 33 zettabytes of data in 2018 alone.	https://edgedelta.com/company/knowledge-center/data-storage-statistics
4	total	The worldwide IDC Global DataSphere is defined as a measure of how much data is created, captured, replicated, and consu	https://my.idc.com/getdoc.jsp?containerId=US53383425
5	total	The amount of data created, captured, replicated, and consumed worldwide is expected to reach 393.9 zettabytes (ZB) in 2	https://www.linkedin.com/posts/data-panacea_the-latest-idc-global-datasphere-forecast-activity-7370878620728250368-5odA
6	total	The IDC Global DataSphere forecast covers a five-year period from 2025 to 2029.	https://www.marketresearch.com/IDC-v2477/Worldwide-IDC-Global-Datasphere-Forecast-41057291/
7	total	The IDC Global DataSphere forecast is segmented by several categories including consumer, enterprise, and other data typ	https://my.idc.com/getdoc.jsp?containerId=US53363625

8	total	The global data storage market size is projected to grow from \$199.32 billion in 2025 to \$225.82 billion in 2026 at a co	https://www.thebusinessresearchcompany.com/report/data-storage-global-market-report
9	total	The global data storage market size is projected to grow from \$298.54 billion in 2026 to \$984.56 billion by 2034, at a C	https://www.fortunebusinessinsights.com/data-storage-market-102991
10	cloud	Global public cloud storage hit 15 EB in 2023.	https://gitnux.org/data-storage-industry-statistics/
11	enterprise	Enterprise storage utilization averages 55%.	https://gitnux.org/data-storage-industry-statistics/
12	total	Over 97% of businesses already invest in Big Data for real results.	https://www.demandsage.com/big-data-statistics/
13	total	Seagate shipped approximately 150.8 exabytes worth of storage in hard disk drives (HDD) worldwide in the second quarter	https://www.statista.com/statistics/795771/worldwide-seagate-hard-drive-capacity-shipped/
14	enterprise	The global hard disk drive (HDD) market is projected to ship approximately 278 million units in 2025, driven by demand f	https://www.industryresearch.biz/market-reports/hard-disk-drive-market-101873
15	total	Worldwide HDD shipments reached 123.9 million units in 2024, with total storage shipped reaching 1,337 exabytes.	https://www.360researchreports.com/market-reports/hard-drives-market-208689
16	total	Total HDD storage shipped reached 1,337 exabytes in 2024, reflecting a 49% surge as data-heavy applications grow.	https://www.360researchreports.com/market-reports/hard-drives-market-208689
17	total	IDC provides preliminary unit shipment ranges for worldwide hard disk drive (HDD) units by drive class and estimated sup	https://www.giiresearch.com/report/id1863635-worldwide-hard-disk-drive-3q25-preliminary.html
18	cloud	Cloud hyperscalers lead data center capacity market share in 2026, while colocation and enterprise on-premise shares are	https://www.crn.com/news/data-center/2026/data-center-market-share-face-off-hyperscalers-vs-colocation-vs-enterprise
19	cloud	Hyperscalers are projected to control approximately 67% of global data center capacity by 2031.	https://hostnoc.com/data-center-market-share-2026/
20	cloud	The hyperscale data center market was valued at USD 80.9 billion in 2025.	https://www.gminsights.com/industry-analysis/hyperscale-data-center-market
21	cloud	The hyperscale data center market is expected to grow at a CAGR of 22.2% between 2026 and 2035.	https://www.gminsights.com/industry-analysis/hyperscale-data-center-market
22	cloud	Hyperscale data centers operated by AWS, Google, and Microsoft reached 1,300 facilities in 2025.	https://www.crn.com/news/cloud/2025/hyperscale-data-centers-hit-1-300-with-aws-google-microsoft-leading-global-boom
23	total	402.74 billion gigabytes of data are created daily worldwide.	https://www.demandsage.com/big-data-statistics/

24	total	147 zettabytes of data are created annually worldwide.	https://www.demandsage.com/big-data-statistics/
25	total	90% of the world's data has been created in the last two years.	https://www.demandsage.com/big-data-statistics/
26	total	Data creation is expected to reach around 221 zettabytes annually by 2026.	https://www.optimixed.com/amount-of-data-created-daily-2026/
27	video	Video content constitutes about 82% of internet data traffic.	https://www.optimixed.com/amount-of-data-created-daily-2026/
28	total	402 million terabytes of data are generated worldwide every day.	https://thebestvpn.com/statistics/how-much-data-is-created-every-day/
29	video	Video usage accounted for 65% of all internet traffic volume in 2022.	https://www.applogicnetworks.com/press-releases/sandvines-2023-global-internet-phenomena-report-shows-24-jump-in-video-traffic-with-netflix-volume-overtaking-youtube
30	video	Video traffic volume grew by 24% in 2022.	https://www.prnewswire.com/news-releases/sandvines-2023-global-internet-phenomena-report-shows-24-jump-in-video-traffic-with-netflix-volume-overtaking-youtube-301723445.html
31	video	Netflix replaced YouTube as the individual app generating the most video volume for the first time.	https://www.applogicnetworks.com/press-releases/sandvines-2023-global-internet-phenomena-report-shows-24-jump-in-video-traffic-with-netflix-volume-overtaking-youtube
32	video	Live streaming of sporting events is driving video traffic to new heights and pushing network capacities.	https://www.sandvine.com/hubfs/AppLogic_Networks/Collateral/Global%20Internet%20Phenomena%20Reports/GIPR%202025.pdf
33	video	QUIC traffic is displacing TCP traffic within the video category.	https://www.sandvine.com/hubfs/AppLogic_Networks/Collateral/Global%20Internet%20Phenomena%20Reports/GIPR%202025.pdf
34	total	Global average internet traffic increased 3.2-fold from 2018 to 2021 with a compound annual growth rate of 26%.	https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2023/reports/Sandvine%20GIPR%202023.pdf
35	video	YouTube stores petabytes and exabytes of video content to cater to its vast user base.	https://dev.to/ujjawaltiyagi/why-youtube-never-runs-out-of-storage-its-not-just-cloud-225f
36	video	People upload 720,000 hours of videos to YouTube every single day.	https://www.reddit.com/r/AskProgramming/comments/vueyb9/how_the_fuck_does_youtube_store_all_of_its_data/
37	video	Approximately 4.3 Petabytes of new data are uploaded to YouTube every day.	https://www.reddit.com/r/AskProgramming/comments/vueyb9/how_the_fuck_does_youtube_store_all_of_its_data/
38	backup	YouTube keeps separate copies of every uploaded video in five other locations.	https://www.reddit.com/r/AskProgramming/comments/vueyb9/how_the_fuck_does_youtube_store_all_of_its_data/

39	video	Processing and redundancy cause one petabyte of daily uploads to become multiple petabytes in total storage requirements	https://www.youtube.com/watch?v=aUsK3yCbi8w
40	video	Five hours of new video are uploaded to YouTube every second.	https://www.reddit.com/r/answers/comments/45egsp/how_much_storage_space_does_youtube_use/
41	video	Netflix hosts almost 35,000 hours of content.	https://www.reddit.com/r/NoStupidQuestions/comments/s/6mfzuj/how_much_data_is_contained_in_the_netflix_library/
42	video	An hour of HD streaming is 3GB according to Netflix.	https://www.reddit.com/r/NoStupidQuestions/comments/s/6mfzuj/how_much_data_is_contained_in_the_netflix_library/
43	video	An hour of Ultra HD streaming is 7GB according to Netflix.	https://www.reddit.com/r/NoStupidQuestions/comments/s/6mfzuj/how_much_data_is_contained_in_the_netflix_library/
44	video	American Netflix had 2.2 million minutes of content as of March 2020.	https://www.reddit.com/r/theydidthemath/comments/qpxtvb/request_how_many_tbs_of_content_does_netflix_s/
45	video	The file size for one minute of Netflix full HD is estimated at 43.5 MB.	https://www.reddit.com/r/theydidthemath/comments/qpxtvb/request_how_many_tbs_of_content_does_netflix_s/
46	video	Based on 2.2 million minutes of content and 43.5 MB per minute, the total library size is estimated at 95.7 TB.	https://www.reddit.com/r/theydidthemath/comments/qpxtvb/request_how_many_tbs_of_content_does_netflix_s/
47	video	Netflix stores 140 million hours of viewing data per day.	https://blog.bytebytego.com/p/how-netflix-stores-140-million-hours
48	video	Netflix has about 17,000 titles globally.	https://news.ycombinator.com/item?id=33456782
49	video	Netflix stores copies of content in SD, 720p, HD, and 4k resolutions.	https://news.ycombinator.com/item?id=33456782
50	video	The global video streaming industry generated \$233 billion in total revenue in 2024.	https://axis-intelligence.com/streaming-statistics-definitive-data-report/
51	video	Streaming captured 47.5% of all U.S. TV viewing in December 2025, marking the highest share ever recorded.	https://axis-intelligence.com/streaming-statistics-definitive-data-report/
52	video	The global video streaming market size was valued at \$811.37 billion in 2025.	https://www.fortunebusinessinsights.com/video-streaming-market-103057
53	video	The global video streaming market is projected to grow from \$969.56 billion in 2026 to \$3,394.56 billion by 2034.	https://www.fortunebusinessinsights.com/video-streaming-market-103057
54	video	Short-form video is projected to account for 82% of global internet traffic by 2025.	https://autofaceless.ai/blog/short-form-video-statistics-2026
55	video	YouTube Shorts is reported to hit 70 billion daily views.	https://autofaceless.ai/blog/short-form-video-statistics-2026

56	video	95% of consumers watch short-form video regularly, and 71% do so daily.	https://www.kapwing.com/resources/short-form-video-statistics-tiktok-reels-and-shorts-by-the-numbers-in-2026/
57	video	90%+ of Gen Z and Millennials watch short-form video weekly across TikTok, Reels, and Shorts.	https://www.kapwing.com/resources/short-form-video-statistics-tiktok-reels-and-shorts-by-the-numbers-in-2026/
58	video	The transition from Full HD (1080p) to 4K Ultra HD and emerging 8K formats is driving substantial increases in bandwidth	https://dataintelo.com/report/data-center-video-on-demand-vod-market
59	video	High-performance object storage solutions are utilized to store and stream 4K/8K video and manage large media libraries	https://datnass.com/solutions/media
60	video	Working with 4K and 8K RAW or uncompressed footage incurs significant storage costs that must be calculated alongside ot	https://wolfcrow.com/the-costs-of-working-with-2k-and-4k-uncompressed-footage/
61	photos	About 1.95 trillion photos will be taken in 2026, which equates to approximately 5.3 billion photos daily.	https://photutorial.com/photos-statistics/
62	photos	Flickr upload statistics indicate that roughly 25 million photos are uploaded daily.	https://www.lightstalking.com/photo-statistics/
63	photos	Approximately 3.5 million photos are shared daily on Flickr.	https://www.lightstalking.com/photo-statistics/
64	photos	Google Photos ended its unlimited free storage policy on June 1, 2021, causing all new uploads to count against the user	https://www.thetechadvocate.org/google-photos-storage-limit-explained/
65	photos	Each Google Account includes up to 15 GB of storage shared across Gmail, Google Drive, and Google Photos.	https://support.google.com/accounts/answer/9312312?hl=en-EN
66	photos	T-Mobile is scheduled to kill the last remaining unlimited Google Photos storage plan on March 31, 2026.	https://www.photoworkout.com/google-photos-unlimited-storage-dead/
67	photos	Instagram enforces a file size limit of 30MB for photos.	https://filesize.org/limits/instagram/
68	photos	Instagram enforces a file size limit of 8MB for photos.	https://blondish.net/instagram-file-size-limit/
69	photos	The Instagram official account has posted 8,550 posts.	https://www.instagram.com/instagram/
70	photos	Cloud storage services for smartphone photos are compared in terms of pricing, quality, and sync features for the year 2	https://pic-tomo.com/en/blog/smartphone-photo-storage-comparison-2026
71	photos	Users perceive cloud storage for photos and videos as expensive, leading to a desire to stop	https://www.cnet.com/tech/services-and-software/stop-paying-for-photo-video-cloud-storage/

		paying for subscriptions.	
72	backup	Software solutions exist that enable automatic backup of photos and videos to both local storage and the cloud regardless	https://www.acronis.com/en/blog/posts/the-best-online-backup-for-photos/
73	photos	There are numerous cloud storage options available for photos that guarantee reliability, security, easy access, and efficiency	https://fixthephoto.com/best-cloud-storage-for-photos.html
74	photos	It is estimated that more than 2.1 trillion photos will be taken worldwide in 2025.	https://phototrend.fr/en/2025/06/more-than-2-trillion-photos-taken-in-2025/
75	photos	94% of the photos taken worldwide in 2025 are captured via smartphones.	https://phototrend.fr/en/2025/06/more-than-2-trillion-photos-taken-in-2025/
76	chats	WhatsApp has over 3 billion users worldwide according to official company statements.	https://expandedramblings.com/index.php/whatsapp-statistics/
77	chats	Meta has stated that WhatsApp has over 2 billion daily active users.	https://expandedramblings.com/index.php/whatsapp-statistics/
78	chats	WhatsApp messaging volume has been reported at roughly 100 billion messages per day.	https://expandedramblings.com/index.php/whatsapp-statistics/
79	chats	WhatsApp users send more than 100 billion messages per day according to WhatsApp statements.	https://thunderbit.com/blog/whatsapp-statistics
80	chats	Roughly 150 billion messages are exchanged each day through WhatsApp as of December 20, 2025.	https://prioridata.com/data/whatsapp/
81	chats	WhatsApp has over 2 billion monthly active users.	https://smallusefultips.com/how-many-gb-does-whatsapp-use/
82	chats	Over three billion people use messaging apps, making them one of the most popular app types.	https://www.businessofapps.com/data/messaging-app-market/
83	chats	For most people, mobile messaging consists of two platforms: Facebook Messenger and WhatsApp.	https://www.businessofapps.com/data/messaging-app-market/
84	chats	WhatsApp sends 100 billion messages globally each day.	https://gitnux.org/mobile-messaging-statistics/
85	chats	US users spend most of their time in text and iMessage is locked at 48% among iOS.	https://gitnux.org/mobile-messaging-statistics/
86	chats	Messaging apps are an essential part of daily lives for billions of people worldwide.	https://prioridata.com/data/messaging-app-stats/
87	chats	There are 4.59 billion email users worldwide, representing more than half of the global population.	https://www.theglobalstatistics.com/email-statistics/
88	chats	Global daily email volume reached 376.4 billion messages in 2025.	https://clean.email/blog/insights/email-industry-report-2026

89	chats	Daily email volume is projected to grow to 392.5 billion by 2026.	https://clean.email/blog/insights/email-industry-report-2026
90	chats	Daily email volume is projected to reach 408.2 billion by 2027.	https://clean.email/blog/insights/email-industry-report-2026
91	chats	Worldwide email usage is projected to climb to 422.4 billion daily emails by 2027.	https://gitnux.org/email-usage-statistics/
92	chats	Gmail provides 15 GB of storage per account.	https://mail.google.com/mail
93	chats	Discord handles trillions of messages, requiring engineering for instant delivery and real-time search at scale.	https://medium.com/stackwise/how-discord-stores-billions-of-messages-real-world-engineering-at-scale-215531a54bb7
94	chats	Discord migrated its message storage from MongoDB to Cassandra to achieve scalability and fault tolerance for billions o	https://discord.com/blog/how-discord-stores-trillions-of-messages
95	chats	Telegram is marketed and used as a cloud storage service, allowing users to store files in 'Saved Messages' without expl	https://www.reddit.com/r/Telegram/comments/kpkrry/how_telegram_affords_so_much_of_storage_space/
96	chats	File upload size limits for chats and media on platforms like Telegram, Discord, WhatsApp, and Slack are defined by plat	https://www.filetoolworks.com/file-size-limits
97	ai	The global AI training dataset market size was valued at USD 3.2 billion in 2024.	https://www.gminsights.com/industry-analysis/ai-training-dataset-market
98	ai	The global AI training dataset market is projected to grow at a CAGR of 20.5% between 2025 and 2034.	https://www.gminsights.com/industry-analysis/ai-training-dataset-market
99	ai	The global AI training dataset market size was estimated at USD 2.82 billion in 2024.	https://www.marketsandmarkets.com/Market-Reports/ai-training-dataset-market-153819655.html
100	ai	The global AI training dataset market is projected to reach USD 9.58 billion by 2029, growing at a CAGR of 27.7%.	https://www.marketsandmarkets.com/Market-Reports/ai-training-dataset-market-153819655.html
101	ai	The AI training dataset market size is projected to grow from \$3.4 billion in 2026 to \$15.42 billion by 2034, at a CAGR	https://straitsresearch.com/report/ai-training-dataset-market
102	science	Increasing demand for complex, multidimensional data is fostering strategic collaborations among biotechnology, pharmace	https://www.grandviewresearch.com/industry-analysis/ai-training-dataset-market
103	ai	AI training dataset market growth is attributed to the rapid adoption of AI technologies.	https://www.fortunebusinessinsights.com/ai-training-dataset-market-109241
104	chats	Chats may be reviewed and used to improve AI models.	https://chatgpt.com/
105	ai	JLL estimates that 7 GW of new global data center capacity built in 2026 will be specifically allocated to AI workloads.	https://www.allaboutai.com/resources/ai-statistics/ai-data-centers/

106	ai	Uptime's 2026 survey indicates rising AI pressure on data center costs and capacity, with modal rack density reaching 11	https://www.datacenterknowledge.com/ai-data-center/s/ai-drives-data-center-uncertainty-in-uptime-s-2026-survey
107	ai	AI workloads are fundamentally resetting 2026 data center capacity forecasts due to sustained load, power constraints, a	https://www.datacenterrealestate.com/news/ai-demand-is-resetting-2026-data-center-capacity-forecasts
108	ai	Global AI data centers consumed 485 TWh of electricity in 2026.	https://axis-intelligence.com/ai-data-center-statistics/
109	ai	Hyperscaler capital expenditure for AI data centers reached \$725B in 2026.	https://axis-intelligence.com/ai-data-center-statistics/
110	ai	The U.S. faced a 9.3 GW power shortfall for AI data centers in 2026.	https://axis-intelligence.com/ai-data-center-statistics/
111	ai	Training large language models (LLMs) imposes higher demands and challenges on data storage capacity as model parameter	https://link.springer.com/article/10.1007/s11280-025-01367-7
112	ai	System memory capacity is limited to the O(1TB) level, which cannot accommodate LLM training datasets that are on the or	https://luosuu.github.io/assets/files/AIOps_24.pdf
113	chats	Curated collections of datasets for chatbot training include information on dataset size, language, and usage.	https://github.com/Zjh-819/LLMDataHub
114	ai	There are potential constraints on LLM scaling posed by the availability of public human-generated text data.	https://arxiv.org/html/2211.04325v2
115	ai	LLM training storage demands involve specific requirements for datasets, checkpoints, and efficient recovery strategies.	https://www.kad8.com/ai/llm-training-storage-demands-data-and-checkpoints/
116	ai	Gartner predicts that synthetic data will account for 60% of the data used in AI and analytics projects by 2024.	https://arxiv.org/html/2310.17848v2
117	ai	Gartner predicts that synthetic data will surpass the utilization of raw data in AI models by 2030.	https://arxiv.org/html/2310.17848v3
118	ai	Synthetic data generation is a component of the synthetic data generation market which is forecasted through 2034.	https://www.fortunebusinessinsights.com/synthetic-data-generation-market-108433
119	ai	The global AI data center market size was valued at USD 17.73 billion in 2025.	https://www.fortunebusinessinsights.com/ai-data-center-market-110845
120	ai	The global AI data center market is projected to grow from USD 21.27 billion in 2026 to USD 133.51 billion by 2034.	https://www.fortunebusinessinsights.com/ai-data-center-market-110845
121	ai	The global AI data center market is projected to exhibit a CAGR of 25.80% during the forecast	https://www.fortunebusinessinsights.com/ai-data-center-market-110845

		period from 2026 to 2034.	
122	ai	AI data centers have crossed from a marginal share of global electricity demand to its fastest-growing component.	https://axis-intelligence.com/ai-data-center-statistics/
123	ai	AI is projected to drive a 165% increase in data center power demand by 2030.	https://www.goldmansachs.com/insights/articles/ai-to-drive-165-increase-in-data-center-power-demand-by-2030
124	ai	Organizations are increasingly investing in specialized infrastructure optimized for AI workloads, including high-densit	https://www.marketsandmarkets.com/Market-Reports/ai-data-center-market-267395404.html
125	ai	The data center construction boom has entered a new chapter with AI adoption accelerating and the floor of potential dem	https://www.bain.com/insights/ai-data-center-forecast-from-scramble-to-strategy-snap-chart/
126	ai	AI training involves writing new, multi-gigabyte checkpoints consisting of intermediary model weights and optimizer stat	https://simplyblock.io/blog/ai-training-checkpoint-model-registry-storage-kubernetes/
127	ai	For mixed-precision training, the combination of model weights and optimizer states results in a surprisingly large stor	https://www.cudocompute.com/blog/storage-requirements-for-ai-clusters
128	ai	Multi-tier checkpointing is used in large AI training jobs to manage the storage of model states.	https://cloud.google.com/blog/products/ai-machine-learning/using-multi-tier-checkpointing-for-large-ai-training-jobs
129	ai	Tiered storage architecture for AI servers separates storage into tiers matched to access patterns, such as distinguishi	https://vrlatech.com/storage-architecture-ai-gpu-servers-2026/
130	ai	Model weights are often distributed via ad hoc scripts, manual copies between storage buckets, or unsecured shared files	https://www.cncf.io/blog/2026/03/27/the-weight-of-ai-models-why-infrastructure-always-arrives-slowly/
131	ai	On-premises AI infrastructure requires specific design patterns for storing, versioning, and recovering large model chec	https://sysart.consulting/insights/checkpoint-model-storage-on-premises/
132	ai	An inference log is defined as a permanent record of all inferences made by an AI system, including input and output dat	https://practical-ai-act.eu/latest/engineering-practice/inference-log/
133	ai	Inference logs serve as the basis for monitoring AI system operation to ensure intended behavior and compliance with leg	https://practical-ai-act.eu/latest/engineering-practice/inference-log/
134	ai	Logging of inference data should allow for the reconstruction of events within the AI system.	https://github.com/aai-institute/practical-ai-act/blob/main/docs/engineering-practice/inference-log.md
135	ai	Middleware can intercept incoming inference requests to log relevant data before passing the request to the model for pr	https://practical-ai-act.eu/latest/showcase/implementation-notes/inference-logging/

136	ai	Comprehensive audit trails for AI inference systems require defining what to log, storage architecture, tamper-evidence,	https://gigagpu.com/ai-audit-trail-logging-inference/
137	social	As of 2020, an estimated third of the global population were using social networks.	https://www.statista.com/topics/1164/social-networks/
138	social	69.6% of people globally use social media as of 2026.	https://resourcera.com/data/social/social-media-usage/
139	total	Meta operates hundreds of exabyte-scale storage clusters that serve all of Meta's external and internal products, includ	https://engineering.fb.com/2026/07/01/data-infrastruct-ure/metas-ai-storage-blueprint-at-scale/
140	enterprise	Facebook leverages Hive and Hadoop to build a petabyte-scale data warehouse, processing approximately 12TB of new data a	https://ecweb.ecer.com/topic/en/detail-320302-facebo-ok-uses-hive-hadoop-for-petabytescale-data-wareho-using.html
141	enterprise	Facebook scaled its data warehouse to manage over 300 PB of data, evolving storage formats from RCFile to ORCFile.	https://www.engineering.fyi/article/scaling-the-facebo- k-data-warehouse-to-300-pb
142	enterprise	Scalable analysis on large data sets is core to the functions of engineering and non-engineering teams at Facebook for a	https://research.facebook.com/publications/data-wareh-ousing-and-analytics-infrastructure-at-facebook/
143	total	Facebook is renowned as one of the big data companies proficient at using big data to drive its business.	https://www.acceldata.io/blog/data-engineering-best-p-ractices-facebook-exascale
144	social	X (Twitter) currently has 561 million active users.	https://backlinko.com/twitter-users
145	social	More than 132 million daily users worldwide access X/Twitter daily.	https://backlinko.com/twitter-users
146	social	X/Twitter is most popular among users aged 18-29 in the US.	https://backlinko.com/twitter-users
147	social	Worldwide, men use X/Twitter more than women.	https://backlinko.com/twitter-users
148	social	X/Twitter statistics for 2026 include data on users, ad reach, revenue, engagement, and demographics drawn from DataRepo	https://onclusive.com/resources/blog/x-twitter-statistic-s-2026/
149	social	Metricool provides X (Twitter) statistics for marketers based on data from over 1 million posts.	https://metricool.com/x-twitter-statistics/
150	social	LinkedIn's data infrastructure supports core features like recommendation and search systems through batch computing for	https://vutr.substack.com/p/diving-deep-into-linkedins-data-infrastructure
151	social	Reddit is described as a social media platform hosting thousands of communities and endless conversation.	https://redditinc.com/

152	social	Reddit is categorized as an American proprietary social news aggregation and forum social media platform.	https://en.wikipedia.org/wiki/Reddit
153	social	Product discussions on Reddit involve raw, honest feedback, while LinkedIn scales these through professional growth and	https://www.linkedin.com/pulse/truth-discussions-start-reddit-scale-linkedin-venu-reddy-pkbfc/
154	total	IDC's Global DataSphere quantifies and analyzes the amount of data created, captured, replicated, and consumed in any gi	https://my.idc.com/getdoc.jsp?containerId=IDC_P38353
155	enterprise	The Global DataSphere is forecast by several segments including consumer and enterprise.	https://my.idc.com/getdoc.jsp?containerId=US53363625
156	enterprise	The Global DataSphere is forecast by data type, location (core, edge, endpoint), and cloud/noncloud status.	https://my.idc.com/getdoc.jsp?containerId=US53363625
157	enterprise	IDC's Global DataSphere looks at how much of that data is stored across various storage media including HDD and SSD.	https://my.idc.com/getdoc.jsp?containerId=IDC_P38353
158	enterprise	The Global DataSphere taxonomy includes Enterprise Computing as a primary segment.	https://www.idc.com/research/viewtoc.jsp?containerId=US53121925
159	enterprise	The Global DataSphere Pivot Table segments data by data creator (consumer and enterprise).	https://mfe-prod.idc.com/getdoc.jsp?containerId=US53404625
160	enterprise	IDC presents a five-year forecast for the worldwide IDC Global DataSphere covering 2025–2029.	https://www.giiresearch.com/report/id1726448-worldwide-idc-global-datasphere-forecast.html
161	enterprise	The dynamics shaping the Global DataSphere are shifting in important ways.	https://my.idc.com/getdoc.jsp?containerId=US53363625
162	enterprise	Globally, about 55% of enterprise data is considered dark, with some organizations reporting up to 75% unused informatio	https://www.datastackhub.com/insights/dark-data-statistics/
163	enterprise	Research from multiple sources including Gartner, Splunk, and DataStackHub estimates that approximately 55% of enterpris	https://arguslabs.ai/dark-data-enterprise-55-percent-problem/
164	enterprise	85 percent of customer data effectively sits in the dark.	https://zipdo.co/dark-data-statistics/
165	enterprise	90 percent of executives say leveraging dark data is critical to long term success.	https://zipdo.co/dark-data-statistics/
166	enterprise	Global dark data is projected to reach 134 zettabytes.	https://worldmetrics.org/dark-data-statistics/
167	enterprise	Only 15% of enterprise data gets active analysis.	https://worldmetrics.org/dark-data-statistics/
168	enterprise	Dark data drives 30% of data center inefficiencies.	https://worldmetrics.org/dark-data-statistics/

169	enterprise	52% of IT leaders rank dark data as a significant issue.	https://worldmetrics.org/dark-data-statistics/
170	backup	The 3-2-1 backup rule is a widely embraced data protection strategy that recommends having three copies of data stored o	https://www.veeam.com/blog/321-backup-rule.html
171	enterprise	Backup creates point-in-time copies of data for recovery after corruption or deletion, while replication maintains synch	https://trilio.io/resources/backup-vs-replication/
172	enterprise	Backup is defined as the ability to maintain a timestamped copy of data used to restore lost data, with redundancy being	https://learn.microsoft.com/en-us/azure/reliability/concept-redundancy-replication-backup
173	enterprise	Amazon S3 led the global enterprise data storage software industry in 2024 with a market share of * percent.	https://www.statista.com/statistics/1258456/enterprise-data-storage-software-market-share-vendor-worldwide/
174	enterprise	Pure and NetApp followed Amazon S3 in the global enterprise data storage software industry market share in 2024, account	https://www.statista.com/statistics/1258456/enterprise-data-storage-software-market-share-vendor-worldwide/
175	enterprise	Enterprise data storage is defined as a centralized repository for information that commonly offers data management, pro	https://solutionsreview.com/data-storage/the-best-enterprise-data-storage-solutions/
176	cloud	The future of enterprise storage is evolving rapidly, driven by the rise of cloud infrastructure.	https://www.statista.com/topics/3150/data-storage/
177	enterprise	Log management and SIEM platforms charge based on the volume of data users ingest, making log data growth a significant	https://chronosphere.io/learn/observability-log-data-trends/
178	enterprise	Full-stack observability models integrate logs, metrics, and traces into a shared storage, schema, and query layer.	https://openobserve.ai/blog/full-stack-observability-logs-metrics-traces/
179	enterprise	Log collectors operate as DaemonSets on every node to read log files from the filesystem and ship them to central storag	https://k8s.info/docs/intermediate/observability
180	enterprise	Data deduplication is a technique for eliminating duplicate copies of repeating data to improve storage utilization and	https://en.wikipedia.org/wiki/Data_deduplication
181	enterprise	Data Deduplication is a feature that reduces the impact of redundant data on storage costs by optimizing free space on a	https://learn.microsoft.com/en-us/windows-server/storage/data-deduplication/overview
182	enterprise	Data deduplication is a data optimization technique that eliminates duplicate copies of data, reducing the overall data	https://www.unitsmash.com/tools/data-deduplication-ratio-calculator
183	enterprise	Data deduplication and other methods of reducing storage consumption play a vital role in affordably managing the explos	https://snia.org/sites/default/files/Understanding_Data_Deduplication_Ratios-20080718.pdf

184	enterprise	Deduplication ratios help quantify how effectively storage systems reduce redundant data, allowing organizations to opti	https://y2calculate.com/deduplication-ratio-to-percenta-ge-calculator/
185	cold	Data center storage is categorized into tiers including hot, warm, cold, and archive.	https://www.solved.scality.com/data-center-storage-tie-rs/
186	cold	Most companies pay hot storage prices for cold data.	https://datastorage.com/articles/hot-warm-cold-archive-the-data-tiering-strategy-that-cuts-storage-bills-by-hal-f/
187	cold	Data tiering strategies can cut storage bills by half.	https://datastorage.com/articles/hot-warm-cold-archive-the-data-tiering-strategy-that-cuts-storage-bills-by-hal-f/
188	cold	Real per-GB prices for storage tiers are available in 2026.	https://datastorage.com/articles/hot-warm-cold-archive-the-data-tiering-strategy-that-cuts-storage-bills-by-hal-f/
189	cold	Cold storage is defined as a tier characterized by cheap bytes with no governance.	https://technewskb.com/cold-storage-vs-archive-tiers/
190	cold	Archiving is defined as a capability that is governed and searchable, distinct from the cold storage tier.	https://technewskb.com/cold-storage-vs-archive-tiers/
191	cold	Retrieval fees, egress costs, and minimum duration requirements are the primary drivers of storage bills rather than jus	https://technewskb.com/cold-storage-vs-archive-tiers/
192	cold	Following data archiving best practices ensures compliance and security during archive transitions.	https://www.solved.scality.com/data-center-storage-tie-rs/
193	total	Approximately 41% of stored data has never been accessed.	https://www.itbear.com/technews/report-reveals-shocki-ng-data-wastage-70-80-of-data-on-some-servers-rema-ins-unused/
194	total	The percentage of unused data in certain areas can be more than twice as high as the initial 41% estimate.	https://www.itbear.com/technews/report-reveals-shocki-ng-data-wastage-70-80-of-data-on-some-servers-rema-ins-unused/
195	total	A significant portion of server data is never accessed.	https://www.techspot.com/news/103497-significant-po-rtion-server-data-never-accessed-netapp.html
196	total	90% of all data is never accessed 3 months after it is stored.	https://www.linkedin.com/posts/lisa-holm1_90-of-all-da-ta-is-never-accessed-3-months-activity-7307387415428587520-QIRq
197	total	68% of the world's data is never accessed or used after its creation.	https://www.netapp.com/blog/data-sustainability-dirty-secret/
198	cold	In 2023, magnetic tape shipments reached a record 152.9 exabytes of compressed capacity, representing a 3.14% year-over-	https://magstor.com/blogs/news/lto-tape-hits-record-152-9-eb-shipped-as-data-growth-accelerates

199	cold	In 2024, LTO tape capacity shipments reached a record 176.5 exabytes, marking a 15.4% growth over 2023.	https://www.lto.org/2025/07/unstructured-data-growth-and-hybrid-cloud-adoption-drive-record-high-lto-tape-capacity-shipments-in-2024/
200	cold	The growth in tape shipments is driven by unstructured data growth and hybrid cloud adoption.	https://www.lto.org/2025/07/unstructured-data-growth-and-hybrid-cloud-adoption-drive-record-high-lto-tape-capacity-shipments-in-2024/
201	cold	Tape storage costs one-fifth of cloud storage over a 10-year period.	https://whatarethebest.com/research/industry-research-cloud-storage-backup-file-management-and-archiving-long-term-storage-solutions/
202	cold	Amazon S3 Glacier provides archive storage classes designed for long-term data retention with the lowest-cost options in	https://aws.amazon.com/s3/storage-classes/glacier/
203	cold	S3 Glacier offers three specific archive tiers: Instant Retrieval, Flexible Retrieval, and Deep Archive, optimized for d	https://docs.aws.amazon.com/AmazonS3/latest/userguide/glacier-storage-classes.html
204	cold	Objects stored in S3 Glacier Flexible Retrieval and Deep Archive are archived and cannot be accessed directly, requiring	https://docs.aws.amazon.com/AmazonS3/latest/userguide/archival-storage.html
205	cold	S3 Glacier Deep Archive is an archival storage class intended for data that is rarely accessed and requires long-term re	https://awsglossary.org/terms/s3-glacier
206	iot	IDC provides a five-year forecast for the Global DataSphere IoT device installed base and data generated market.	https://mfe-prod.idc.com/getdoc.jsp?containerId=PRF004560
207	iot	The global surveillance camera market size was valued at USD 47.9 billion in 2025.	https://www.grandviewresearch.com/industry-analysis/surveillance-camera-market-report
208	iot	The global surveillance camera market is projected to grow from USD 52.8 billion in 2026 to USD 118.1 billion by 2033.	https://www.grandviewresearch.com/industry-analysis/surveillance-camera-market-report
209	iot	The global surveillance camera market is projected to grow at a CAGR of 12.2% from 2026 to 2033.	https://www.grandviewresearch.com/industry-analysis/surveillance-camera-market-report
210	ai	Municipal procurement teams now insist on cameras that run object detection, license plate recognition, and crowd-densit	https://www.mordorintelligence.com/industry-reports/cctv-market
211	iot	Leading automotive companies manage over 20 million connected vehicles globally, processing 110 terabytes of data daily.	https://www.axrail.ai/post/copy-of-connected-car-technology-how-20-million-vehicles-generate-110tb-of-data-daily
212	iot	Modern connected vehicles generate approximately 25GB of data per hour, which amounts to 600GB daily if running continuo	https://outposts.io/article/connected-cars-generate-massive-data-decentralized-storage-6db4f3d2-e2a6-4dc0-a3f4-8e083d79b7c6
213	iot	At a fleet scale of 1000 vehicles, connected car data generation amounts to 219 petabytes	https://outposts.io/article/connected-cars-generate-massive-data-decentralized-storage-6db4f3d2-e2a6-4dc0

		annually.	-a3f4-8e083d79b7c6
214	iot	Autonomous cars generate more than 300 TB of data per year.	https://www.tuxera.com/blog/autonomous-cars-300-tb-of-data-per-year/
215	science	CERN has collected and stored one exabyte of experimental data from the Large Hadron Collider after running for over 15	https://home.cern/cern-hits-one-exabyte-stored-experimental-data-lhc/
216	science	The Worldwide LHC Computing Grid (WLCG) offers distributed computing infrastructure with storage totaling roughly an exa	https://cds.cern.ch/record/2912513
217	science	Magnetic tapes are used as the main long-term storage medium at CERN, with data continuously migrated to newer technolog	https://home.cern/science/computing/storage/
218	science	Data filtered by the trigger at LHCb is transmitted to the CERN computing centre at a rate of 2,000 events per second.	https://lhcb-outreach.web.cern.ch/data-collection/storage-and-analysis/
219	science	Genomic data storage is characterized as a lifecycle problem rather than merely a capacity problem, requiring hot data f	https://www.solved.scality.com/genomics-data-storage/
220	science	There is an anticipated growth in genomic data storage needs, driving the development of cost-effective, sustainable, an	https://www.genomebc.ca/blog/building-better-storage-for-genomics-data
221	science	Genomic data forms the foundation of precision medicine by providing information on structure, function, and variation w	https://sanogenetics.com/resources/blog/navigating-the-data-deluge-in-genomics-report
222	science	Daily Earth observation constellation data volumes reached 50 TB in 2023.	https://gitnux.org/earth-observation-industry-statistics/
223	science	Real-time Earth observation streaming reached an aggregate capacity of 10 Gbps with an average latency of one hour in 20	https://gitnux.org/earth-observation-industry-statistics/
224	science	The Climate Data Store (CDS) publishes new datasets including multi-model hydrological seasonal forecasts and historical	https://cds.climate.copernicus.eu/
225	science	The National Centers for Environmental Information (NCEI) provides a repository containing weather model output datasets	https://www.ncei.noaa.gov/products/weather-climate-models
226	science	NCEI provides near-real-time access to forecast data from organizations sharing the weather and climate model repository	https://www.ncei.noaa.gov/products/weather-climate-models
227	science	The ERA5 reanalysis dataset provides high-resolution, hourly global climate and weather data from 1950 to the present.	https://www.nature.com/articles/s41597-026-07381-2
228	science	Climate data storage optimization strategies are utilized for climate data analytics and data	https://datacalculus.com/en/blog/climate-data-and-analytics/climate-data-analyst-data-warehousing/climate-d

		warehousing to boost busin	ata-storage-optimization-guide
229	crypto	The Bitcoin blockchain distributed ledger took up less than a terabyte of space as of 2025.	https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size/
230	crypto	Blockchain.com provides real-time tracking of the Bitcoin blockchain size in megabytes and gigabytes.	https://www.blockchain.com/charts/blocks-size
231	crypto	Bitcoin blockchain size is tracked hourly from a live Bitcoin Knots full node to determine storage requirements for runn	https://www.blockchainsize.org/
232	crypto	Comparisons of average block size exist for major networks including Bitcoin, Ethereum, and Solana to highlight differen	https://coinledger.io/research/bitcoin-blockchain-size-and-growth-over-time
233	crypto	Decentralized storage networks utilize blockchain technology to facilitate storage operations and validate transactions	https://www.techtarget.com/searchstorage/tip/Comparing-4-decentralized-data-storage-offerings
234	crypto	Filecoin (FIL), Arweave (AR), and Storj (STORJ) are identified as top decentralized storage crypto projects with growing	https://bingx.com/en/learn/article/top-decentralized-storage-crypto-projects-to-know
235	crypto	Crypto mining data centers differ from standard facilities in hardware, cooling, and scalability requirements.	https://selectrow.com/crypto-mining-data-centers/
236	crypto	The combination of cheap energy and deregulated markets has made Texas a top choice for large, industrial-scale crypto m	https://journal.uptimeinstitute.com/crypto-data-centers-the-good-the-bad-and-the-electric/
237	gaming	Game updates can result in significant disk usage, with users reporting issues where installation processes take a long	https://steamcommunity.com/discussions/forum/1/360671352685682883/
238	gaming	Users may have large amounts of data (e.g., 75GB) sitting in the Steam 'downloading' folder while only a smaller portion	https://steamcommunity.com/discussions/forum/1/4625854988055219244/
239	gaming	Game updates increase the total file size of a game, with examples citing a 30GB game receiving a 10GB update.	https://www.reddit.com/r/Steam/comments/pm57rh/how_do_updates_actually_work_in_terms_of_file_size/
240	gaming	Cloud gaming allows users to play high-quality games on various devices without needing local storage for game installat	https://www.gameslearningsociety.org/wiki/how-much-data-is-required-for-cloud-gaming/
241	gaming	Understanding system requirements for cloud gaming involves evaluating internet connection, device compatibility, operat	https://gamesreq.com/understanding-the-system-requirements-for-cloud-gaming/
242	gaming	Minimum system and network requirements for cloud gaming vary by platform and include specific internet speed tips and s	https://cloudloadout.com/minimum-system-network-requirements-for-cloud-gaming/

243	gaming	A cloud gaming PC requirements checklist prioritizes processing power, memory, storage, and graphics card specifications	https://thetechgenie.in/cloud-gaming-pc-requirements-checklist-a-guide-to-a-seamless-experience/
244	web	Common Crawl maintains a free, open repository of web crawl data.	https://commoncrawl.org/
245	web	Common Crawl is a 501 (c) (3) non-profit founded in 2007.	https://commoncrawl.org/
246	web	Common Crawl makes wholesale extraction, transformation and analysis of open web data accessible.	https://commoncrawl.org/
247	web	Common Crawl releases monthly archives with statistics on size, number of pages, unique URLs, hosts, domains, and top-le	https://commoncrawl.github.io/cc-crawl-statistics/
248	web	The number of released pages per month in Common Crawl fluctuates over time.	https://commoncrawl.github.io/cc-crawl-statistics/plots/crawlsize
249	web	Fluctuations in Common Crawl page counts are due to changes in available seeds, scheduling policy for page revisits, and	https://commoncrawl.github.io/cc-crawl-statistics/plots/crawlsize
250	chats	WhatsApp Web allows users to send and receive messages and files.	https://web.whatsapp.com/
251	video	WhatsApp Web supports one-on-one or group video and voice calls directly from the browser.	https://www.whatsapp.com/
252	web	As of October 16, 2024, the compressed size of Wikipedia's current text version is approximately 24.7 GB.	https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia
253	web	Wikipedia contains over 66 million pages, with articles constituting 10.93 percent of the total page count.	https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia
254	web	Wikipedia is approximately 122.96 times the size of Encyclopædia Britannica, with a total volume size estimated at 9.84	https://en.wikipedia.org/wiki/Wikipedia:Size_in_volumes
255	web	Determining the complete storage size of Wikipedia is challenging due to the inclusion of text, images, videos, media, a	https://takingabyte.com/2025/01/the-size-of-wikipedia/
256	web	Wikipedia publishes approximately 55 million articles across 306 languages.	https://www.vle.lt/straipsnis/wikipedia/
257	web	A single copy of the Internet Archive library collection occupies 200+ Petabytes of server space.	https://archive.org/about/
258	backup	The Internet Archive stores at least 2 copies of its library collection.	https://archive.org/about/

259	web	The Internet Archive system is designed to store many billions of objects and petabytes of content.	https://www.cs.huji.ac.il/~kirk/IAArchitecture.pdf
260	science	Capricorn replicates the Internet Archive's successful deployment of the PetaBox for major academic institutions, digital	https://archive.org/web/petabox
261	web	An employee carried the equivalent of 10 petabytes of hard drives purchased to integrate the Internet Archive infrastruc	https://impreza.host/discover-the-internet-archive-storage-infrastructure/
262	adult	Pornhub generated over \$100 million yearly in revenue as of 2023, driven primarily by advertising.	https://zipdo.co/porn-hub-statistics/
263	adult	Pornhub traffic topped 129 billion visits in 2023.	https://zipdo.co/porn-hub-statistics/
264	adult	Pornhub released its 12th annual 'Year in Review Insights' report for 2025, which includes data analysis and infographic	https://www.xbiz.com/news/294273/pornhub-releases-2025-year-in-review-report
265	adult	Pornhub's 2025 Year in Review report analyzes traffic fluctuations and cultural obsessions, including national data for	https://www.pornhub.com/insights/2025-year-in-review
266	total	Data centers exist because consumers, businesses, governments and communities depend on 'digital everything'.	https://www.coresite.com/hubfs/-website-documents/Fact%20Sheets/CoreSite-FS-MythsTruths.pdf
267	crypto	Data centers differ from crypto mining in terms of energy, water, noise, demand, and regulation.	https://datacenterfacts.org/
268	total	The construction of data centers is subject to state and local zoning and environmental regulations.	https://www.coresite.com/hubfs/-website-documents/Fact%20Sheets/CoreSite-FS-MythsTruths.pdf
269	total	There are real-world failures and energy consumption figures associated with data centers that challenge common myths.	https://www.linkedin.com/pulse/what-ever-told-you-data-centres-myths-realities-rodriguez-antibon-m4x5f
270	total	Projections for data center energy and demand extend up to 2030.	https://www.linkedin.com/pulse/what-ever-told-you-data-centres-myths-realities-rodriguez-antibon-m4x5f
271	ai	Rapid adoption of AI among enterprises and consumers is fueling the growth of data centers.	https://programs.com/resources/data-center-statistics/
272	total	The data center market is forecasted to reach USD 584.86 Billion by 2032, showing a massive 140% growth.	https://www.sci-tech-today.com/stats/data-centers-statistics/
273	video	Images, videos and memes of domestic cats make up some of the most viewed content on the World Wide Web.	https://en.wikipedia.org/wiki/Cats_and_the_Internet
274	video	Cats and the Internet denotes the pervasive dominance of domestic cat-themed content	https://gropedia.com/page/Cats_and_the_Internet

		across digital media, encompassing	
275	video	The internet is flooded with cat content including memes, cute videos, and cat-themed video games.	https://blog.everpuredata.com/perspectives/how-much-of-the-worlds-data-is-cat-content/
276	web	Dogs lead on ownership and scale online, while cats still rule Wikipedia and much of Europe.	https://dogwithblog.in/cats-vs-dogs-internet-popularity/
277	cloud	The decoupling of storage from processing (compute) has emerged as a transformative paradigm in modern computing archite	https://www.linkedin.com/pulse/storage-vs-compute-why-splitting-up-wins-cloud-game-pmp-csp-po--n6gec
278	ai	Treating storage as a secondary capacity purchase creates long-term risk because compute and data scale through differen	https://blog.westerndigital.com/ai-data-center-storage-strategy-compute-vs-data-scaling/
279	ai	AI data center planning requires tier-aware retention strategies for training data and inference workloads.	https://blog.westerndigital.com/ai-data-center-storage-strategy-compute-vs-data-scaling/
280	enterprise	Separating storage and compute allows businesses to consume data and scale storage independently of the compute required	https://builtin.com/articles/knowledge-compute-storage-data-platforms
281	cloud	In modern data centers and cloud environments, the machine running the database is typically separate from the storage d	https://www.thenile.dev/blog/storage-compute-separation
282	ai	AWS, Azure, and Google Cloud have rewritten the rules of cloud block storage to accommodate AI workloads, agentic system	https://www.lucidity.cloud/blog/the-2026-cloud-storage-reset-why-aws-azure-and-google-just-made-your-storage-bill-bigger
283	cloud	Cloud storage pricing comparisons are available for AWS S3, Azure Blob, and Google Cloud Storage, including per-GB rates	https://n2ws.com/blog/cloud-storage-cost
284	cloud	Cloud regions for Google, AWS, and Azure can be geographically pinpointed to understand hardware specs like high-density	https://github.com/LUDOWIZ/the-storage-hierarchy
285	cloud	AWS services are delivered via a network of server farms located throughout the world.	https://en.wikipedia.org/wiki/Amazon_Web_Services
286	enterprise	Large enterprises held the largest share of cloud storage in 2025 due to managing high volumes of data and complex IT sy	https://www.polarismarketresearch.com/industry-analysis/cloud-storage-market
287	cloud	North America had the highest share of the cloud storage market by region.	https://www.polarismarketresearch.com/industry-analysis/cloud-storage-market
288	cloud	The global cloud storage market is expected to experience significant growth driven by increasing demand for scalable an	https://www.fortunebusinessinsights.com/cloud-storage-market-102773
289	cloud	Dropbox has over 700 million registered users and serves 600,000 teams.	< https://play.google.com/store/apps/details?id=com.dropbox.android&hl=en-US >

290	video	Dropbox allows users to send videos exactly as captured without compression.	https://www.dropbox.com/
291	enterprise	Dropbox is positioned as a solution for syncing large files and scaling with teams.	https://www.dropbox.com/compare/onedrive
292	cloud	Cloud storage pricing comparisons for services like Dropbox and OneDrive are available as of July 2026.	https://www.cloudslinker.com/projects/cloud-storage-pricing-comparison/
293	enterprise	OneDrive and Dropbox are compared for cloud storage, file sharing, collaboration, security, and Microsoft 365 integration	https://www.m365.fm/blog/onedrive-vs-dropbox-which-cloud-storage-platform-is-better/
294	backup	Dropbox enables users to backup essential files to the cloud for access from any location.	< https://play.google.com/store/apps/details?id=com.dropbox.android&hl=en-US >
295	video	The CDN market is segmented by content type, specifically distinguishing between Video CDN and Non-Video CDN.	https://www.mordorintelligence.com/industry-reports/content-delivery-network-market
296	gaming	Increased demand for latency-free gaming experiences and enhanced video content is projected to drive CDN market growth.	https://www.emergenresearch.com/industry-report/content-delivery-network-market
297	web	CDN usage analysis includes a breakdown of requests for different types of content such as HTML, Subdomain, and Third-party	https://almanac.httparchive.org/en/2025/cdn
298	web	cdnjs serves over 200 billion requests each month, indicating significant volume for open-source library content.	https://cdnjs.com/
299	video	At the end of 2025, video traffic accounted for around 75 percent of all mobile data traffic.	https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-update
300	video	As of February 2025, video apps accounted for around 76 percent of global mobile data usage every month.	https://www.statista.com/statistics/383715/global-mobile-data-traffic-share/
301	social	Social networking accounted for eight percent of global mobile data volume as of February 2025.	https://www.statista.com/statistics/383715/global-mobile-data-traffic-share/
302	video	Video consumption remains dominant at 75 percent of mobile traffic in the Ericsson Mobility Report 2026.	https://telecomlead.com/5g/ericsson-mobility-report-2026-global-mobile-data-traffic-to-reach-515-eb-by-2031-as-india-leads-smartphone-data-consumption-126290
303	total	Global network traffic is projected to reach 515 EB per month by 2031.	https://telecomlead.com/5g/ericsson-mobility-report-2026-global-mobile-data-traffic-to-reach-515-eb-by-2031-as-india-leads-smartphone-data-consumption-126290
304	cloud	The distribution of Content Delivery Networks (CDN), cloud, and content provider capacity across Internet Exchange Point	https://blog.apnic.net/2026/06/26/peering-capacity-at-public-internet-exchanges-what-the-data-reveals/

305	cloud	Peering in 2025 covers Content Delivery, Business to Business Services, Cloud Interconnect, and Traditional Peering.	https://www.ciscolive.com/c/dam/r/ciscolive/global-event/docs/2025/pdf/BRKMSI-2005.pdf
306	total	Public peering policies are generally determined by the size and nature of traffic for access and content networks.	https://lac-ix.org/wp-content/uploads/2026/02/Peering_market_at_a_glance.pdf
307	cloud	Major CDNs, cloud providers, and content providers participate in public Internet exchange ecosystems at an enormous scale	https://pulse.internetsociety.org/en/blog/2026/07/public-peering-remains-important-in-the-hyperscale-era/
308	total	Global unstructured data volumes are projected to surge from 64 zettabytes in 2020 to 181 zettabytes by 2025.	https://worldmetrics.org/unstructured-data-statistics/
309	total	Unstructured data is projected to account for 75% of all global data by 2025.	https://worldmetrics.org/unstructured-data-statistics/
310	total	Approximately 90% of global data is unstructured.	https://www.demandsage.com/big-data-statistics/
311	social	70% of global data is user-generated.	https://www.demandsage.com/big-data-statistics/
312	total	In 2024, the global volume of data created, captured, copied, and consumed was 149 zettabytes.	https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/
313	total	By the end of 2025, the global volume of data is projected to rise to 181 zettabytes.	https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/
314	total	Global data storage reached approximately 120 zettabytes in 2023.	https://www.komprise.com/glossary_terms/zettabyte/
315	total	Global data storage is projected to approach 200 zettabytes by 2026 based on IDC research.	https://www.komprise.com/glossary_terms/zettabyte/
316	total	Only about 2% of all data created is actually stored and retained, while the vast majority is transient.	https://www.statspanda.com/live-counters/terabytes-of-data-created-globally
317	total	Global data creation is projected to reach 284 zettabytes annually by 2027.	https://www.statspanda.com/live-counters/terabytes-of-data-created-globally
318	ai	Seagate shipped 130 exabytes of storage in Q1 FY2026, with AI nearline HDD demand identified as the primary driver.	https://abhs.in/blog/seagate-q1-2026-ai-storage-hdd-data-center-nearline-devops-playbook
319	total	Seagate has delivered more than four billion terabytes of capacity over the past four decades.	https://www.seagate.com/
320	video	The uncompressed source copies of all movies available for viewing on Netflix total 3.14 petabytes.	https://www.themeasureofthings.com/singleresult.php?comp=data&unit=pb&amt=01&i=549
321	video	Videos streamed from Netflix accounted for about 29% of all Internet traffic as of 2013.	https://www.themeasureofthings.com/singleresult.php?comp=data&unit=pb&amt=01&i=549

322	video	Netflix operates a big-data platform that handles billions of events per day and has grown into petabytes/exabytes scale	https://www.linkedin.com/pulse/how-netflix-builds-runs-petabyte-to-exabyte-big-data-platform-i8f5f
323	video	Cloudinary transformed its petabyte-scale streaming data lake using Apache Iceberg and AWS analytics.	https://aws.amazon.com/blogs/big-data/how-cloudinary-transformed-their-petabyte-scale-streaming-data-lake-with-apache-iceberg-and-aws-analytics/
324	photos	In 2026, around 1.95 trillion photos are estimated to be taken worldwide.	https://photutorial.com/photos-statistics/
325	photos	More than 2.1 trillion photos are estimated to be taken worldwide in 2025, with 94% captured via smartphones.	https://phototrend.fr/en/2025/06/more-than-2-trillion-photos-taken-in-2025/
326	photos	Approximately 1.72 trillion photos were taken in the world in 2023.	https://www.lapseoftheshutter.com/photography-statistics/
327	photos	Up to 1.96 trillion photos are estimated to be taken in 2026.	https://www.lapseoftheshutter.com/photography-statistics/
328	photos	Rise Above Research estimates that the annual number of photos captured globally will approach 2.5 trillion by 2030.	https://riseaboveresearch.com/rise-above-research-forcasts-a-global-surge-in-photos-captured/
329	photos	5.3 billion selfies are taken daily, accounting for 4% of all photos taken.	https://photutorial.com/selfie-statistics/
330	photos	The statistic of 92 million selfies taken daily is considered outdated.	https://photutorial.com/selfie-statistics/
331	photos	Individuals spend 54 hours a year (or 7 minutes a day) taking selfies.	https://photutorial.com/selfie-statistics/
332	social	Selfies make up 30% of Instagram uploads daily.	https://wifitalents.com/selfie-statistics/
333	social	92% of Gen Z takes a selfie every week.	https://wifitalents.com/selfie-statistics/
334	social	58% of US adults ages 18 to 29 share selfies.	https://gitnux.org/selfie-statistics/
335	chats	Google Messages is used by a billion users and utilizes Rich Communication Services (RCS) as the industry standard for t	https://play.google.com/store/apps/details?id=com.google.android.apps.messaging&hl=en-US
336	chats	Google Messages for web allows users to send SMS, MMS, and RCS messages from a computer.	https://messages.google.com/web/
337	chats	Google SMS applications enable users to send and receive SMS text messages via Google Voice.	https://www.google.co.il/sms/
338	ai	AI training and HPC applications generate massive I/O patterns that impose unique storage demands differing substantially	https://itctshop.com/gpu-cluster-design-guide/
339	ai	Private LLM training hardware decisions include specific checkpoint storage requirements, such	https://www.rack2cloud.com/gpu-cluster-architecture-private-llm/

		as a 2.5TB requirement.	
340	ai	AI training storage requires sustained sequential write throughput measured in hundreds of GBps for checkpoint operation	https://www.softwareseni.com/ai-training-and-inference-storage-performance-requirements-benchmarked/
341	ai	Production GPU clusters benefit from storage architectures designed specifically for the demands of accelerated computing	https://massedcompute.com/faq-answers/?question=What+are+the+recommended+storage+solutions+for+NVIDIA+GPU+clusters
342	ai	AI has become the main growth engine for US data centers, reshaping data center economics, power planning, and leasing	https://www.mckinsey.com/featured-insights/charts/the-future-of-ai-workloads
343	ai	Demand for AI compute is split between two workloads: training and inference.	https://www.mckinsey.com/featured-insights/charts/the-future-of-ai-workloads
344	ai	Inference workloads are outpacing training workloads in terms of growth and volume.	https://www.delloro.com/2026-predictions-data-center-infrastructure/
345	ai	Inference is projected to account for roughly two-thirds of all AI compute in 2026.	https://agentmarketcap.ai/blog/2026/04/17/ai-inference-overtakes-training-two-thirds-2026-compute
346	ai	The share of AI compute dedicated to inference grew from one-third in 2023 to one-half in 2025, reaching two-thirds in 2026	https://agentmarketcap.ai/blog/2026/04/17/ai-inference-overtakes-training-two-thirds-2026-compute
347	ai	The inference-optimized chip market is projected to clear \$50 billion in 2026.	https://agentmarketcap.ai/blog/2026/04/17/ai-inference-overtakes-training-two-thirds-2026-compute
348	ai	Inference accelerators rely on scale-out fabrics to support utilization, redundancy, and ultra-low latency.	https://www.delloro.com/2026-predictions-data-center-infrastructure/
349	ai	The modest requirements of inference workloads are enabling smaller, distributed data center networks compared to training	https://www.bain.com/insights/ai-data-center-forecast-from-scramble-to-strategy-snap-chart/
350	ai	Data centers are being designed to accommodate flexibility between training and inference workloads.	https://www.bain.com/insights/ai-data-center-forecast-from-scramble-to-strategy-snap-chart/
351	enterprise	90% of enterprise data is unstructured.	https://shelf.io/wp-content/uploads/2024/10/idc-unstructured-data-report-sept-2024.pdf
352	enterprise	Healthcare generates 85% of its data as unstructured.	https://worldmetrics.org/unstructured-data-statistics/
353	enterprise	Government stores 90% of non-sensitive data as unstructured.	https://worldmetrics.org/unstructured-data-statistics/
354	enterprise	40% of survey respondents find it difficult to integrate unstructured data.	https://resource.itbusinesstoday.com/whitepapers/46231-Box-CPL-Q2-Q3-ABM-DTG-CAN-3.pdf
355	enterprise	28% of survey respondents cite the fast growth of connections between unstructured data sources as a top challenge.	https://resource.itbusinesstoday.com/whitepapers/46231-Box-CPL-Q2-Q3-ABM-DTG-CAN-3.pdf

356	video	The Internet Archive hosts a digital library containing texts, movies, music, audio, moving images, and software.	https://en.wikipedia.org/wiki/Internet_Archive
357	total	The Internet Archive holds approximately 624 billion archived items.	https://archive.org/
358	cloud	The Internet Archive is a non-profit digital library offering free universal access to its collection.	https://archive.org/
359	web	The Wayback Machine is an initiative of the Internet Archive that builds a digital library of Internet sites.	http://web.archive.org/
360	science	The Internet Archive includes the NASA Images project as part of its hosted projects.	https://en.wikipedia.org/wiki/Internet_Archive
361	web	Archive It is a service provided by the Internet Archive to help organizations and individuals archive web content.	https://ru.wikipedia.org/wiki/%D0%90%D1%80%D1%85%D0%B8%D0%B2_%D0%98%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82%D0%B0
362	web	Archive.today functions as a time capsule for web pages by taking snapshots of URLs.	https://archive.is/
363	other	The Rahvusarhiiv (National Archives of Estonia) provides access to digitized archive sources and databases.	https://www.ra.ee/en
364	iot	The edge computing market is expanding rapidly, driven by the efficient handling of massive data generated by IoT device	https://www.grandviewresearch.com/industry-analysis/edge-computing-market
365	iot	The forecast global revenue for the edge computing market is set to reach 350 billion U.S. dollars by 2027.	https://www.statista.com/topics/6173/edge-computing/
366	iot	Edge computing market size and share reports cover forecasts from 2026 to 2031.	https://www.mordorintelligence.com/industry-reports/edge-computing-market
367	iot	Edge computing market size and share reports cover forecasts from 2026 to 2033.	https://www.grandviewresearch.com/industry-analysis/edge-computing-market
368	iot	Federated learning is proposed as a technique to forecast energy power consumption using data from edge devices while ad	https://www.sciencedirect.com/science/article/pii/S2542660523003736
369	science	The Square Kilometre Array (SKA) Observatory is considered one of science's biggest data challenges due to the sheer vol	https://www.skao.int/en/explore/big-data
370	science	Information collected by the world's largest radio telescope will be stored and processed by global data centres.	https://theconversation.com/information-collected-by-the-worlds-largest-radio-telescope-will-be-stored-and-processed-by-global-data-centres-255268
371	science	The SKA Observatory is a global radio telescope project that will go online later this decade, creating significant data	https://phys.org/news/2025-09-world-largest-radio-telescope-global.html

372	science	The SKA project faces big data challenges in radio astronomy regarding the transport, processing, storage, and distribut	https://indico.kit.edu/event/277/contributions/1478/attachments/962/1364/Barr_BigData.pdf
373	science	NASA's Earthdata Search provides access to more than 119 petabytes of Earth observation data in its archive.	https://www.earthdata.nasa.gov/data/tools/earthdata-search
374	science	Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale	https://developers.google.com/earth-engine/datasets/
375	science	The Sentinel-2 Earth observation archive alone stands at a volume that contributes to the world's largest environmental	https://noc.ac.uk/news/noc-leads-ps7-million-project-petabytes-environmental-data-everyones-pocket
376	science	Earth Compress is a project using AI to compress and unlock the world's largest environmental datasets because satellite	https://noc.ac.uk/news/noc-leads-ps7-million-project-petabytes-environmental-data-everyones-pocket
377	crypto	Most NFT metadata is stored off-chain rather than directly on the blockchain due to cost and size limitations.	https://enjin.io/blog/the-ultimate-solution-for-nft-metadata-storage/
378	crypto	NFT metadata serves as a pointer to digital content, containing attributes like name, description, and the location of t	https://thecoincourse.com/educational-guides/tutorials/nfts/a-deep-dive-into-nft-data-what-metadata-means-and-where-it-lives
379	crypto	Decentralized storage solutions such as Arweave, Filecoin, and Sia are used to store NFT metadata and media files in Web	https://nftbirdies.com/article/arweave-filecoin-and-sia-where-to-store-nft-metadata-and-files-in-web3/
380	crypto	On-chain metadata storage is presented as a solution that solidifies valuable aspects of NFTs compared to off-chain stor	https://enjin.io/blog/the-ultimate-solution-for-nft-metadata-storage/
381	gaming	Modern games now exceed 100GB install sizes due to textures, live services, and parity demands reshaping storage require	https://www.techtimes.com/articles/313751/20260105/why-modern-games-are-getting-bigger-explaining-massive-install-sizes-storage-demands.htm
382	gaming	Subsequent game updates often result in games consuming more storage space as developers add content or higher-quality a	https://www.techspot.com/article/2680-game-install-sizes/
383	gaming	The 2025 Global Games Market Report provides key statistics on seven game ecosystems including Steam, Epic Games Store,	https://investgame.net/wp-content/uploads/2025/09/2025-Newzoo-Free-Global-Games-Market-Report.pdf
384	web	Common Crawl's multi-billion page web archives can be hundreds of terabytes in size.	https://commoncrawl.org/get-started
385	web	Prior to Summer 2013, Common Crawl data was stored in the ARC format.	https://commoncrawl.org/get-started
386	web	The WARC format allows for more efficient storage and processing of Common Crawl's web archives.	https://commoncrawl.org/get-started

387	web	Statistics of Common Crawl Monthly Archives track the size of crawls including number of pages, unique URLs, hosts, and	https://commoncrawl.github.io/cc-crawl-statistics/
388	web	The number of released pages per month in Common Crawl fluctuates over time due to changes in available seeds, schedulin	https://commoncrawl.github.io/cc-crawl-statistics/plots/crawlsizes
389	cloud	Data centers power cloud computing, online banking, emergency communications, telehealth, streaming services, and artifi	https://www.agc.org/datacenters/facts
390	total	Data centers house servers and storage infrastructure.	https://datacenterfacts.org/
391	crypto	Data centers differ from crypto mining operations.	https://datacenterfacts.org/
392	total	Data centers are critical infrastructure supporting the modern economy.	https://www.agc.org/datacenters/facts
393	total	Misinformation about data centers is often aimed at slowing their development.	https://www.goldwaterinstitute.org/the-data-center-debate-fact-vs-fiction/
394	other	Data sets include examples such as price indices, unemployment rates, literacy rates, and census data.	https://en.wikipedia.org/wiki/Data
395	cloud	Amazon S3 stores more than 500 trillion objects taking up hundreds of exabytes of capacity.	https://www.emergenresearch.com/industry-report/cloud-object-storage-market
396	cloud	Amazon S3 serves more than 200 million requests per second globally across hundreds of exabytes of data in 123 Availabil	https://www.theregister.com/off-prem/2026/03/16/aws-s3-turns-20-and-reaches-hundreds-of-exabytes/5219872
397	cloud	The S3-compatible object storage software market was valued at \$8.2 billion in 2025.	https://dataintel.com/report/s3compatible-object-storage-software-market
398	cloud	The S3-compatible object storage software market is projected to reach \$18.7 billion by 2033, growing at a 10.8% CAGR.	https://dataintel.com/report/s3compatible-object-storage-software-market
399	total	The 2026 Global Internet Phenomena Report (GIPR) provides detailed insights into global internet traffic patterns and ap	https://www.applogicnetworks.com/phenomena
400	video	Global internet traffic in 2026 is shaped by streaming and mobile behaviors.	https://wifitalents.com/global-internet-traffic-statistics/
401	ai	The Digital 2026 Global Overview Report highlights the rise of AI as a key digital trend alongside social media use.	https://datereportal.com/reports/digital-2026-global-overview-report
402	total		https://worldmetrics.org/global-internet-traffic-statistics/
403	total	Global data traffic is reported to exceed 150 (unit unspecified in snippet, likely ZB or EB	https://www.businessresearchinsights.com/market-reports/cloud-content-delivery-network-cdn-market-12792

		based on context of global t	1
404	web	Cloudflare leads deployment with 84.1% of identifiable reverse proxy usage, while Akamai leads revenue at \$4.208 billion	https://commandlinux.com/statistics/cdn-market-share-cloudflare-akamai-fastly/
405	web	cdnjs serves over 200 billion requests each month for library files, trusted by over 12.5% of all websites.	https://cdnjs.com/
406	ai	Wall Street analysis suggests agentic use of the internet and web applications will drive CDN traffic higher.	https://www.streamingmediablog.com/2026/07/2025-cdn-market-sizing.html
407	cloud	CDN platforms play a central role in data delivery by caching content closer to end users and dynamically routing traffi	https://www.marketsandmarkets.com/Market-Reports/content-delivery-networks-cdn-market-657.html
408	total	The Datasphere Market Size was valued at 71.8 USD Billion in 2024.	https://www.wiseguyreports.com/reports/datasphere-market
409	total	The Datasphere Market is expected to grow from 78.9 USD Billion in 2025 to 200 USD Billion by 2035.	https://www.wiseguyreports.com/reports/datasphere-market
410	total	IDC's Global DataSphere forecast analyzes changes to the underlying mix of data by type, including structured and unstru	https://techcontentwave.com/files/1763389937_64b594fd4f7f5607256a.pdf
411	total	IDC's Global DataSphere Forecast predicts that global data creation and replication will experience compound growth.	https://www.red-gate.com/blog/whats-the-real-story-behind-the-explosive-growth-of-data/
412	total	The IDC Global DataSphere measures how much data is created, captured, replicated, and consumed each year.	https://my.idc.com/getdoc.jsp?containerId=US53383425
413	ai	Worldwide spending on artificial intelligence is projected to reach \$632 billion in 2028, representing a compound annual	https://futurecio.tech/global-spending-on-ai-to-reach-632-billion-in-2028-idc-projects/
414	total	IDC provides a five-year forecast for the worldwide Global DataSphere from 2025 to 2029, measuring data created, capture	https://my.idc.com/getdoc.jsp?containerId=US53363625
415	ai	IDC's Global DataSphere Forecast for 2024–2028 indicates an 'AI Everywhere' trend, though the overall upsurge in data vo	https://www.marketresearch.com/IDC-v2477/Worldwide-IDC-Global-DataSphere-Forecast-36971010/
416	total	The total amount of data created, captured, copied, and consumed globally is forecast to increase rapidly through 2029.	https://www.statista.com/statistics/871513/worldwide-data-created/
417	ai	The explosion of data created and consumed by AI models is fueling AI's massive growth, creating a cycle where more data	https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/western-digital/collateral/infographic/infographic-ai-data-cycle.pdf

418	backup	The 3-2-1 backup rule is a widely used best practice for protecting data against loss and ensuring recoverability.	https://gitprotect.io/blog/3-2-1-backup-rule-complete-guide/
419	backup	The 3-2-1 backup rule defines a framework to keep three data copies on two different media types, with one stored off-si	https://www.veeam.com/blog/321-backup-rule.html
420	backup	The 3-2-1 backup rule is a data protection strategy that reduces the risk of total data loss from hardware failure, rans	https://www.avepoint.com/blog/backup/3-2-1-backup-rule
421	backup	The 3-2-1 backup rule is a practical backup strategy guide that explains how to create an offsite copy and improve the s	https://www.mrplanb.com/storage/3-2-1-backup-rule
422	backup	Storage Replica in Windows Server protects critical data by replicating volumes between servers or clusters for disaster	https://learn.microsoft.com/en-us/windows-server/storage/storage-replica/storage-replica-overview
423	backup	Storage Replica safeguards data against hardware failures, natural disasters, or planned maintenance.	https://learn.microsoft.com/en-us/windows-server/storage/storage-replica/storage-replica-overview
424	cloud	Azure Site Recovery provides an architecture for setting up disaster recovery between Azure regions for Azure VMs.	https://learn.microsoft.com/en-us/azure/site-recovery/azure-to-azure-architecture
425	backup	Cohesity offers data replication and disaster recovery solutions to ensure business continuity and minimize downtime.	https://www.cohesity.com/solutions/disaster-recovery/

Appendix: collected evidence
