



The Rising Demand for Data: Oxylabs' 2020 Trend Report

What does 2020 hold for web scraping, and what can we learn from 2019?

Key Insights	3
Data Market Overview	4
What to Expect From This Report	6
Data Center Proxies Prove to Be Sufficient for the Majority of Web Scraping Tasks	7
Most popular segments	8
Most prominent segments by single client requests	9
Global Companies Indicate Increasing Demand for Residential Proxies	10
Requests & traffic trends	10
The Rising Need for Scraping Tools: Real-Time Crawler	11
A substantial increase of scraping requests in the web scraping market	11
Most popular targets: search engines & e-commerce websites	11
Most popular months	12
The future of web scraping tools	13
Web Scraping Trends Using Data Center Proxies in E-commerce Market: 2017-2019 Case Studies	14
Case Study 1: Valentine's Day	15
Case Study 2: Mother's Day	15
Case Study 3: Black Friday	15
Lessons learned	16
Emerging Markets of Asia and Northern America: Web Scraping Activity Rises	18
Asia	18
North America	19
Thank you	20

Key Insights

The already substantial global demand for data is poised for significant further growth. This increasing demand, we assume, is partially reflected in the growing usage of all Oxylabs' public web data gathering products. All growth statistics in this section are presented in comparison to last year.



Data center proxies

- In 2019 there was a substantial **22.7% growth** in the **total number of requests**.
- Total data center proxy **traffic volume** grew by **45.8%** in 2019. This suggests that Oxylabs clients gathered more data per request on average than in 2018.
- The **retail** segment continues to dominate, and it is also the fastest-growing vertical, registering a **95%** surge in the number of **requests** issued via data center proxies. The **finance** segment is second, recording a noteworthy **86%** increase.



Residential proxies

- In 2019, there was a **165.3%** growth in the total **number of requests**.
- Meanwhile, total residential proxy **traffic volume** grew by an impressive **177.5%** in 2019.
- The 2019 spikes in traffic volume were directly related to the e-commerce industry's economic stimulation, starting with the initial preparations for Black Friday to the Christmas period.



Real-Time Crawler

- In 2019, there was a **97.6% growth in the number of requests** via Real-Time Crawler.
- Compared to the prior year, the number of requests in January 2019 was higher by a staggering **454.1%**. This is a clear indication of the **increasing overall demand for automated web data gathering tools such as Real-Time Crawler**.
- Real-Time Crawler can be separated into two main verticals: **search engines and e-commerce websites**. Comparing 2019 with historical data, the number of requests increased by **71% for search engines; 260% for e-commerce websites**.

71%
search
engines

260%
e-commerce
websites

Regions

Comparing the growth in data center traffic in Asia and North America, the latter grew by a sharp **35.96%**. Meanwhile, we recorded a steady **20%** increase from clients located in Asia.

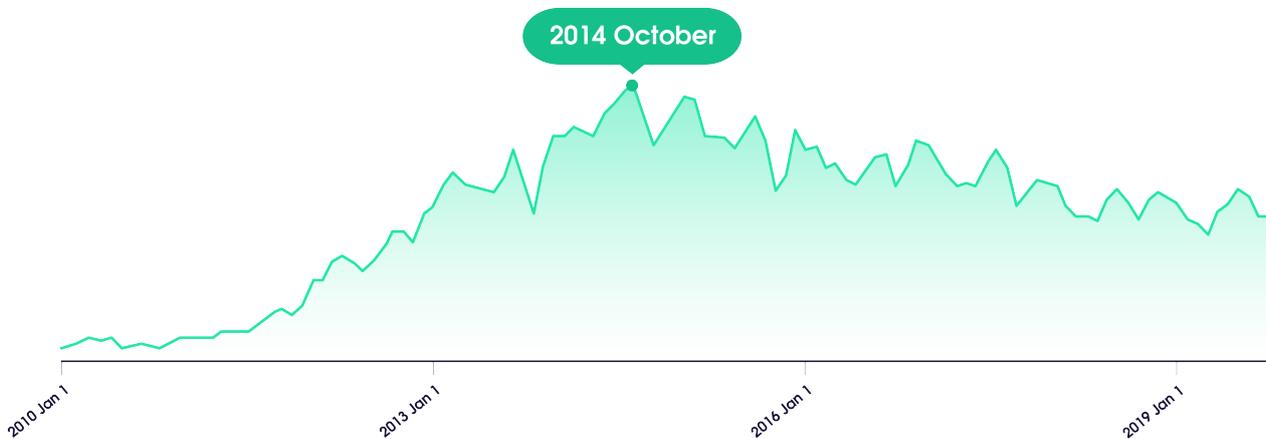


Main 2019 lesson: to carry out web scraping operations successfully and avoid a significant number of blocks, businesses need to invest in adequately upscaling their proxy resources in advance of significant market events and based on the size of the operation in mind.

Data Market Overview

Unlike just ten years ago, today, the term “big data” is familiar to laymen and specialists alike. Back in 2015, Gartner, a global research and advisory firm, [dropped “big data”](#) from its popular Hype Cycle methodology, claiming that no longer a hyped technology with untested value, it instead became **the new normal**.

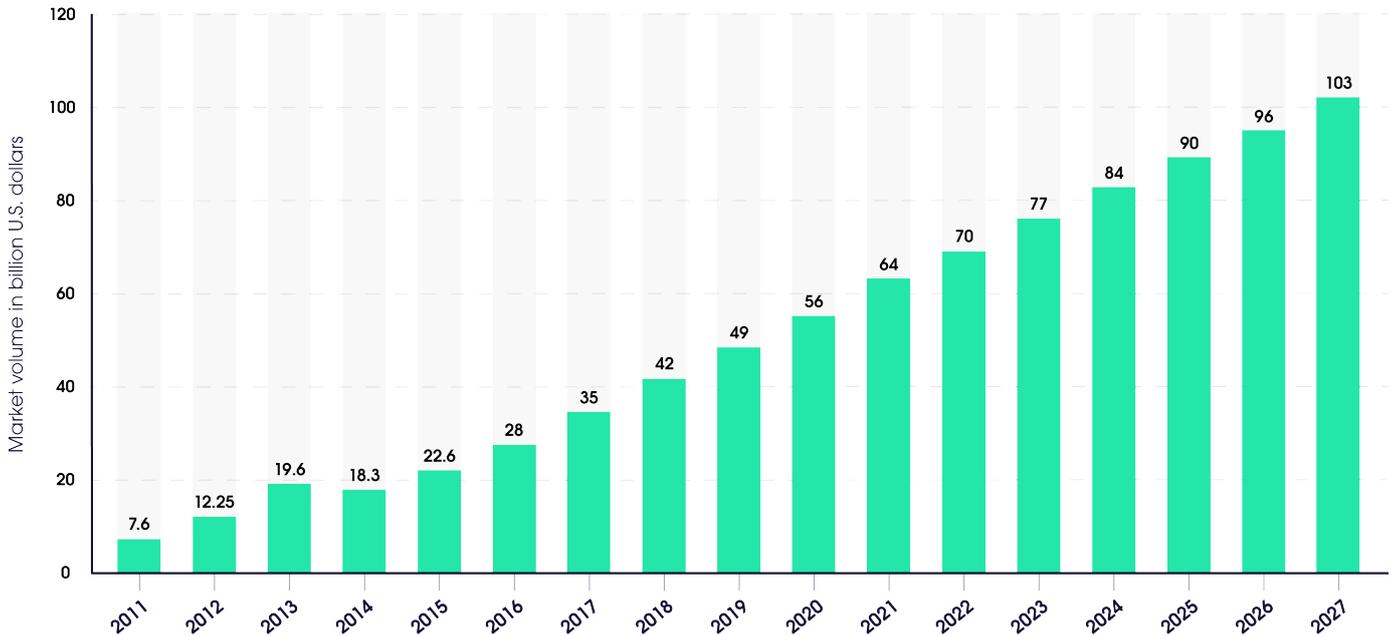
Interestingly enough, [Google Trends’ U.S. 2010-2020 search data](#) corroborates Gartner’s position by showing that the term received peak interest in October 2014 and then steadily declined in popularity in the upcoming years.



Popularity of the term “big data” in the U.S. Source: Google Trends, 2010-2020 data.

Not as new and exciting, big data matured and branched into other areas, such as **data science** and **business intelligence**, becoming irreplaceable and delivering tangible value for businesses in all types of industries worldwide.

Proving the point, the big data analytics industry continues to grow steadily. In fact, in 2019, it reached a value of \$49 billion. Assuming a steady growth rate, it should **more than double in just seven years**, reaching an [impressive \\$103 billion](#) by 2027.



Forecast of Big Data market size, based on revenue, 2011-2027. Source: Statista (2018)

The demand for data and data analysis for business intelligence is perfectly illustrated by the fact that data scientists are currently one of the most in-demand specialist positions around the world, and this demand is **growing faster than ever before**.

For example, LinkedIn's [2020 U.S. job trends report](#) features the increasing demand for artificial intelligence and data science positions as **the main 2020 job trend**, claiming that these roles "continue to proliferate across nearly every industry."

According to an [industry report](#) by Fortune Business Insights, some of the key market drivers accounting for this growth are **AI solution implementation, customer** and **operational analytics**, the latter defined by Techopedia as "a type of business analytics which focuses on improving existing operations."

Naturally, with the increasing demand for data by businesses worldwide, web scraping, the primary method of collecting public data from online sources, is also experiencing steady growth in popularity.

Especially for online-based businesses, such as e-commerce sites or travel fare aggregators, web scraping is an integral part of operations. Even more, it is now established **as a standard practice** in multiple use cases across various industries, for example, as a method to ensure that product prices are in line with the market average and determine which products have the highest demand. Forward-looking companies also leverage scraped data to ensure higher SERP ranking by optimizing their SEO efforts or check whether their ads are displayed correctly in multiple locations, among many other uses.

Further, in the report, we will outline findings based on our internal data and present some key insights related to web scraping in 2019 and the use of web data extraction solutions offered by Oxylabs.

What to Expect From This Report

Based on Oxylabs Data Analysis Department's extensive research overlooking 2019 and its comparison to historical data, web scraping firmly stands its ground in 2020 and continues to grow its influence in nearly every industry, setting a firm tone for the future of data-driven business development.

In 2019, Oxylabs' clients used Oxylabs' web data extraction solutions for a variety of business use cases. This report aims to provide thorough insights into web scraping statistics of 2019, as well as explore historical data to identify future data gathering trends by traffic and volume for 2020.

The findings of this document are supported by our aggregated client usage statistics, namely, those of data center proxies, residential proxies, and the Real-Time Crawler solution, all of which were used in varying capacity to carry out web scraping operations throughout the year of 2019.

* Traffic volumes and request amounts are summarized into trend graphs, ensuring anonymity for our clients. The findings are supported by data recorded through the use of our products, not directly our clients. These findings do not include usage statistics of clients who prefer that their usage statistics are not recorded.

Data Center Proxies Prove to Be Sufficient for the Majority of Web Scraping Tasks

Research conducted by our Data Analysis Department has found that in 2019 there was a substantial **22.7% growth** in **total request volume** via data center proxies, compared to 2018 recorded metrics.

22.7%
growth in total request volume

Taking a closer look at the data center proxies' request graph below, we can see that the most significant increase in the number of **requests** was recorded in **June**, showing an impressive **47.2% surge**.

47.2%
requests spike in June



Data center proxies' requests change by month, 2019 vs. 2018 data. Source: Oxylabs

On the subject of data center proxy **traffic** in 2019, there was a **45.8% increase** noted in total traffic load, compared to the year before. Based on this significant growth, we can expect that 2020 will also see a double-digit increase in traffic load.

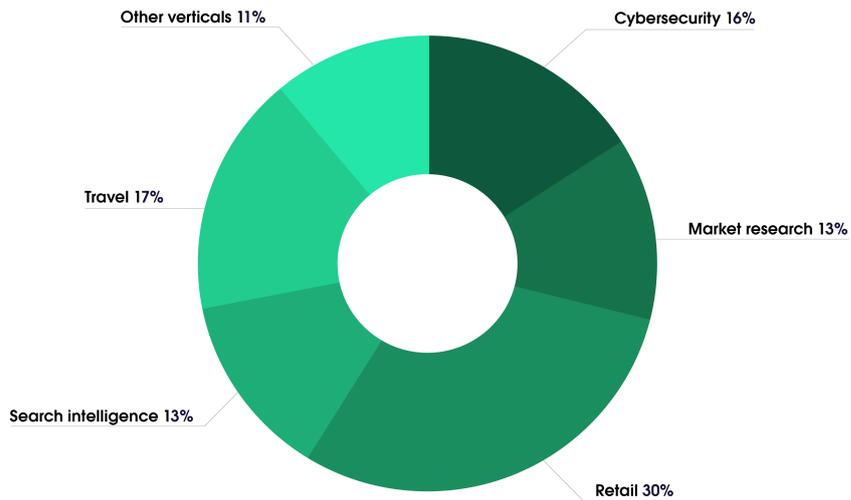
45.8%
growth in total traffic load



Data center proxies' traffic change by month, 2019 vs. 2018 data. Source: Oxylabs

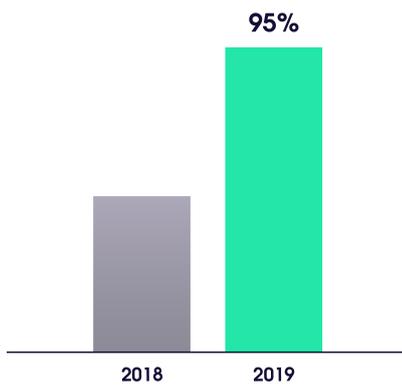
Most popular segments

Furthermore, our Data Analysis Department revealed that the most prominent segments in 2019 by the overall data center proxy **request volume** were as following:

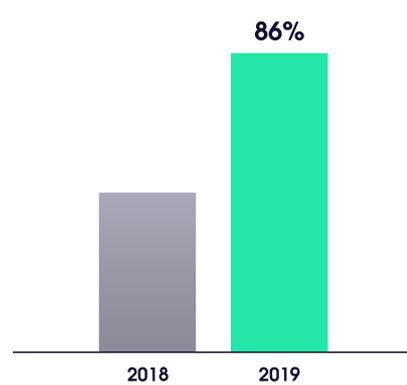


Most popular segments, based on data center proxies' requests, 2019. Source: Oxylabs

As expected, the **retail** segment continues to dominate, and it is also the fastest-growing vertical, registering a **95%** surge in the number of **requests** processed via data center proxies, compared to 2018 data.



Data center proxies' requests change in retail, 2019 vs. 2018 data. Source: Oxylabs



Data center proxies' requests change in finance, 2019 vs. 2018 data. Source: Oxylabs

More surprisingly, the second most rapid growth in the number of requests processed was noted in the **finance** segment, recording a noteworthy **86%** increase, compared to 2018.

Most prominent segments by single client requests

Naturally, companies in each industry have diverse external data requirements, as evidenced by the differences between requests generated by a single client via data center proxies.

Taking into account 2019 recorded statistics and exploring one client's average generated data center proxy **requests** in their sector, the **cybersecurity** segment's client noted the highest number of requests, followed by the **retail** and **travel** sectors' client, based on an average.

To put this into perspective, the **cybersecurity** sector's client on average processed **25%** more requests per month compared to the retail sector's client and **31%** more requests than the travel segment's customer.



Unsurprisingly, the **retail** and **travel** segments continue to increasingly rely on web scraping tasks, harnessing real-time data, which aids businesses in these segments in quickly reacting to dynamic market conditions.

Whereas, against the majority of expectations, the **cybersecurity** segment's clients are proving the point that the web scraping practice can be an effective countermeasure against various online threats for global companies.

Global Companies Indicate Increasing Demand for Residential Proxies

Requests & traffic trends

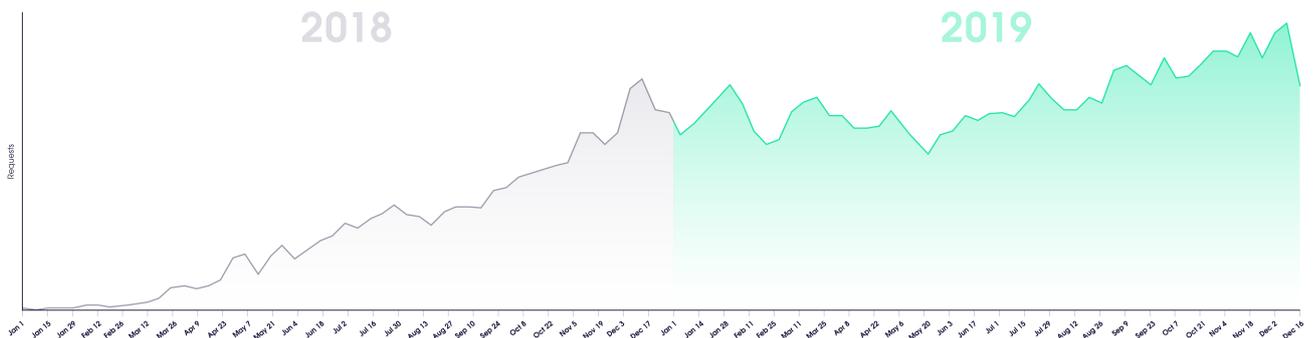
Research conducted by Oxylabs Data Analysis Department has found that in 2019, there was an astonishing **165.3% growth** recorded in total **requests** volume via residential proxies, compared to 2018 historical data.

165.3%
growth in total requests

On the subject of residential proxy traffic in 2019, there was a substantial **177.5% surge** noted in total **traffic** volume, compared to 2018 residential proxies' recorded metrics.

177.5%
growth in traffic volume

Taking a closer look at the residential proxy requests and the traffic graph below, we can see that the most prominent period was noted from **mid-November** until the **end of 2019**, which correlates with data center proxies' recorded metrics.



Residential proxies' requests by week, 2018-2019 data. Source: Oxylabs



Residential proxies' traffic by week, 2018-2019 data. Source: Oxylabs

The increase in traffic volume is directly related to the e-commerce industry's economic stimulation, from the initial preparations for Black Friday to the Christmas period. Our clients were mainly gathering data for product and pricing intelligence, in order to stay competitive and increase the volume of sales.

Residential proxies remain an excellent solution for almost any web scraping job due to the low probability of them being identified as automated traffic by even the most protected data sources, as statistics reveal.

The Rising Need for Scraping Tools: Real-Time Crawler

A substantial increase of scraping requests in the web scraping market

Research carried out by Oxylabs Data Analysis Department has found that in 2019, there was a staggering **97.6% request growth** registered via Real-Time Crawler, compared to 2018 recorded data.

Compared to the prior year, the number of requests in January 2019 started at a staggering **454.1%**. Later throughout the year, the number went down and stabilized, finally reaching **20.72%** in December 2019, as seen in the graph below.

This is a clear indication of the **overall demand for automated web data gathering tools such as Real-Time Crawler** over the past year within the web scraping market.



Real-Time Crawler requests change by month, 2019 vs. 2018 data. Source: Oxylabs

Most popular targets: search engines & e-commerce websites

Real-Time Crawler can be separated into two main verticals: **search engines** and **e-commerce websites**. Comparing 2019 with historical data, the number of requests increased by:

71%

for search engines

260%

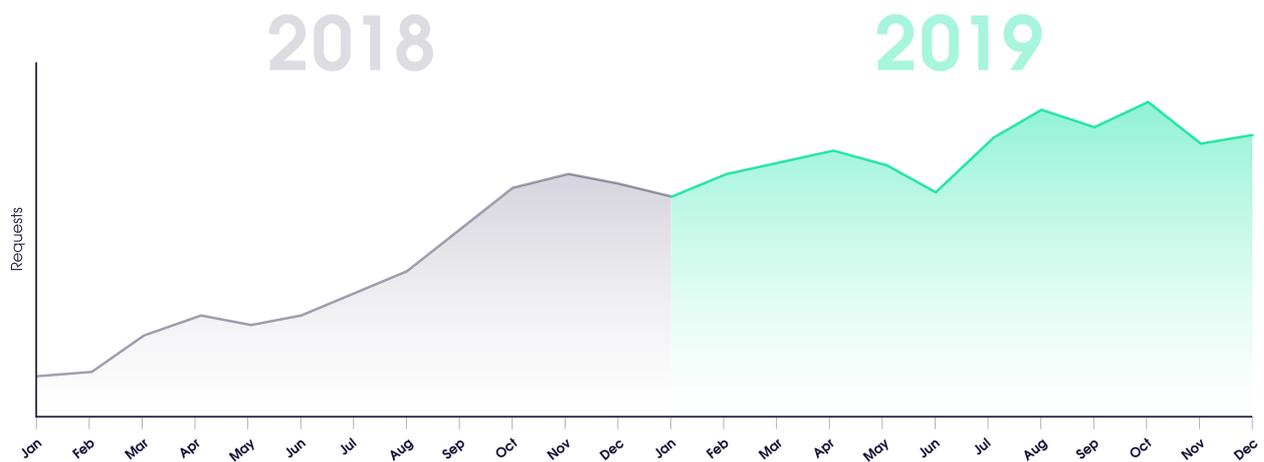
for e-commerce websites

The impressive growth of search engine, and especially e-commerce website requests indicate that the demand for **SERP** and **pricing data** is increasing due to this intelligence becoming more valuable for the development of business. As such data is time-sensitive, and the competition is continuously growing, the best chance to stay one step ahead is by applying real-time scraping strategies.

By using Real-Time Crawler, our clients are increasing the efficiency of data collection and significantly reducing their spending on resources. This enables companies to focus on data analysis rather than data scraping and thus overcome the market competition.

Most popular months

As we can see from Real-Time Crawlers' request graph below, there was a substantial number of requests processed via Real-Time Crawler from 2018 to 2020.



Real-Time Crawler requests by month, 2018-2019 data. Source: Oxylabs

There are two major spikes that should be taken into account. One was around mid-July and the other - in August. The former might have been caused by e-commerce retailers launching their sales events, similar to those of Black Friday.

As for the spike in August, it most likely occurred due to the summer holidays ending and the new school year beginning. A prime time for parents to stock up on uniforms, stationery, and other school year essentials. Real-Time Crawler was used to scrape the competition for retail sites to gain competitive pricing intelligence and provide their customers with the best deals.

However, the highest number of requests can be seen in October 2019. This increase is most likely due to the biggest event of the year in Western countries - Christmas. It is widely reported that the business preparations for the biggest season of Christmas shopping start in Autumn. Hence, retail sites also begin getting ready.

Other massive retail events that need to be taken into consideration together with Christmas were Black Friday and Cyber Monday. These cyber holidays undoubtedly impacted the request spike seen in the graph above.

These request spikes reappear annually, and each year the number gradually increases, so it is safe to assume that they are to be expected in 2020 too, with additional growth likely.

The future of web scraping tools

With Oxylabs' Real-Time Crawler, we continue seeing the growth of requests rise and slowly stabilize, indicating the **overall demand for automated web scraping tools in the market**.

The clearly favorite use cases for Real-Time Crawler continue to be **e-commerce websites** and **search engines**, confirming that the demand for SERP and pricing data is becoming crucial for business development.

In the words of our client, **Aurélien Remy Dionisi, the CEO of Digidian**, Real-Time Crawler proves to be a **"Very good alternative to proxies"**, he continues:

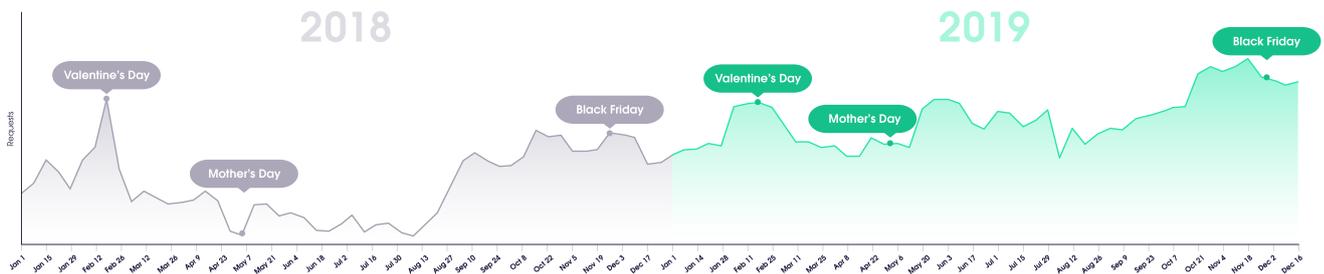
We have replaced our old proxy-based infrastructure with the Real-Time Crawler solution provided by Oxylabs. The service is very fast and reliable. The results we obtained by now are far better than the old ones. We are very happy with the product, and so are our customers.

Web Scraping Trends Using Data Center Proxies in E-commerce Market: 2017-2019 Case Studies

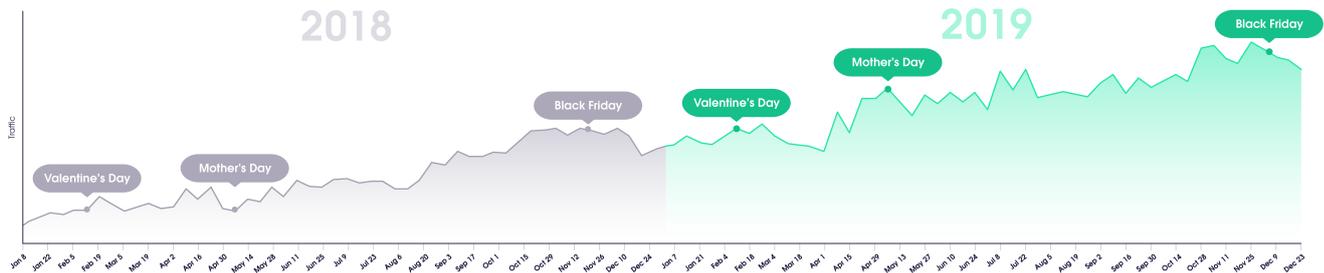
Businesses from the e-commerce market choose data center proxies to carry out various data gathering operations. They are regularly used for market research, brand protection, pricing intelligence or SEO monitoring, just to name a few use cases. Data center proxies are more affordable to buy in bulk for collecting vast amounts of relevant data from all types of sources. Although data center proxies are easier to block, this risk can be mitigated with IP rotation and other protective strategies.

Over the years, while following the web scraping trends of e-commerce market players, we have noticed some changes in scraping behavior. In 2019 more careful and diligent web scraping practices were abandoned by some companies, most likely due to their success the year before.

Data seems to suggest that these companies assumed that the strategy of using more proxies was unnecessarily careful and they then experimented with a different approach that ultimately resulted in much higher block rates compared to prior years.



Data center proxies' requests by week, 2018-2019 data. Source: Oxylabs



Data center proxies' traffic by week, 2018-2019 data. Source: Oxylabs

Looking at both of these graphs above, in the case of Valentine's Day, for example, even though the request growth in 2019 was significant during the time around the event, the spike in traffic was not as big. From this we can assume that in 2019 the clients, based on prior experience, optimized their data gathering procedures, requiring less data to extract relevant insights.

Nevertheless, we can see that both for Mother's Day and Valentine's Day, retail sites intensified their traffic to prepare for both events in 2019. An especially significant rise of both requests and traffic is seen before Black Friday, indicating the extensive preparations made beforehand.

Case study 1: Valentine's Day

The first big event that hits the e-commerce market every year is Valentine's day. Days prior to February 14th, e-commerce websites start preparing in advance by scraping relevant sources.

For example, in 2018 on Valentine's Day, clients **increased** their proxy **resources by an average of 13.03%**, but they also increased **traffic per IP by 27.71%**, which likely led to **12.14% higher block rate** on average.

On Valentine's Day 2019, Oxylabs e-commerce clients decided to increase their web scraping **resources by just 6.79%** and raise the **traffic per IP by 26.61%**, which might have led to **51.65% higher block rate**. This block rate was **more than 4 times higher** than in 2018 - a consequence of an insufficient amount of data center proxies used prior to the event.

51.65%
higher block
rate

Case study 2: Mother's Day

The second huge e-commerce market event of the year - Mother's Day. Looking at choices made both in 2018, and 2019 - the same mistakes were made as on Valentine's Day.

In 2018, on Mother's Day, companies managed even to drop the block rate **by 9.34%** while still increasing their **traffic per IP by 21.83%**, most likely because they expanded their data center proxy resources **by an average of 32.3%**.

In 2019, however, our clients sent **72.86% more traffic** per IP despite the fact that the data gathering resources **were only increased by 14.37%**, which probably led to a **25.41% higher block rate**.

25.41%
higher block
rate

Case study 3: Black Friday

Last but not least, the biggest e-commerce event of the year - Black Friday. Every year this event keeps growing, but in 2019 it became the busiest day for online shoppers in the U.S., attracting over **93M people**, thus taking the crown from Cyber Monday (±83M).

93M
people

Looking through the history of Oxylabs' client web scraping trends, it is easy to conclude that the e-commerce market takes this event very seriously.

On Black Friday that took place on November 24th, 2017, Oxylabs' clients generated **40.9% more traffic per IP** than usual, which most likely resulted in **the average block rate jumping from 3.91% to 14.89%**. Prior to the event, the web scraping **resources were increased by an average of 7.22%**.

14.89%
average block
rate

On November 23rd, Black Friday of 2018, preparations started earlier. Furthermore, before the event, our clients **increased** their scraping **resources by 29.12%** and spread their data collection operations through a couple of days. This way, on Black Friday, they generated 31.8% more **traffic per IP** than usual and raised the **block rate by only 7.15%**.

This brings us to November 29th, Black Friday of 2019. Companies started their web scraping operations a week prior to the main event by **increasing their resources by 12.67%** and spreading out the traffic through that whole week. On Black Friday, the **traffic per IP** was only raised **by 9.7%** and the **block rate** remained almost the same (**+1.22%**).

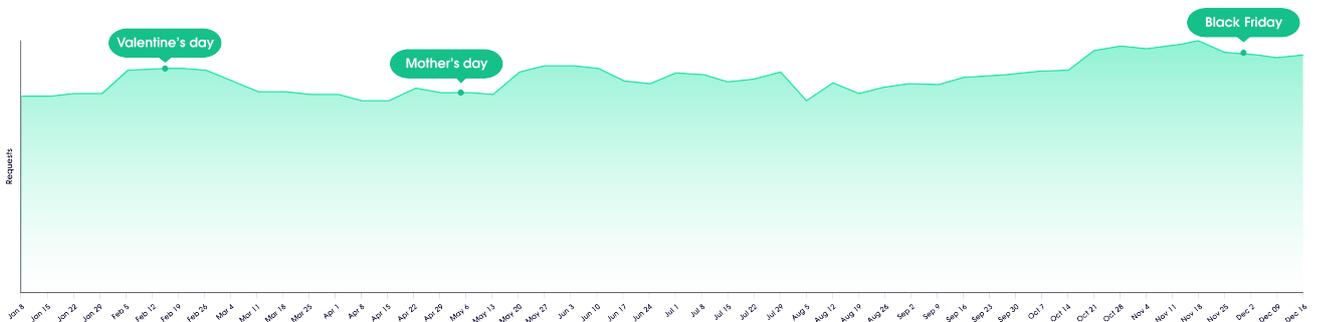
+1.22%

minor block rate increase

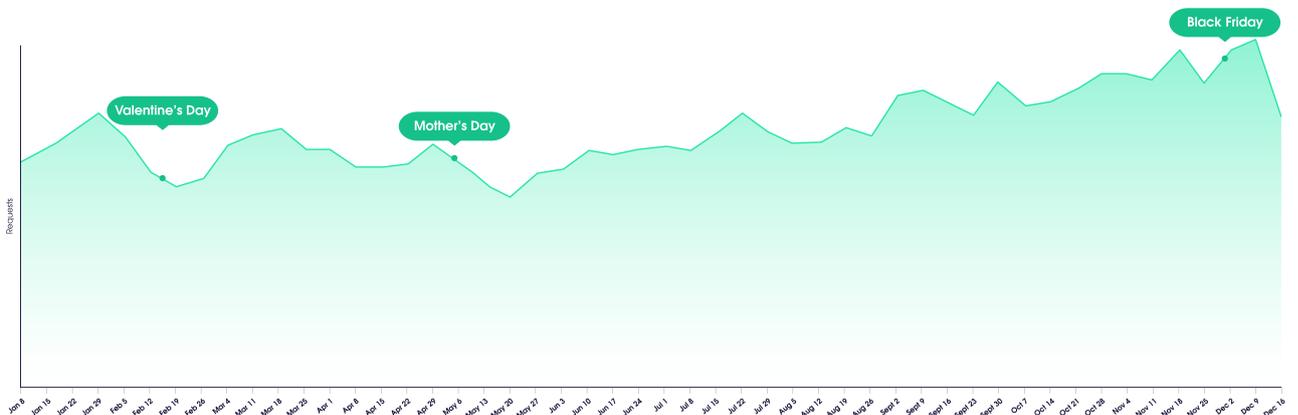
Lessons learned

All in all, what is evident from the case studies analyzed above - to carry out successful data-gathering operations with data center proxies, businesses ought to prepare in advance. Not only is it essential to spread out more intense web scraping through multiple days leading up to the event, but, also, to prepare better by upscaling data center proxy resources so that the significant increase in traffic could be handled without issues.

Going further, we present graphs indicating the request and traffic volume forecasts for 2020. The estimates were made on the assumption of similar growth continuing into 2020.

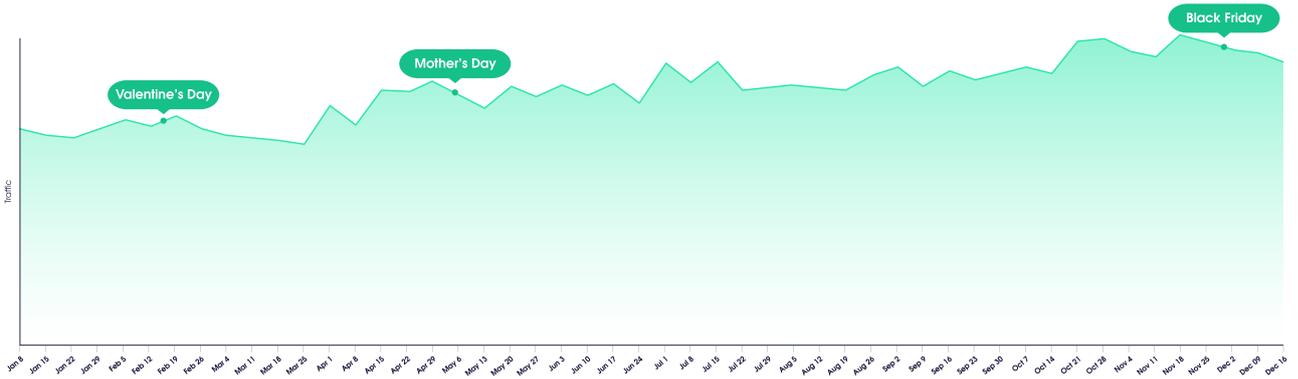


Data center proxies' requests forecast for 2020 by week. Source: Oxylabs

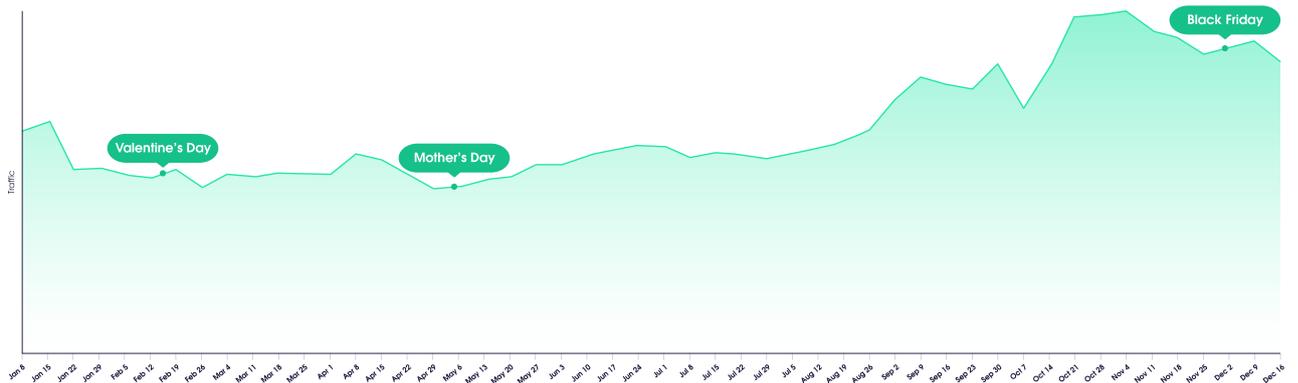


Residential proxies' requests forecast for 2020 by week. Source: Oxylabs

Unsurprisingly, our Data Analysis Department predicts that similar growth will happen around significant events such as Valentine’s Day, Mother’s Day, and Black Friday. Assuming similar scraping behavior, we will see more requests before, during, and after Valentine’s Day. With Mother’s Day, it is likely that the request rate will once again grow after the event, whereas for Black Friday, it is safe to assume that e-commerce market players will once again prepare in advance and start their operations before the major event.



Data center proxies' traffic forecast for 2020 by week. Source: Oxylabs



Residential proxies' traffic forecast for 2020 by week. Source: Oxylabs

As was the case with the forecast for the number of requests for the year 2020 for both data center and residential proxies, we predict similar growth trends for traffic volume. The only exception can be predicted in the case of Valentine’s Day, where the traffic in 2019 was not as significant as with the other two events, and we believe the same will repeat in 2020.

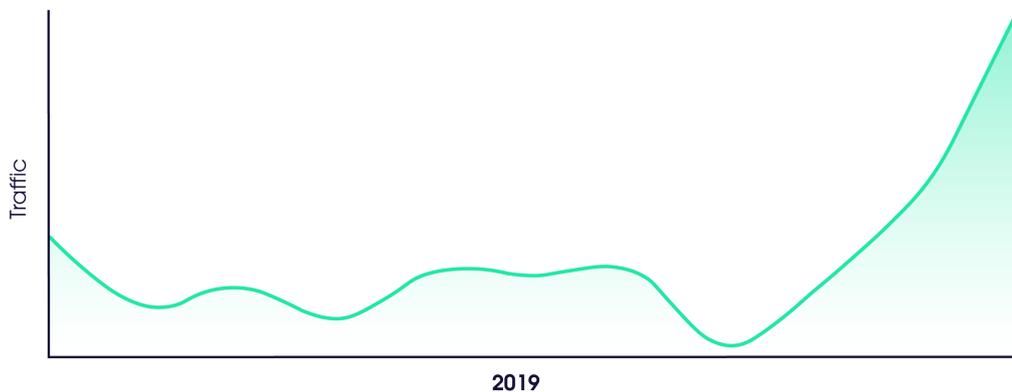
Emerging Markets of Asia and Northern America: Web Scraping Activity Rises

The year 2020 will be challenging for the global consumer markets, according to the ["Industries in 2020" report](#) by the Economist. In volume terms, growth will slow to 2.2%, compared with 2.5% in 2019. That said, the volume growth is anticipated to be led by Asia, and Oxylabs Data Analysis Department's research indicates growth in web scraping operations for this particular region.

Asia

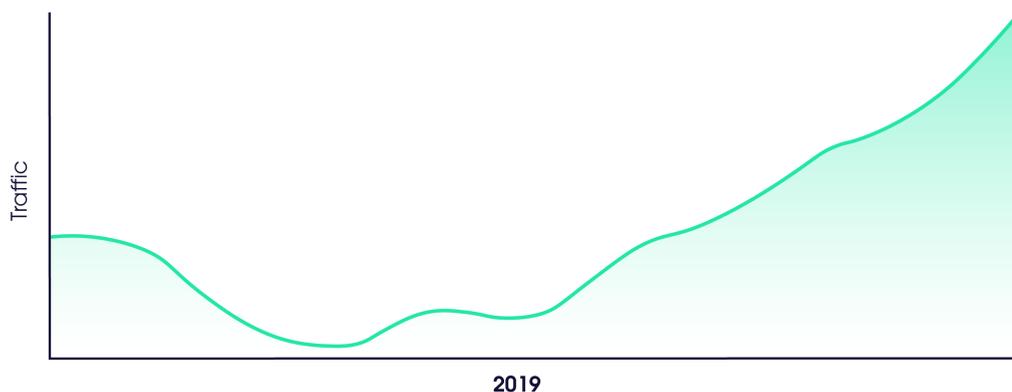
A conducted research by the Data Analysis Department revealed that Oxylabs' clientele located in Asia are steadily growing their web scraping operations via both residential and data center proxies' overall usage.

On the subject of data center proxies' traffic, Oxylabs' clients located in Asia are recording on average **12.5%** growth from late September, as the in-depth analysis revealed.



Data center proxies' traffic growth in Asia, 2019. Source: Oxylabs

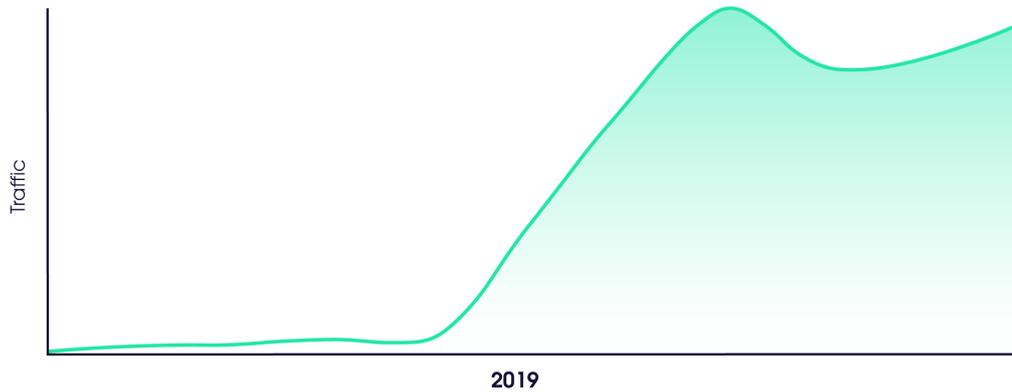
When it comes to the residential proxies' traffic for web scraping operations, clients located in Asia are recording a steady **20%** per month growth since the beginning of August 2019.



Residential proxies' traffic growth in Asia, 2019. Source: Oxylabs

North America

A conducted analysis by the Data Analysis Department revealed that Oxylabs' North America customer base recorded a sharp **35.96%** growth in data center proxy traffic, compared to August when traffic was well established.



Data center proxies' traffic growth in North America, 2019. Source: Oxylabs

With that said, while Oxylabs' North America clients have indicated positive growth in data center proxy traffic in **early 2019**, the growth significantly slowed down later on in the year. Presently, a slight traffic increase has been observed since the beginning of November 2019, yet the Data Analysis Department is not forecasting any significant growth soon.

It is worth mentioning that in the case of both Asia and North America, the significant growth in data center proxies' usage for the web scraping tasks began in the **August-October period**. This indicates early preparations in external intelligence gathering and the placement focus in these regions for the year 2020 by Oxylabs' clients.



Thank you

Our team of over 120 people would like to thank you for taking the time to go over our extensive web scraping trend report of 2020. Here at Oxylabs, we pride ourselves on providing the right scraping tools to businesses, as we believe everyone has the right to access public data and strive to achieve a competitive advantage in their respectful markets.

If you have any further questions or comments regarding our 2020 trend report, or would like to inquire about our products, feel free to contact us at marketing@oxylabs.io.