



Extensive web scraping experience

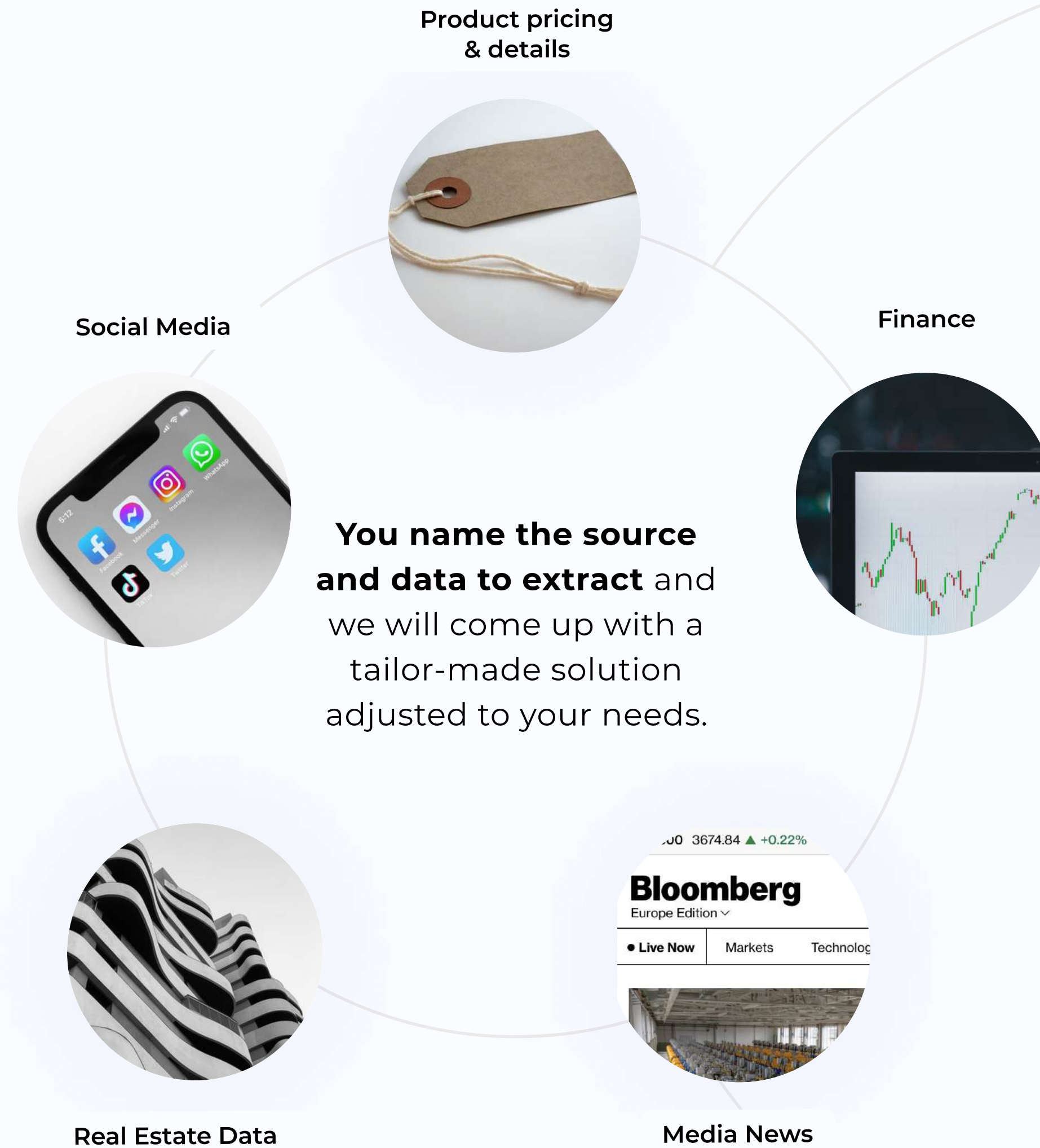
DevsData LLC
1820 Avenue M, Suite 481
Brooklyn, NY 11230, USA

(917) 300 1752
general@devsdata.com

You can find us here:
devsdata.com

We offer **premium** technology services for business

DevsData LLC is a boutique software & recruitment agency, with Google-level engineers and a vast network of senior expert contractors.



“DevsData demonstrated a strong degree of proactivity, taking time to thoroughly understand the problem, business perspective and continuously suggesting performance and usability enhancements. I’ve worked with them on numerous projects over the last 3 years and I’m very happy.”



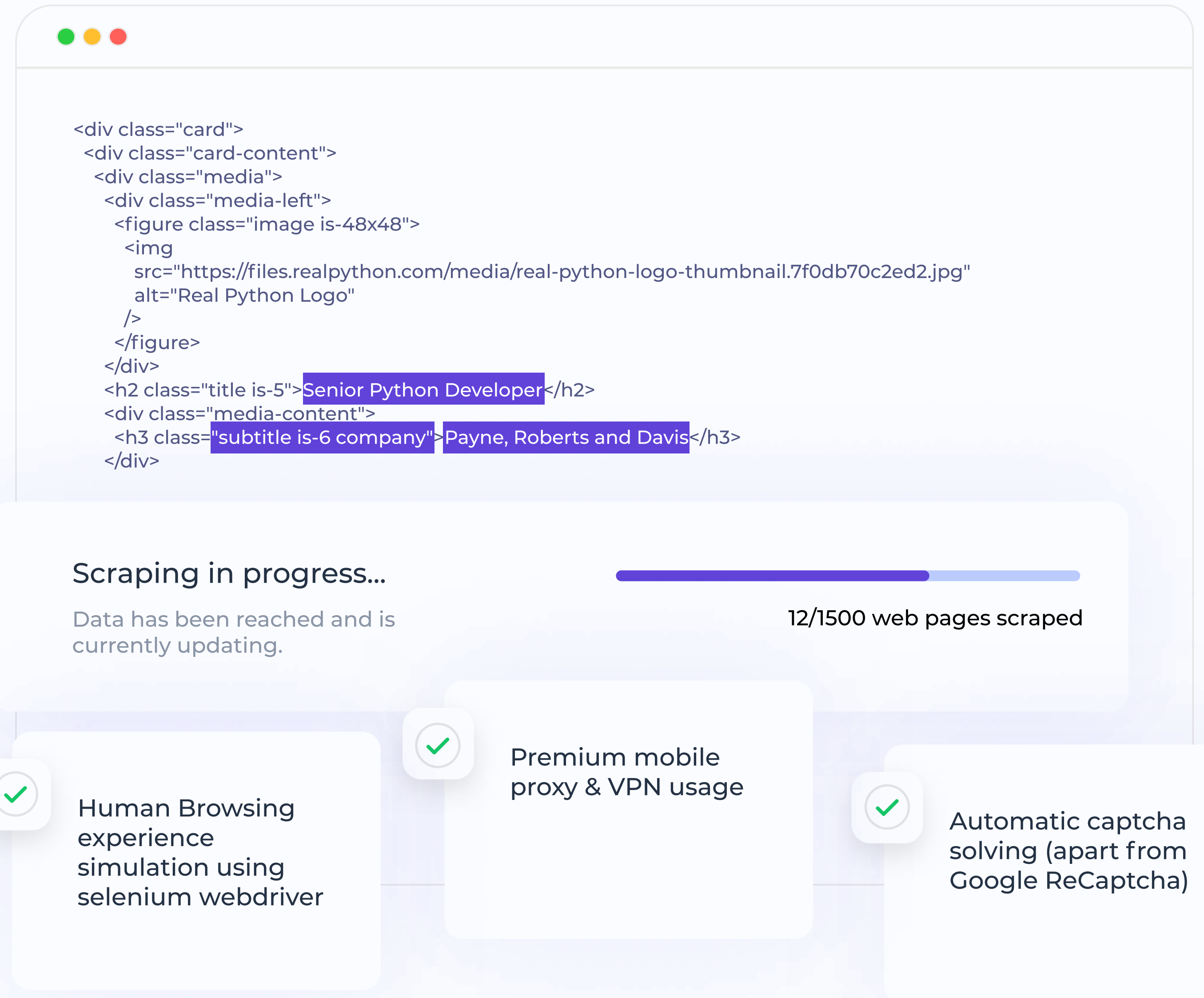
Jonas Lee, Partner & Executive Vice President of Verus Financial LLC

VERUS™

Battle-tested at **web scraping**

Our engineers have an in-depth understanding of complex databases and broad experience in processing them.

We are able to extract data from multiple challenging sources, even scrape-proof websites. To achieve such results, we use the most advanced tech solutions such as:



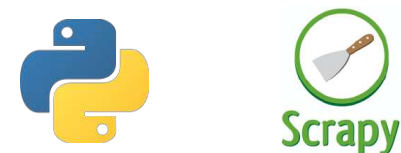
Throughout the years our engineers have gathered extensive hands-on experience



Creating and maintenance of several small scrapers on short notice

We designed and maintained several small scrapers for a company project on short notice. Our task was to extract the data as quickly as possible and filter it to obtain only the essential information. **One of the projects was a Natural Language Processing scraping engine for a London-based hedge fund** - it scraped and scored news articles based on precise criteria given by the client.

Technologies:



Scraping and processing confidential data

We created a scraper for a US-based client. **It required as few requests as possible to collect responses** for 300m SSNs under the protected form on the website. The obvious choice for scraping technology was the low-level request package. The system was set on ten small machines on a Google cloud.

Technologies:



Scraping data stored deep inside HTML

We worked on extracting data from **Filmweb - the second biggest movie database in the world**. It required as few requests as possible to collect all data about every movie/TV series on the website. The data was stored deep inside HTML. Beautifulsoup was used to collect essential information and parts of the website.

Technologies:



Communication is the key

We always make sure to be on the same page with our clients as we strongly believe that communication is the key to fruitful cooperation.

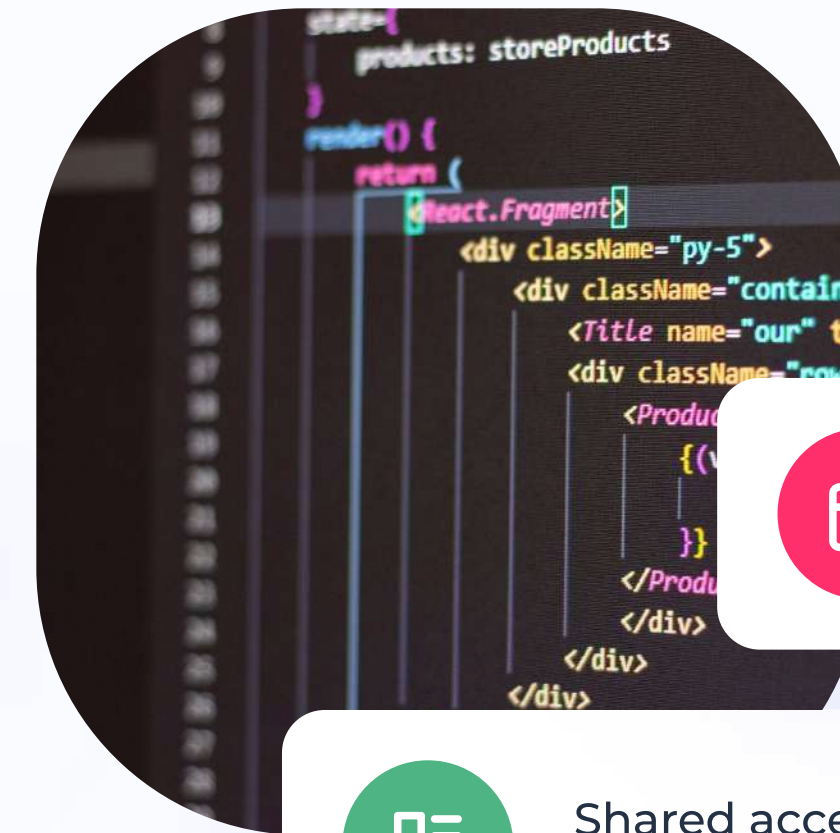
Most of our specialists work remotely from our European office, however, we are open to permanent, cross-border relocation of selected engineers. For longer projects, we usually start full-time engagement with 2 weeks of onboarding, locally at the client's office.



Regular video calls
(twice a week)



Frequent in-person meetings



Project planning and
grooming every week



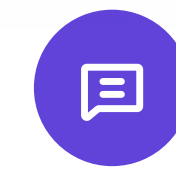
Shared access to a project management
tool and code repository



Time overlap with
the eastern time zone



Dedicated Slack Channel



Engineers and PMs replying instantly
to messages and emails



Our way of working is very transparent.
We plan everything in advance and all
the planning is visible in our PM tool so
that the client can access it anytime.





Scraping unstructured data from Wikipedia

We created a scraper running on Wikipedia to collect a data set regarding movies/television series and their cast. The biggest threat in this project was that the website was non-structured, so links to other subpages could have been located everywhere. Scrapy, which memorizes visited subpages and schedules pages to visit, was the most efficient technology to use.

Technologies:



Boosting the efficiency of an existing scraping engine

We took part in the maintenance and modification process of the company's scraping engine. It was responsible for collecting profile data about people and companies from about ten confidential sources. The data had been purchased before, so our task was to collect what was either not yet available to buy or to extend the possessed data.

Technologies:



We took part in the maintenance and modification process of many scraping engines.



Gathering data from numerous websites

Our client needed to collect data on clothing products, with the main focus being their categorization and prices. There were about 30 websites with varying depth of information and protection against scraping.

Technologies:



More opinions from clients **delighted** with our performance

// Experienced in large-scale projects

The pace was very fast, with tight project milestones and deadlines as well as quick iterations. **The machine learning approach was well thought out and the advantages, as well as drawbacks, were well-explained.** The project was a huge success.”



Angela Sy, Head of Strategy at SkyCatch, Inc.



// Flexible and dynamic

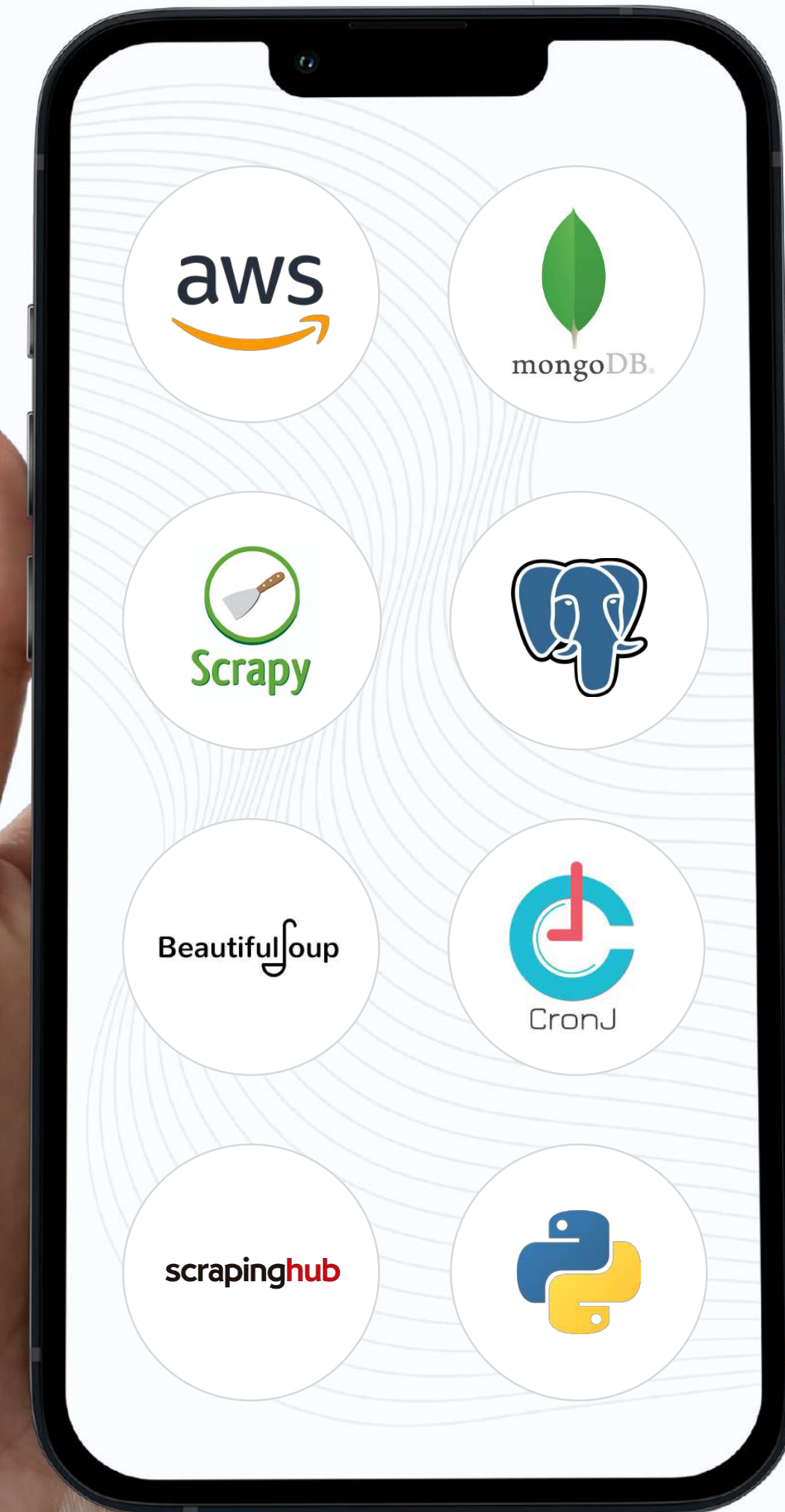
Tom runs a unique IT consulting firm, DevsData. I have the pleasure of working with him as business partners on a big-scale digital transformation project for Varner - a European fashion Group with 11 thousand employees worldwide - **they have assembled the core Machine Learning and Software team for the operation.** I can personally guarantee their dedication, ability to understand the business needs and track down those exceptional people (we're very picky) wherever they are located. ”



Nikolai Fasting, CEO of Syncano, Inc.
Executive VP of Varner digital transformation



Selected tech stack we are familiar with





DevsData LLC - a premium technology partner

- ✔ Worked with **premium & extremely demanding clients** before (hedge funds, fast-moving VC-backed startups).
- ✔ **100% client satisfaction**, 5.0/5.0 score on clutch.co.
- ✔ Strong **business & marketing** focus in addition to high engineering skills.
- ✔ **Extensive experience** in very recent technology tools, frameworks and languages for innovative software development.
- ✔ Ultra **transparent** way of work.

TRUSTED BY

VARNER

Cubus

 **SKYCATCH**

 **NOVARTIS**

REVIEWED ON

Clutch



5/5 | 31 REVIEWS



STRONG WORK EXPERIENCE FOR THE US STARTUP ECOSYSTEM

techstars_

idealab

[/] Devs_Data

Let's talk!

general@devsdata.com
(917) 300 1752

calendly.com/devsdata/hello

REVIEWED ON
Clutch ★★★★★
5/5 | 31 REVIEWS

TRUSTED BY

VARNER

Cubus

SKYCATCH

NOVARTIS

DevsData LLC
1820 Avenue M, Suite 481
Brooklyn, NY 11230, USA

(917) 300 1752
general@devsdata.com

You can find us here:
devsdata.com

