



Datamam

Building a World Inspired by Data

A Case Study:

Custom Data Scraping Software for High-Volume Contact Information Extraction

September 2021

Overview

A major player in web analytics sought our help to extract contact data from an extensive number of websites.

They aimed to provide businesses and individuals access to a rich database of updated contact information.

Datamam's solution was a finely crafted pipeline built to extract, clean, and deliver contact information from **over 50,000 domains** weekly.

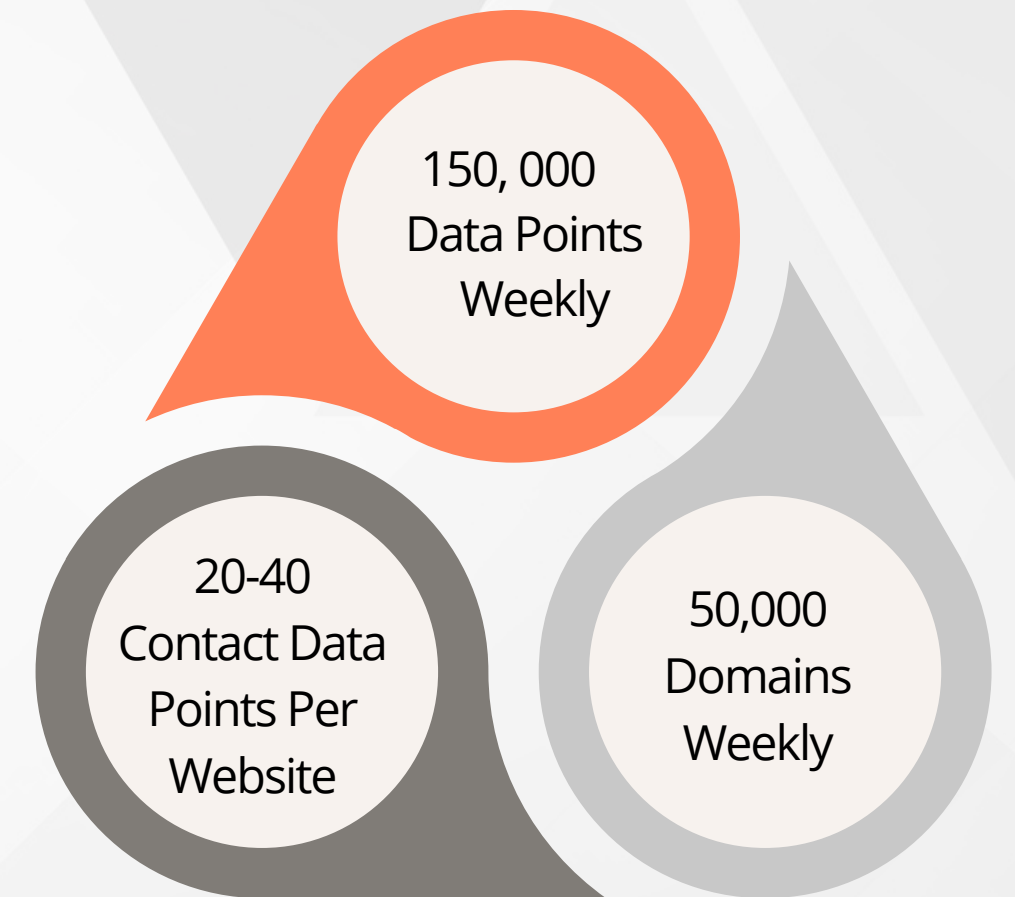
This significantly improved our client's operational efficiency and resulted in cost savings, thanks to the power of automation in data extraction.



Challenges

Our team came across numerous challenges in this project:

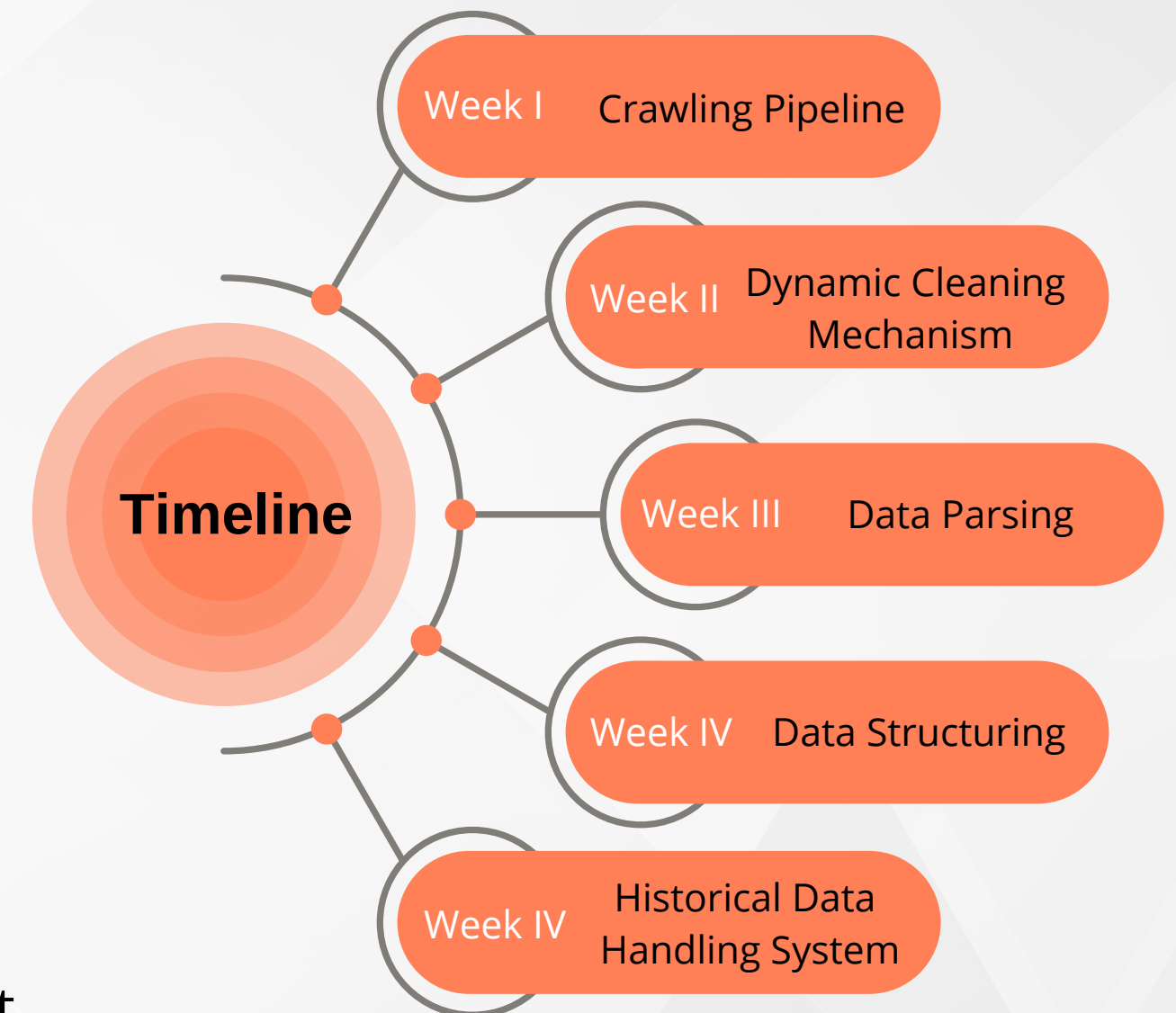
- Website Accessibility: Diverse technological implementations across domains
- Data Inconsistency: The varying structure and design of websites
- Data Accessibility: Poorly structured HTML or dynamic JavaScript elements
- Datapoint Recognition: The structure of contact information varied widely
- Historical Data Management: Need for tracking updates to contact information across time



Solution

Datamam's custom solution addressed these challenges through:

- Crawling Pipeline: Script to fetch data from various domains
- Dynamic Cleaning Mechanism: An adaptable algorithm to deal with data structures from diverse websites
- Data Parsing: System to manage the identification of contact data and ensure accuracy
- Data Structuring: Unification of data formats and automated JSON delivery
- Historical Data Handling: System to track changes and updates in contact information



Impact

Operational efficiency spiked by an estimated **60%** - By delivering structured and updated contact data, the client was able to make faster and more informed business decisions.

Cost reduction by **30%** through automated data collection - Our custom-built scraping solution significantly reduced the client's resource allocation, thereby boosting operational efficiency and data accuracy.



Cost reduction
(30% inc.)



Operational Efficiency
(60% inc.)



Key Takeaways

This case study emphasizes the significance of:

- **Custom-Built Solutions:** Tailored solutions for unique data extraction problems are key to success in web data extraction
- **Adaptability:** Flexible systems to handle changing data structures and volumes, showcasing our agility in managing data variability
- **Data Standardization:** The transformation of diverse data into a unified format simplifies data analysis and usage, adding more value to the extracted data
- **Large-Scale Data Handling:** Our proficiency in handling high-volume data extraction tasks is distinctly showcased in this project
- **Future-Proof Growth:** Our solutions are scalable and adjustable, safeguarding client investments and enabling easy future transitions

Contact



+1 650 405 0504



info@datamam.com



datamam.com



Menlo Park, CA 94025, USA

Thank You For Your Attention

