



Datamam

Building a World Inspired by Data

A Case Study:

Custom Data Scraping Solution for Procurement Notices

February 2021

Overview

The leading entity in the procurement domain, intent on centralizing and digitalizing its comprehensive tender information, approached us with the challenge of modernizing and streamlining its expansive tender data.

Their objective was to enhance the efficiency, transparency, and accessibility of their procurement offering for governments, businesses, and individuals alike.

Datamam's solution encompassed scraping, normalizing, parsing, and delivering approximately **1.4 mil data points** daily.

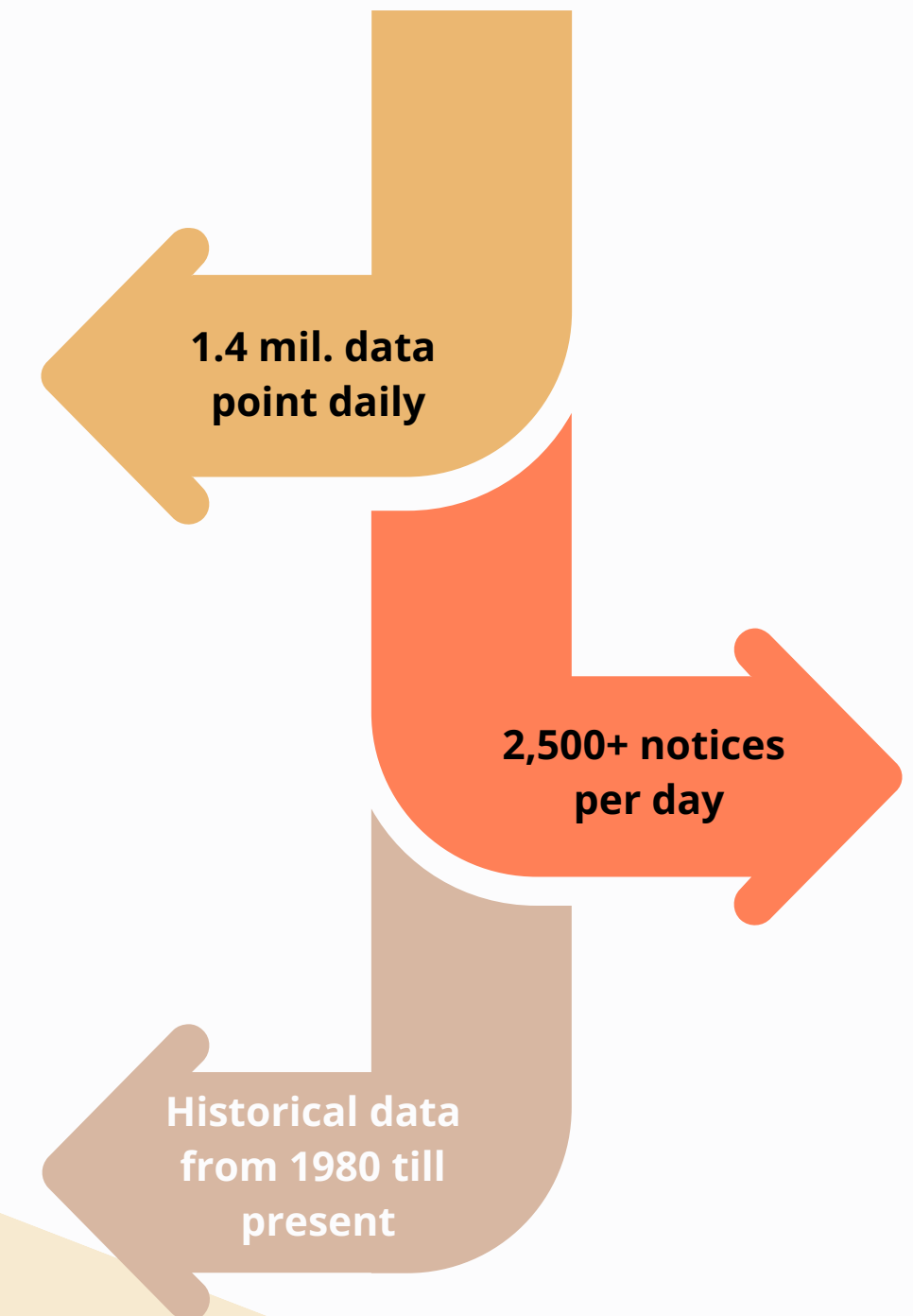
This fortified our client's position, providing an unmatched, efficient, and transparent procurement data service in the industry.



Challenges

During the project, our team faced several challenges:

- Data Accessibility: Data accessibility was hindered by an advanced anti-bot detection system
- Data Inconsistency: Inconsistent data formats in procurement notices
- Data Complexity and Volume Variability: Fluctuating volumes of procurement notices daily, different structures, ranging up to **300 data points**
- Data Format: A preference for data in Open Contracting Data Standard (OCDS) in JSON format and daily uploads to AWS S3
- Historical data: Extracting and formatting the shifting structure of historical data **from 1980 onwards**

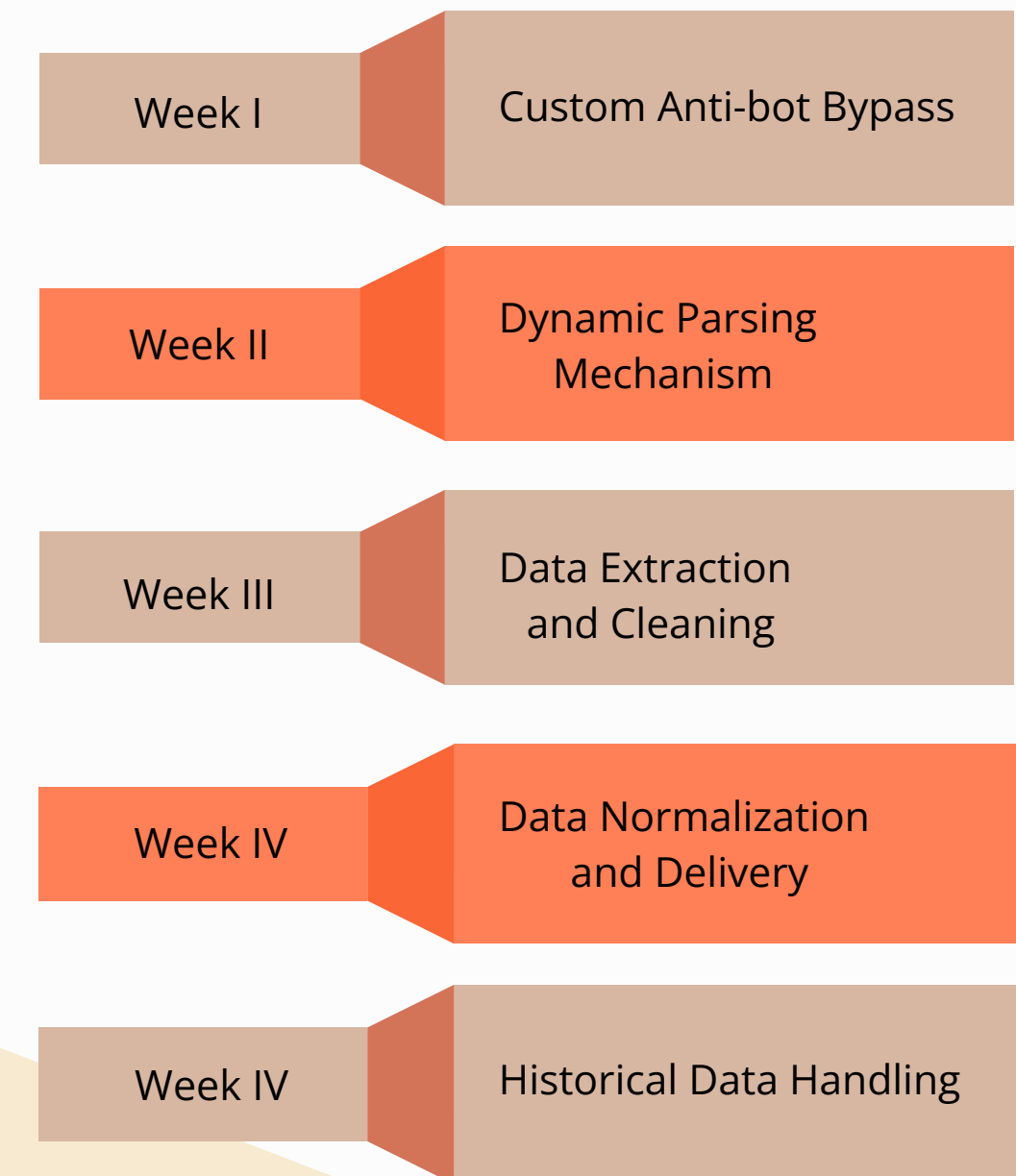


Solution

Datamam's custom solution addressed these challenges through:

- Custom Anti-bot Bypass: Mechanism to bypass advanced anti-bot detection
- Dynamic Parsing Mechanism: Dynamic parsing algorithm to adapt to evolving data structures
- Data Extraction and Cleaning: Extracted variable volumes of daily data using rigorous data cleaning and random inspection system
- Data Normalization and Delivery: Normalization of data into OCDS format and automated delivery to an AWS S3 bucket
- Historical Data Handling: Extraction, cleaning, and normalization of historical data

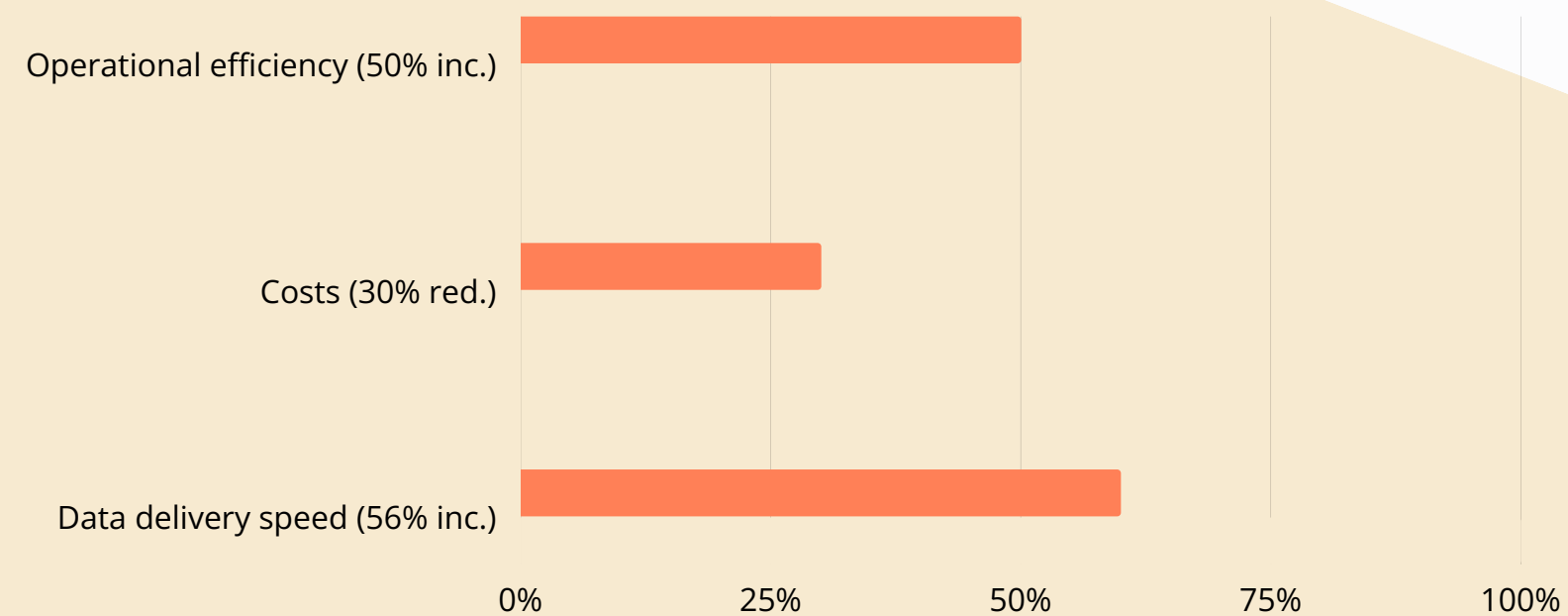
Timeline



Impact

Operational efficiency surged by **45%** with our accurate data streamlining client operations. Costs were notably reduced by **30%** due to our automation approach. Data delivery quickened by **56%**, ensuring faster and precise deliveries.

We handled a daily delivery of **1.4 mil data points**, showcasing our high-volume capability. Additionally, we made historical insights **from 1980** accessible, deepening the client's analytical reach.



Key Takeaways

The case underscores the importance of:

- **Customized Approach:** The power of bespoke solutions cannot be understated when addressing intricate data challenges. Tailoring solutions to specific client needs ensures optimized results and greater efficiency
- **Adaptive Capabilities:** The necessity to adapt to fluctuating data structures and volumes is paramount. A system that can seamlessly adjust to changing data dynamics is crucial in today's fast-evolving data landscape
- **Standardization Significance:** The critical role of data standardization streamlines analysis and ensures consistent interpretation across different platforms and stakeholders
- **Historical Data Mastery:** Proficiency in handling vast spans of historical data not only offers insights into past trends but also sets the stage for predictive analytics and trend forecasting

Contact



+1 650 405 0504



info@datamam.com



datamam.com



Menlo Park, CA 94025, USA

Thank You For Your Attention

