

## Scrape Clothing data from Fashion sites via API

**Client:** A popular Ecommerce platform from USA operating in the fashion products niche

**Domain:** Ecommerce

**Challenge:** The client wanted to [scrape clothing data](#) from around 100 fashion sites like GAP, Macys and Nordstrom. The required data points were product data along with all possible variants of a particular product like different colors and sizes.

```
02 <?xml version="1.0" encoding="utf-8"?>
03 <root>
04 <page>
05   <pageurl>http://www.amazon.com/dp/B007S0LCAC</pageurl>
06   <amazon_info>
07     <product_detail>Deuce MX Grip,Scott USA,219627-1043</product_detail>
08     <url>http://www.amazon.com/dp/B007S0LCAC</url>
09     <price>$12.95</price>
10     <asin>B007S0LCAC</asin>
11     <uniq_id>a0c661bb59736b8fe2bffd6396348016</uniq_id>
12     <product_id>10004</product_id>
13     <shipping_price>$7.18</shipping_price>
14   </amazon_info>
15 </page>
16 <page>
17   <pageurl>http://www.amazon.com/dp/B0050H9XWY</pageurl>
18   <amazon_info>
19     <product_detail>Deuce AIV Grip,Scott USA,217892-1010</product_detail>
20     <url>http://www.amazon.com/dp/B0050H9XWY</url>
21     <price>$13.00</price>
22     <asin>B0050H9XWY</asin>
```

Ecommerce sample data

The client provided us with the list of sources to be crawled and the data points required. The extraction was to be done on daily basis which meant fresh data sets have to be provided everyday. Our team set up crawlers to fetch the required data fields from the source sites provided by the client. This use case comes under our site specific crawl offering since the websites in the list had different structuring and design. The client needed the extracted data in CSV format and be uploaded to their S3 servers. The initial setup was complete in a few days and the crawlers started delivering data immediately. About 200 k records were delivered to the client during the first crawl.

### The Solution:

Client shared the list of source websites and the data points to be extracted. The frequency of data was daily, meaning fresh data was needed from all the sources everyday. Our team set up crawlers for the source websites to extract the required data fields like product name, description, specifications, price, discounts for each color and size variation. Site specific crawl was used for this since every site in the list had a different structure. The details were extracted using the specifically programmed web crawlers and delivered to the client in their desired frequency and file format directly onto their S3 locations. The data was large in quantity with 1 Million records being scraped and delivered in clean and structured format daily.

### Benefits:

- Client started receiving data in a matter of few days after the initial setup was complete
- All the technical aspects of the process were taken care of by our team
- We set up monitoring for the source websites to track changes being made to them that require modification of the crawling setup
- Large volumes of data were handled with ease using PromptCloud's advanced tech stack
- Client was able to launch their fashion app marketplace within a short period of time after starting the data acquisition project