Price Monitoring for Electronic Components Manufacturers and Distributors

Domain: Electronics

Challenge:

The client was looking to gather competition information from a list of competitors to be fed to them on a monthly basis. They were expecting to reach about 250 million SKUs a year. The data fields to be extracted were product part number, quantity breaks and the price associated with these quantity breaks, stock levels and lead times. They wanted to use this data to run their pricing strategy as well as the competitor monitoring systems. Easy access to the data was one of the main challenges that prompted them to go with a managed service provider rather than stick with their current internal web scraping setup. With their internal crawling setup, resource requirements were too high to the point of adversely affecting their core business operations. The client was particular about getting clean and ready-to-use data which can be directly uploaded to their database in order to run the comparison engine and perform other monitoring activities.

Solution:

The client provided the list of competitor sites to be crawled and the data points to be extracted. As indicated by them, we set the crawl frequency to monthly, meaning fresh data sets would be extracted on a monthly basis. Our team then set up the crawlers for the project which took just 2 days. 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 Dropbox account. Our team set up a crawler that can carry out the extraction according to the client requirements, in an automated manner. A template was also created based on the schema provided by the client and data structuring would happen as per this template. Once the setup was complete, we started delivering data to the client in their preferred format and delivery method. There was no need for any further manual intervention as we setup automated monitoring for the source sites in order to detect changes. The first data delivery consisted about 200 k records.

Benefits to the client:

- 100% API availability and continuous data feeds
- Zero data processing efforts at client's end
- Scalable infrastructure reduced client's costs
- Client's analysts only focused on querying final datasets and running analyses
- Clean and ready-to-use data
- Client could focus on their core business functions with the easy access to data