TREND ANALYSIS DATA AGGREGATOR

The Client is a New York based provider of reports on market trends as hosted service since 2004 having Fortune 500 companies from finance consultants to jobsite domains as clients.

Overview

The Trend Analysis Data Aggregator is the base component of the client's business model that involves collection of data from websites related to Job Search and related services, Real Estate, second hand Motor Vehicles sales and second hand Books sales.
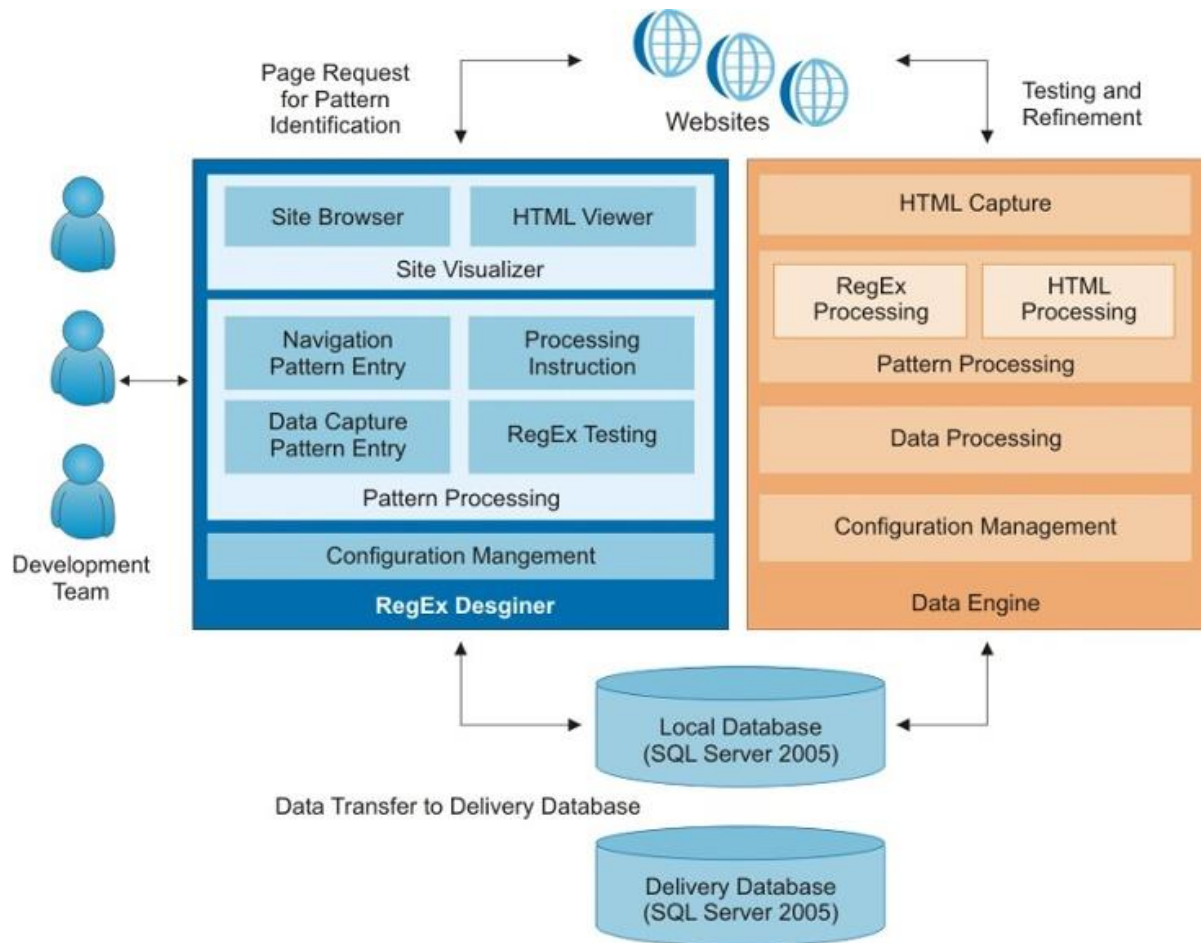
Challenges

The vision behind Trend Analysis Data Aggregator project was to outsource data aggregation from different websites to e-Zest so that client could focus on core business area of selling the data after due processing and analysis.

Solution

RegEx Designer serves as a front end for developers to develop data capture patterns for new sites as well as to change patterns for sites coming in for maintenance. Site Visualizer gets the website pages to be screen scraped and shows its HTML within the front end. Using this visualization, developers can go through the site updating data capturing and site navigation patterns and specifying processing instructions.

Patterns and instructions created by developers are tested using local database by the Data Engine. The data engine displays screen scraping results on console and errors encountered can be analyzed to test and refine URLs, patterns and processing instructions. Once the site scraping is successfully executed with local database, the patterns are transferred to client's delivery database using remote Data Transformation Services of SQL Server 2005.

Technology

Microsoft.NET, SQL Server 2005

Conclusion

Outsourcing the data aggregation function of client's business model to e-Zest enabled them to build a reliable offshore resource base specializing in screen scraping using a custom-made data engine that helped them cut costs, implement a flexible resource ramp-up or ramp-down plan based on projections of business expansions, focus their efforts on data analysis and sales, and enabled client to undertake ambitious growth strategy without worrying about operational issues related to core data aggregation function.