



Case Study



Client Profile:

Immumetrix is an early stage biotech company at Palo Alto. Their focus is on DNA in the immune system and they work on health related projects such as non-invasive monitoring of organ transplant rejection. More details about them at www.immumetrix.com

Tecnology Used:

C++, Python, PHP, HipHop PHP, Biopython, Matplotlib, Gearman and pdftk

Business Situation:

Immumetrix does analysis of DNA sequences and the immune system using multiple lab techniques. They needed to use computational techniques to analyze DNA through methods like high throughput analysis. They would analyze the DNA sequence and would represent each element through ASCII values. By studying the DNA sequence thoroughly, they were able to understand human DNA through various algorithms. What they needed was sequence alignment analysis and similarity searches in biological databases. Each sample would contain approximately 400 MB to 1 GB of data and would need customization based on the data collected. The need was to coordinate with the scientists from Immumetrix in order to understand the complexities and align the analytical algorithm accordingly. The objective of the project was to analyze the DNA data and provide outputs in CSV and graphical formats.

Solution Approach:

Immumetrix needed an application that would analyze the matches in the DNA sequences and produce the output of these analysis as charts. The studying of DNA and matching of certain sequences helped to identify if anti bodies were present and hence judge the possibility of certain types of diseases.

Compassites did an initial data analysis of the data that was currently in the system. We did multiple reads of barcodes, isotypes and primers. We aligned sequences using concepts such as local alignment and dynamic programming. Some of the processing revealed details such as 'primary antibody repertoire' and the amino acid usage.

The tool built by Compassites was on the cloud. We created an Excel file where the scientists had to simply add their data in a DropBox. The application would go onto read this Excel sheet and get to know what data had to be processed. Post processing, the results of the analysis was dumped back into the DropBox folder. The algorithms of Immumetrix

were written in C++ and Python. The application created by Compassites generated charts using the Python MatPlotLib library. Compassites used Gearman for parallel processing of the data.

Benefits & Results:

- ❑ The early system used to take a week to do the processing of the data. The current system does it in a few hours.
- ❑ With the new system, a lot of processes can be done parallelly instead of in serial order.
- ❑ Compassites helped to improvise the code and the algorithms in order to make them run faster.
- ❑ Initially there were a lot of manual tasks to be done in order to run each program. Now a lot of the tasks are automated.
- ❑ Since the application is on the cloud, Immumetrix is also paying "by use" and is not having to pay for physical hardware and servers at their location.

Bangalore, India

IBC Knowledge Park, Tower C, 4th Floor,
Bannerghatta Road, Bangalore -560029
+91-80-4663 7200

United States

3500S, Dupont Highway
Dover, Delaware -19901
+1 408 708 9090

Singapore

Compassites Technology Solutions Pte Ltd,
International Plaza, 10 Anson Road,
#03-50, Singapore 079903
+65-67186204, +65-81574120