



Case Study



Friedrich Miescher Institute scales up bioinformatics web sites with Digipede

FMI Researchers deploy the Digipede Network to increase throughput of genome searches on demand

The Organization

The Friedrich Miescher Institute (FMI) as a part of the Novartis Research Foundation is devoted to fundamental biomedical research. Located in Basel, Switzerland, the FMI employs new technologies to explore basic molecular mechanisms of cells and organisms in health and disease. With a staff of 250 in 22 research groups, the FMI is an internationally recognized research center that has initiated key developments in molecular biology over the past three decades.

When scientists at the FMI began analyzing transcription factor searches, they quickly identified the need for better methods to visualize search results. The result was PromoterPlot, a graphical display and processing tool for identifying similarities in promoter sequences.¹ With this tool, researchers can quickly see which transcription factors are the most likely to bind to certain promoter sequences and identify other unknown sequences with matching patterns from a promoter database.

The Challenge

Alessandro Di Cara, a postdoctoral fellow at the FMI working together with Andrija Tomovic and Edward Oakeley (Head of the PromoterPlot team), quickly realized that the usefulness of this tool brought with it a need for increased performance and scalability. PromoterPlot searches a database of transcription factors and their DNA binding sequences; and while each search is relatively quick, the database already has thousands of entries, and is growing monthly. Furthermore, Di Cara wanted to expose PromoterPlot as a publicly available web site, so that researchers around the world could benefit from its unique capabilities. However, running on a single server, even a single user might wait for hours for results, depending on the size of her search. Di Cara developed some clever workarounds, allowing researchers to check back hours or days later to find their results, but he realized that this approach was too limited. He needed a way to increase the throughput of transcription factor searches. He began working on a distributed computing approach, breaking the process into multiple independent searches. He tried various approaches to handling the distribution of workload, but was unsatisfied with the stability of the system, and his application was in danger of

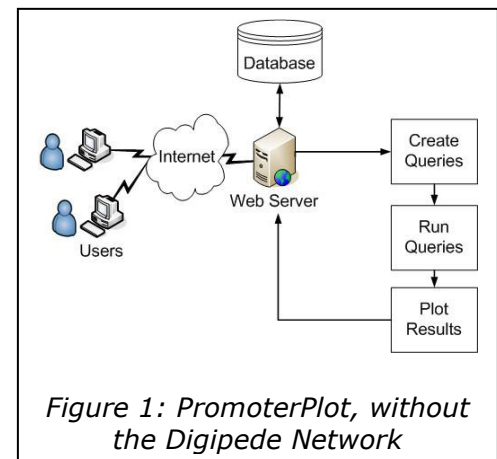


Figure 1: PromoterPlot, without the Digipede Network

"Our developers can focus their efforts on science, not grid computing."

— Dean Flanders, Head of Informatics at FMI

¹ A promoter sequence is a segment of DNA outside of a gene that controls when that gene produces one or more proteins.

becoming far too complex. He began to search for a more robust solution.

The Solution

Di Cara contacted Digipede, and attended a Webcast at which he learned how the [Digipede Framework SDK](#) could help him grid-enable his Web site with just a few lines of code. He downloaded a trial version of the Digipede Network, tested it on his application, and liked what he saw. He was developing PromoterPlot in VB and C#, using .NET 2.0, and was particularly impressed with the Digipede Network's ability to distributed classes of .NET objects. He purchased the [Digipede Network Professional Edition](#), and had a beta of his Web service up and running with a 5-node system in a few days.

"We had used an open-source grid solution when we were prototyping our application, but found that the system became unstable and difficult to manage," said Di Cara. "With Digipede, we got the stability we needed right away." "Our developers can focus their efforts on science, not grid computing," said Dean Flanders (Head of Informatics at FMI).

Because the application was exposed to the public on the FMI Web site, Di Cara and his team also needed flexible capabilities to monitor the system and kill off jobs if necessary. Digipede Control, the Web-based administrative interface, provided this capability, along with the ability to allocate multiple pools of computing resources and to add resources dynamically to accommodate increasing load.

Using a grid had other benefits as well. Running the compute intensive searches on the web server wasn't just slow for the user doing the search--it also degraded performance for every other user on the website. "By using the Digipede Network to distribute the computational load, our Web server is far more responsive to the user," added Di Cara. "We've also greatly increased the throughput of our service, both in terms of the total volume of analysis we can perform, and the total number of simultaneous users we can support. Further growth is straightforward – we can add more Digipede Agents as our needs expand, both on PromoterPlot and on additional grid-enabled applications."

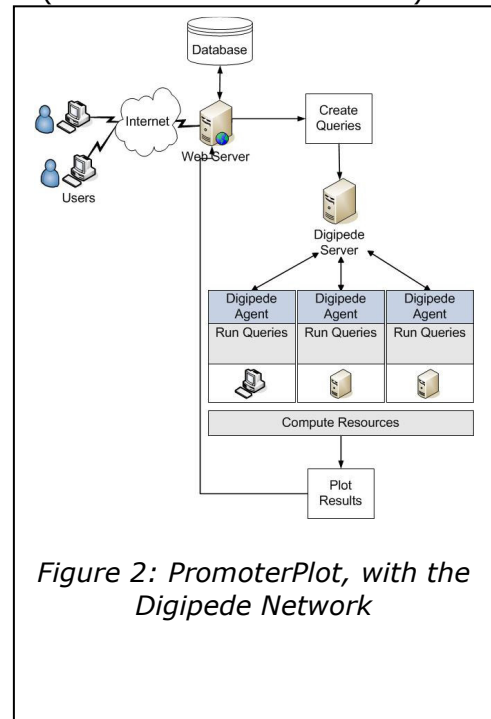


Figure 2: PromoterPlot, with the Digipede Network