

GenomePro 3.0

Student: Daniel Gonzalez, FIU

Mentors: Michael Robinson & Guido Ruiz, FIU

Instructor: Masoud Sadjadi, FIU



PostgreSQL



ubuntu



Apache



PROGRAMMING



HTML5



School of Computing & Information Sciences

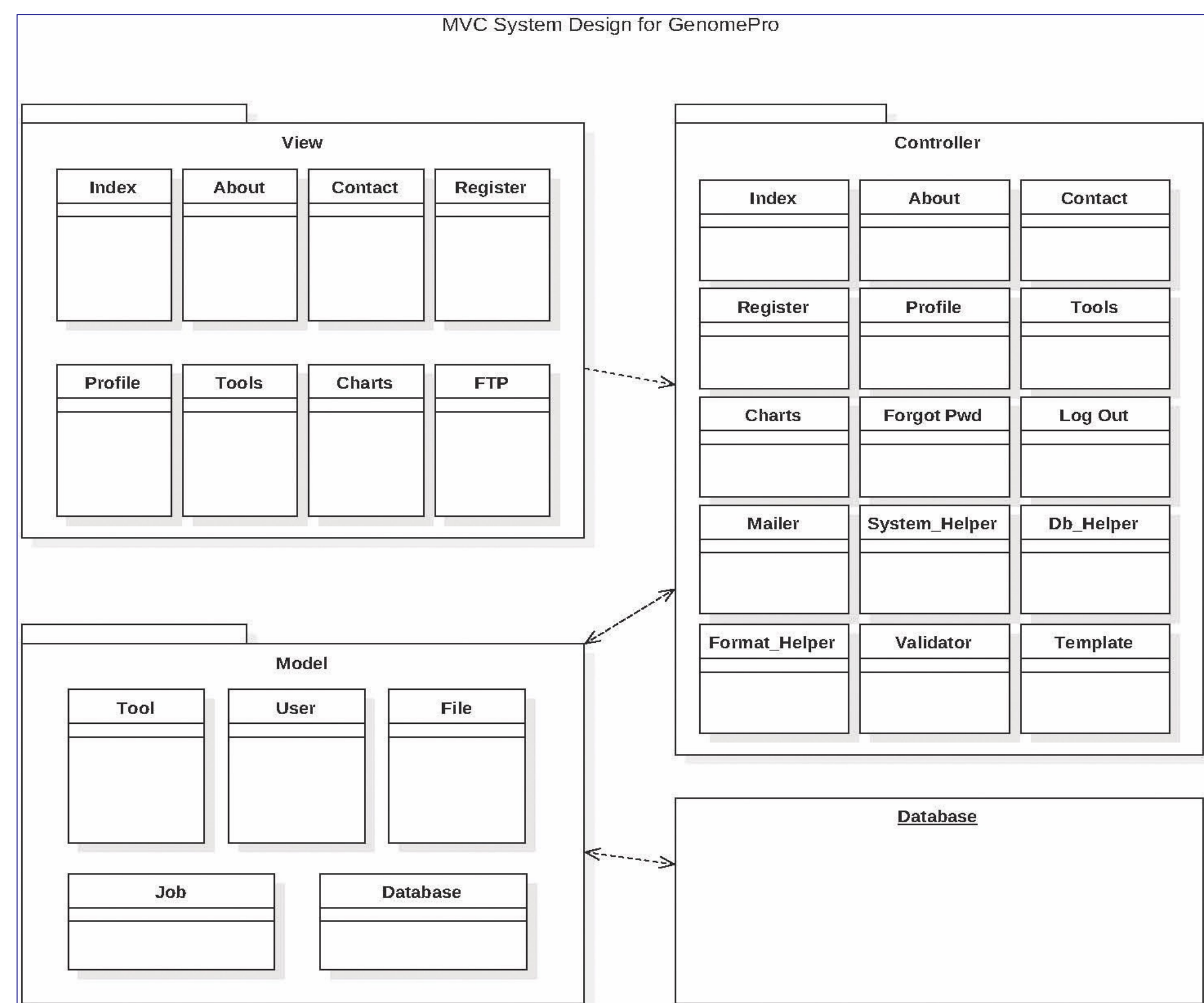
Senior Project, 2016, Spring

Problem

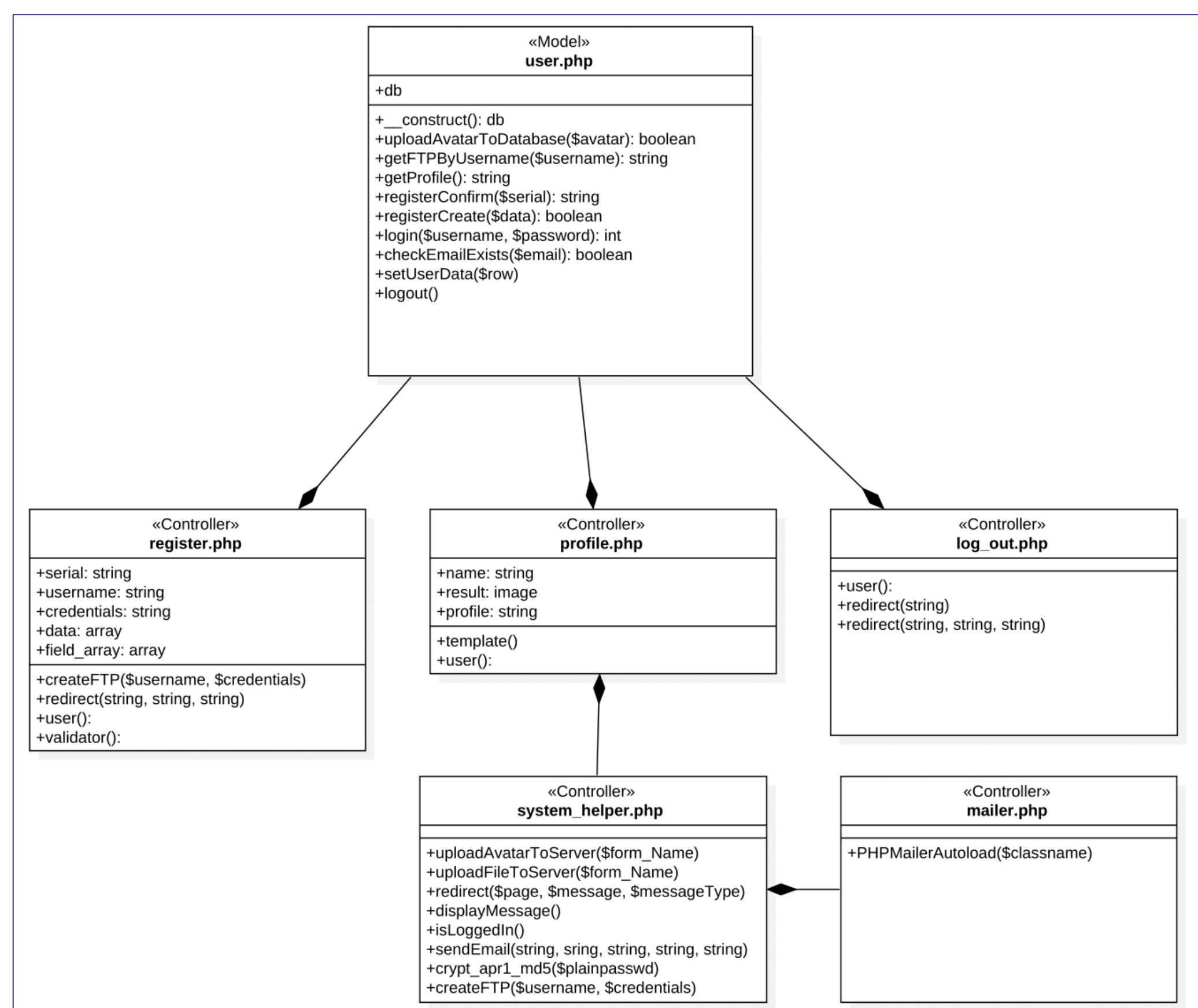
Many Biomedical Researchers today struggle to find critical patterns in genomic data indicative of life-threatening conditions like cancer. Genomic data are both plentiful and intricate; the calculations modern researchers perform are best handled by tailor-made programs with simple, intuitive, and clear interfaces.

We are developing a sophisticated web framework that can perform calculations on genomes automatically, thus automating the complicated tasks that are usually done by hand and allowing researchers to find breakthroughs faster.

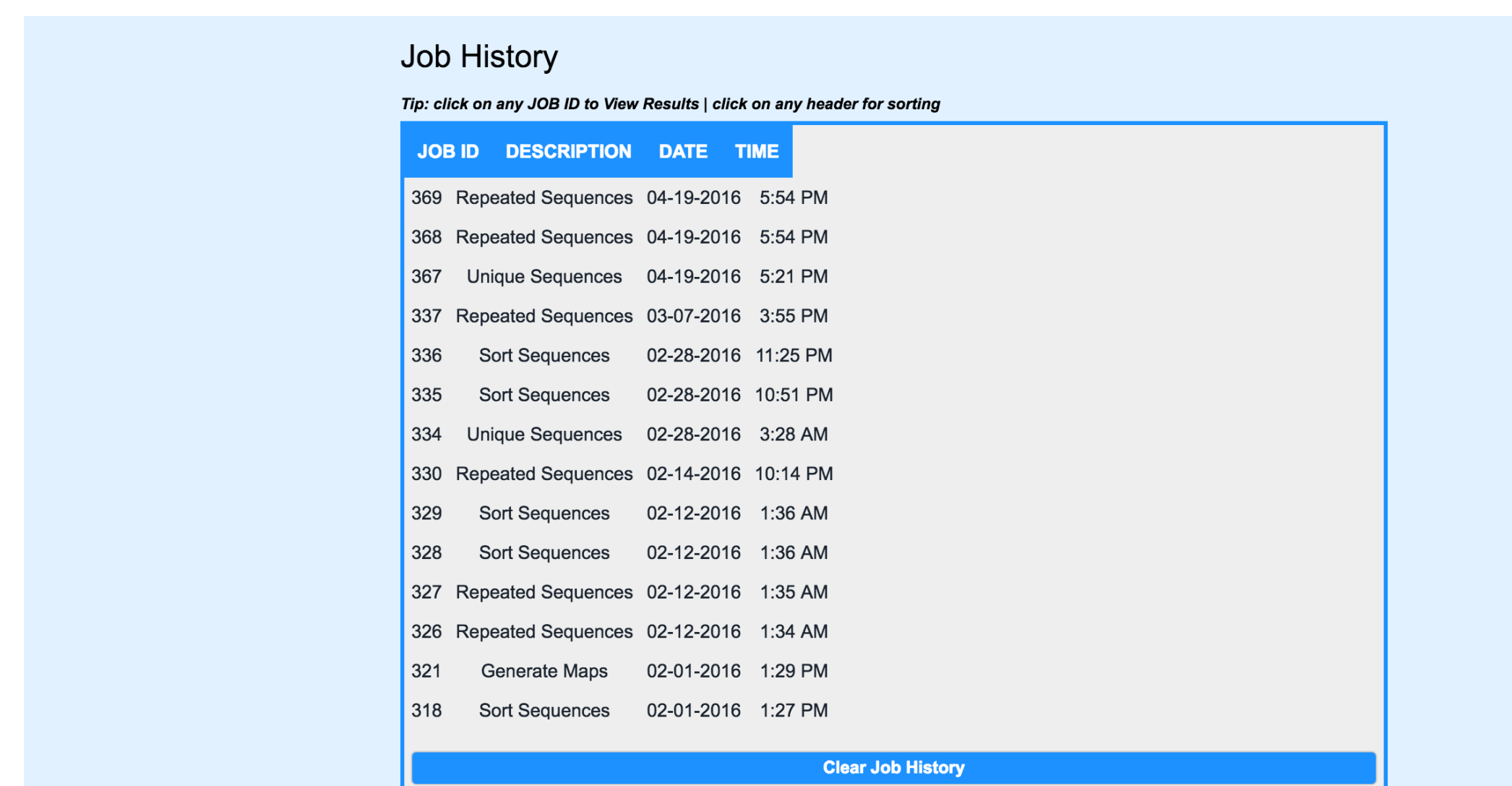
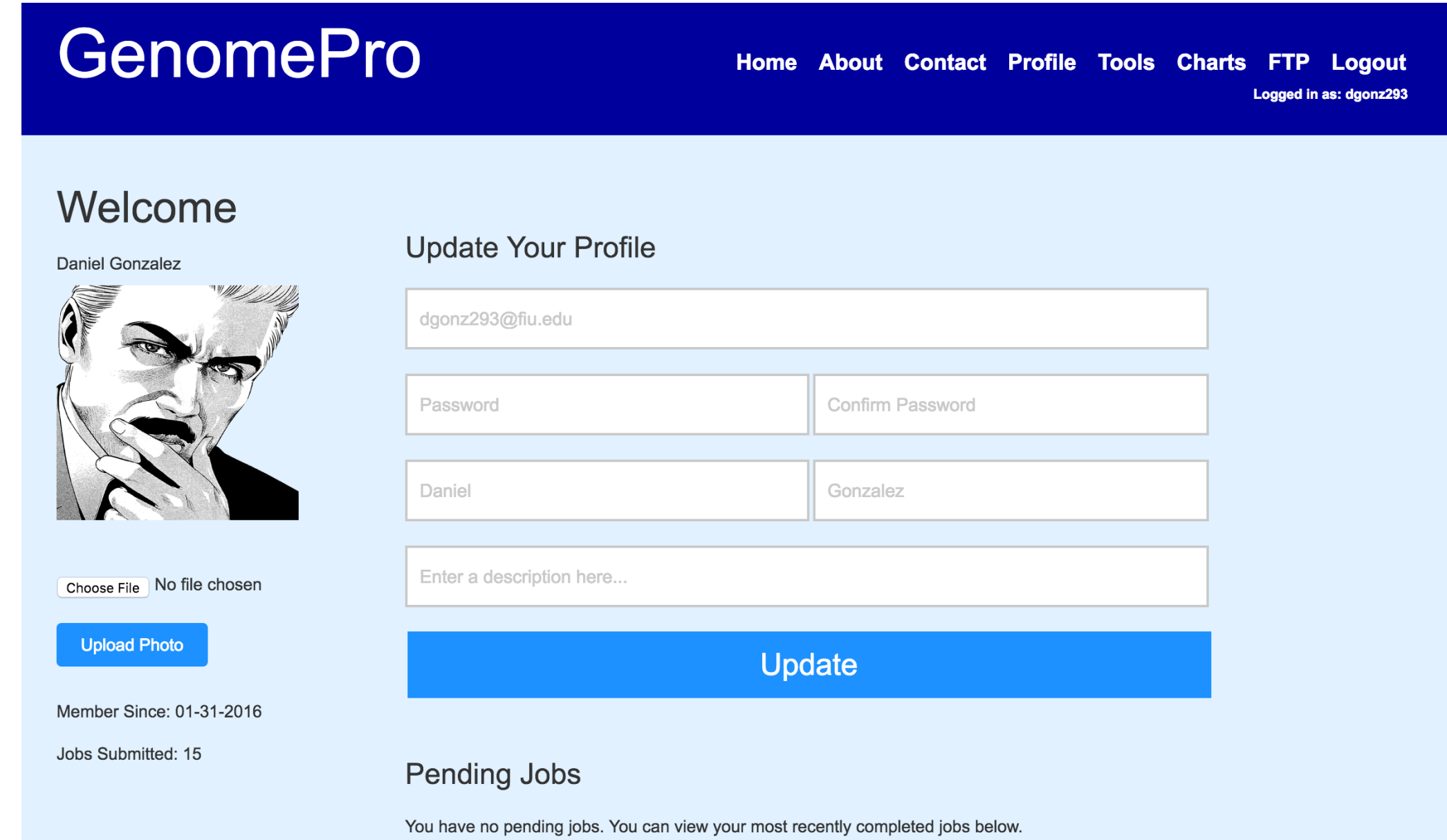
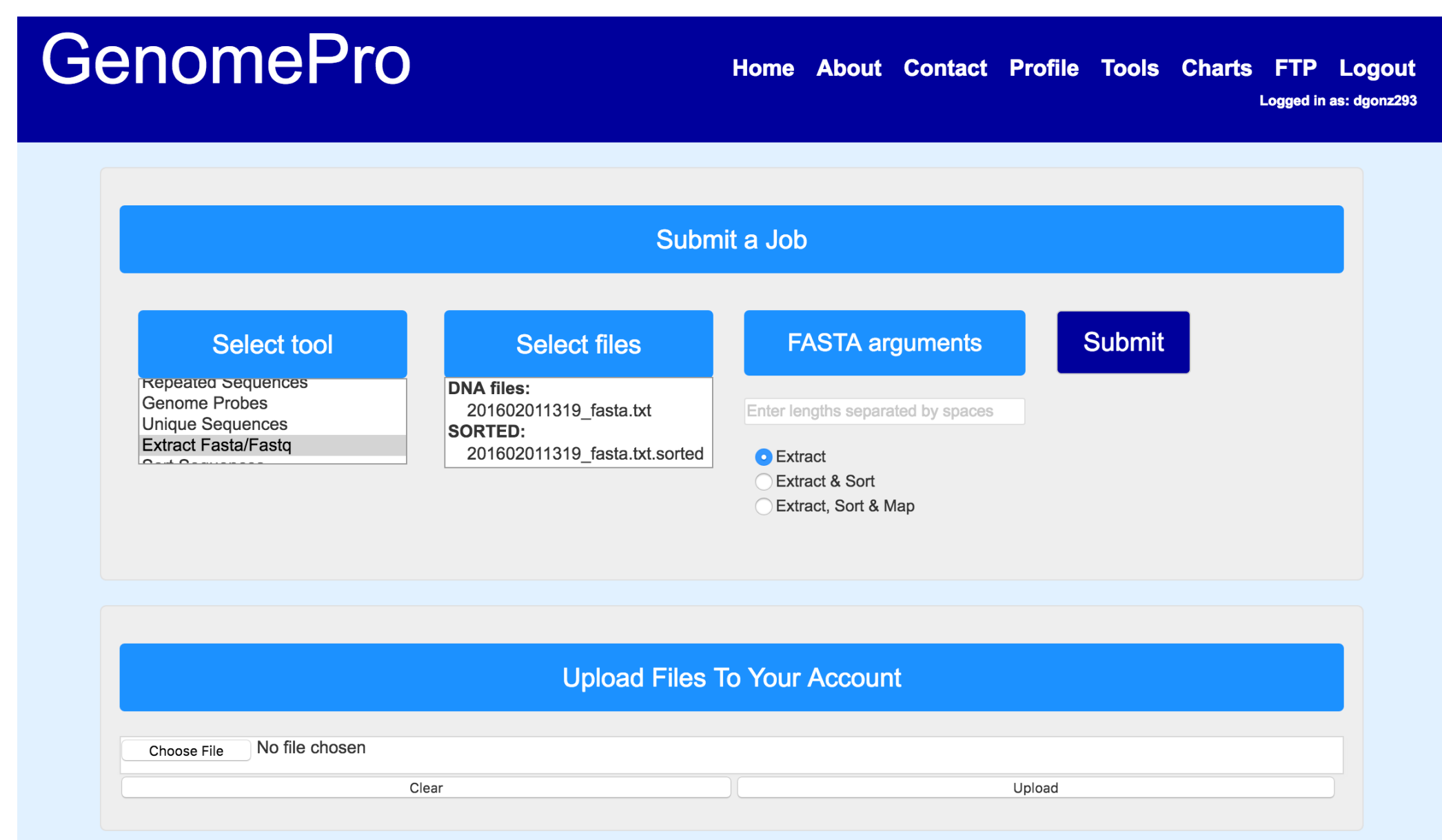
System Design (MVC)



Object Design



Screenshots

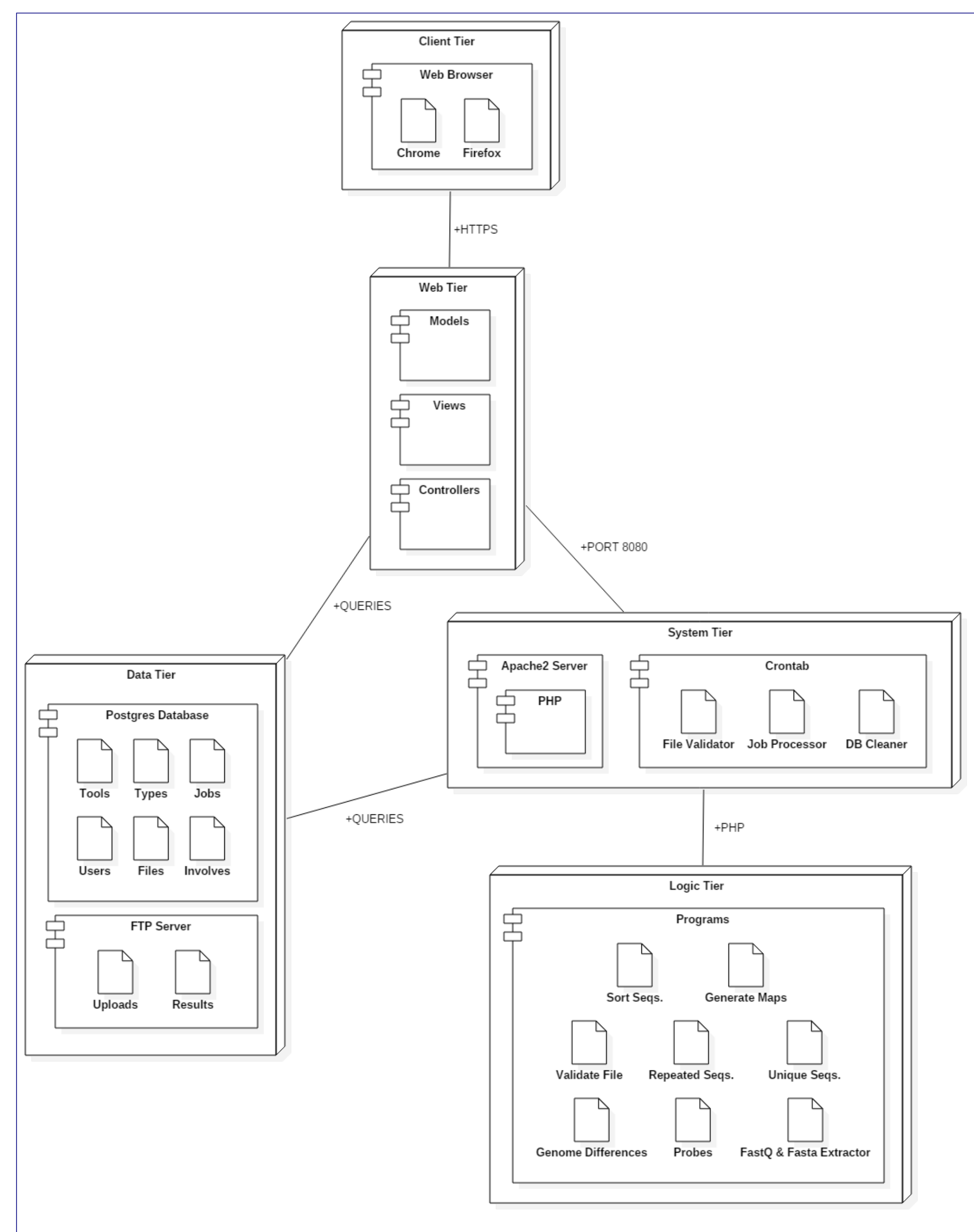


Requirements

- Prepare profiles for eventual social-media integration
- C Program: File Validator
- Tool: Genome Signatures
- Tool: Unique Sequences
- Tool: Genome Probes
- Tool: Extract Sequences
- Tool: Sort Sequences
- Tool: Create Maps
- Tool: Generate Charts
- Tool: Upload Files
- FTP results with encryption
- Save results securely
- Parallel multiple job submission

Implementation

All programs that perform calculations on genomes to produce results are implemented in **ANSI C**. Using our web system, the user can submit multiple jobs on multiple files. Every minute, new jobs are selected from the database and are processed by their respective ANSI C programs through the use of **PHP** scripts. Adding new jobs to the database after a user has submitted them are also procedures done in PHP.



Verification

Unit, integration, and validation testing was done to make sure all the elements of GenomePro 2.0 worked as intended. Extensive manual **unit** testing was done to make sure all aspects of the system, including the ANSI C programs behind the tools, worked by themselves. **Integration** testing was then done manually by ensuring that no functionality was lost or corrupted as a result of adding a new feature, such as multiple file submission.

Both **malicious and ordinary users** were tested, making sure the system was (and is) resilient against SQL injection, Cross-Site Scripting, uploading malicious files, etc.

Once we validated the functionality of the new Tools page, we looked at the system as a whole and performed **validation** testing to make sure all previous performance and functionality was still in tact. We **benchmarked** the performance of all of the tools on the Tools page, the charts on the Charts page, and the general loading times of the pages on the site. Comparing the times to previous benchmarks from GenomePro 2.0, we confirmed GenomePro 3.0 was still reliable and efficient. GenomePro understands that its audience is Biomedical Researchers, performing critical calculations, who want their results quickly and efficiently; the emphasis placed on validation is well-deserved.

Summary

GenomePro 3.0 is a sophisticated web service that hopes to be as useful to a Biomedical Researcher as a scientific calculator is to a mathematician. It provides the efficiency and accessibility that is so critical to biomedical research applications today. With features including visualization, chart generation, and a responsive job history, Biomedical Researchers can see exactly where repeats in their genomic data happen, comparing different sequences of DNA, RNA, and proteid data side-by-side. An FTP interface allows for and facilitates the storage and retrieval of all resultant files.

With GenomePro 3.0, our team hopes that Biomedical Researchers make more exciting breakthroughs to help find cures for many diseases affecting the world today, using our very secure and intuitive software as an assistant to get there. We hope that GenomePro 4.0 implements and extends social-interaction functionality so that researchers can more easily review and share results with their peers, thus pushing the boundaries of scientific and medical knowledge.

Acknowledgement

The material presented in this poster is based upon the work supported by Michael Robinson and Guido Ruiz. I am also thankful for the contributions of my teammate Roberto Arciniegas.