CASE STUDY

High-Performance Computing
Life Sciences

Managing an Ocean of Information

intel Software

Lab7 Systems optimizes BioBuilds™ tools for superior performance using Intel® Parallel Studio XE and Intel® C++ Compiler

"Intel® compilers optimize the BioBuilds™ packages for superior performance on the Intel64 architecture, including auto-vectorization and auto-parallelization for additional performance gains on modern, multi-core CPUs."

—Cheng Lee
Principal Software Architect
Lab7 Systems

Finding efficient ways to manage the massive amounts of data generated by new technologies is a key concern for many industries. It's especially challenging in the world of life sciences, where research breakthroughs are based on an ever-expanding ocean of information.

With help from Intel and Intel® Parallel Studio XE, Lab7 Systems is optimizing the open-source BioBuilds™ tool collection to make life easier for bioinformaticians, scientists, and IT teams.

Open Source Solution

Based in Austin, Texas, Lab7 Systems is a software company working to help reduce the need for hands-on data management of the massive amounts of data generated by new life sciences technologies, and to get the best possible performance from constantly evolving hardware. The company's core commercial product is the Lab7 Enterprise Science Platform™ (ESP™), the first complete sample-to-answer software platform for data-intensive laboratory operations.

Lab7 Systems also curates, maintains, and supports BioBuilds, a collection of open-source bioinformatics tools, pre-built for Linux* on both x86 and IBM Power Systems* platforms and for OS X*. By including all supporting libraries, BioBuilds eliminates the need to maintain and build custom versions of these tools—providing a path out of "dependency hell," a colloquial term for the frustration of those who have installed software packages with dependencies on specific versions of other software packages. BioBuilds also enables reproducibility by providing a reference point that ensures collaborators are all using the same versions of tools. BioBuilds uses the Conda* package manager to ensure consistent computing environments across all supported platforms. The standard BioBuilds releases are free to download and use. Lab7 Systems also offers paid options for customers needing additional support and/or custom packages.

Maximizing Performance

To maximize performance on the Intel64 platform, Lab7 Systems used the tools in Intel Parallel Studio XE. This toolset makes it easier for developers to deliver top C++, Fortran, and Python* application performance that scales on today's and next-gen processors. It also simplifies the process of creating fast, reliable parallel code.

Lab7 Systems modified its upstream build systems to support use of Intel® compilers instead of GNU Compiler Collection* (GCC*). Specifically, Lab7 Systems used the Intel® C++ Compiler to build select binaries in the BioBuilds 2017.05 release—significantly improving their performance compared to their GCC-built counterparts.

The company chose Intel C++ Compiler because of a combination of factors:

Lab7

- Standard C/C++ language support
- GCC compatibility
- Better optimization for Intel64
- Automatic SIMD vectorization and parallelization for Intel64
- Interprocedural optimization
- Profiler and analysis tools for generating and/or improving parallel code

Types of optimizations included explicit targeting of the SSE4.2 architecture ("-x"); alternative code paths for the AVX, CORE-AVX-I, and CORE-AVX2 architectures ("-ax"); optimizations enabled by the "-O3"; interprocedural optimizations ("-ip" and "-ipo"); and auto-parallelization ("-parallel").

Lab7 Systems is in the process of benchmarking several applications to quantify the performance gains from building with Intel Parallel Studio XE, which have been significant.

"Intel compilers optimize BioBuilds packages for superior performance on the Intel64 architecture," explained Cheng Lee, principal software architect at Lab7 Systems, "including auto-vectorization and auto-parallelization for additional performance gains on modern, multi-core CPUs. Additionally, these compilers generate binaries with multiple, auto-dispatched code paths that ensure optimized performance across a range of hardware."

Ongoing Support

Another key factor was Intel's support for Lab7 Systems in this open source project, which included providing both funding and Intel Parallel Studio XE licenses.

In accordance with BioBuilds release policies, Lab7 Systems is making the bioinformatics tools optimized using Intel's compiler freely available. "This reflects Lab7 Systems' commitment to supporting open source in bioinformatics and provides users access to pre-built, higher-performance binaries they would normally not have access to," explained Lee.

In addition, Lab7 Systems provides paid support contracts for users who need additional optimizations for specific Intel64 CPUs, or who would like optimized packages that cannot be included in the public (free) release due to issues such as licensing restrictions.

Building the Future

Lab7 Systems will continue to use the Intel® compilers to build performance-tuned packages for the 2017.11 release. In addition, Lab7 Systems plans to use other Intel Parallel Studio XE components to further improve the performance of certain applications:

- **Intel® VTune™ Amplifier:** An advanced performance profiler for tuning application performance, scalability, and memory access. Enables accurate profiling of C, C++, Fortran, Python*, Go*, assembly, and Java* or a mix of languages.
- **Intel® Advisor:** A vectorization optimization and threading advisor tool for C, C++, C#, and Fortran applications to optimize vectorization and quickly prototype threading designs.

Learn More

Intel® Parallel Studio XE >

Intel® C++ Compiler >

BioBuilds >