Intersect360 RESEARCH

Actionable Market Intelligence for High Performance Computing

ROGUE WAVE TOOLS AND LIBRARIES FOR ACADEMIA AND RESEARCH

Michael Feldman

White paper

December 2014

MARKET DYNAMICS

Scientific modeling and simulations have been the foundation of high performance computing (HPC) since the first supercomputers were built in the 1960s. Today supercomputers are used for both technical and business computing, but the legacy science applications are still the considered the mainstay of the industry 50 years later.

Today technical computing drives research in every scientific domain, including physics, chemistry, life science, mathematics, and earth science. Although much of this supports basic research at government labs and universities, the work is the foundation of practical applications that fill critical needs in both the public and private sectors. Climate and weather modeling, genomic analysis, energy system development, biological simulations, and nuclear weapons support, are just a few of the applications areas that rely on technical computing to fulfill their needs. Taken together, these research areas are the foundation for trillions of dollars in the global economy.

The principal users for these applications are researchers working in government labs and universities. Together they drive a large part of the HPC market, with a total annual spend of approximately $4 billion (2010) for servers alone[1]. We estimate another $2 billion is being spent on software, which includes application codes and system software, as well as software development tools and libraries.

.

PRODUCTIVITY DRIVERS: SOFTWARE TOOLS AND LIBRARIES

> **"The value of software development tools and libraries is tied to how well they contribute to the productivity of application developers."**

The latter two categories of tools and libraries represent a fairly small portion

[1] High Performance Computing Forecast for 2011 through 2015: Economic Sectors and Vertical Markets, Intersect360 Research, 2010.

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124 | www.Intersect360.com | info@Intersect360.com

of total HPC spending, but they are critical to application development. Compilers, debuggers, code profilers, and analysis tools are the main tools software developers employ to create applications, that is, transform source code into working executable code. Libraries, on the other hand, are prepackaged general-purpose functions that can be built into these same applications. Since they relieve the developer from coding these functions themselves, using them can speed development as well as improve code quality and maintainability. Together, tools and libraries provide the basic componentry of software development.

Development of HPC application software is one of the most time-consuming, and therefore costly activities at government and universities facilities where this work is taking place. The need to constantly upgrade legacy software at these sites, as well as developing new applications, takes significant resources. Moreover, since this work is dependent on highly trained professionals, it tends to be expensive and stay expensive.

In general, the value of software development tools and libraries is tied to how well they contribute to the productivity of application developers. Unlike hardware, which gets faster and less expensive with each passing year, application development does not. It requires human effort and interaction, which is resistant to automation. To the extent it can be automated though, the software development process can be accelerated and otherwise improved, and this becomes more critical as software costs outpace hardware costs.

There is another aspect of productivity that comes into play – that of run-time productivity. The end user doesn't just want an application that executes error-free, but one that has other desirable properties. Often that means just fast performance, but users may also desire applications that use minimal resources (for example, memory capacity, I/O, network bandwidth, and power usage) or can be ported easily to other architectures. In the latter two cases, performance may be sacrificed somewhat.

In research and academia, there are also critical application requirements for both accuracy and reproducibility. Scientific simulations rely on high levels of mathematical precision and determinism to model physical systems authentically. From the standpoint of accuracy, this often presents a further challenge to performance, since such precision requires more computational intensity. To support these capabilities, the entire development tool chain – from compiler to debugger – needs to be designed with these issues in mind.

Also, given that government labs and universities use some of the most advanced HPC systems in the world, developers who build technical computing applications for computers have particular demands when it comes to tools and libraries. Often paramount among these is the need for highly scalable applications, which necessitates support for extensive parallelism across the entire development tool chain. For example, debuggers and analysis

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

tools must be able to manage thousands of threads, while libraries must support parallel execution with the ability to scale up and down automatically, based on the underlying hardware.

HPC systems in government and academia also tend to incorporate some of the newer technologies, like GPUs or other accelerators. In the case of supercomputers, there are often a non-standard architectures and technologies to contend with, such as a custom-built networks and proprietary software interfaces. Generic development tools and libraries are less likely to support such exotic technologies.

OPPORTUNITY FOR ROGUE WAVE

Rogue Wave Software offers a unique set of tools and libraries for software engineers who develop compute and data-intensive applications. Foremost among these is the TotalView debugger and the IMSL Numerical Libraries. Additionally, Rogue Wave also offers a static code analysis tool, Klocwork, and an open source management tool, OpenLogic, which are more general-purpose utilities that support large-scale application development.

TotalView Debugger

For HPC application development, the TotalView debugger is one of the most popular tools cited by end users. In our 2013 site census survey[2], TotalView was the top mentioned software tool, cited by 41 percent of respondents. TotalView is a full-featured debugger, and one that is specifically built for high performance computing applications running on a variety of HPC platforms.

In particular, TotalView provides a robust debugging capability for multithreaded applications, giving developers complete control over execution on a thread-by-thread basis. It supports all of the most popular parallel computing frameworks used in HPC, including MPI, OpenMP, OpenACC, and CUDA, as well as the most common programming languages used in technical computing: Fortran, C, and C++.

It also supports the latest accelerator hardware from Intel and NVIDIA, namely the Xeon Phi coprocessor and the Tesla GPUs, in addition to standard multicore x86 processors. The debugger can also deal with Cray and the IBM Blue Gene platforms, the two most popular capability-class supercomputers used at national labs and universities.

Another useful feature is a reverse debugging capability, known as ReplayEngine, which allows the developer to trace execution backward in time in a deterministic manner. This is

[2] HPC User Site Census: Middleware, Intersect360 Research, April 2014

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

especially useful after a crash, where the source of the problem is often many instructions back in the execution stream.

Memory debugging is also available with a tool known as MemoryScape. It allows the developer to find heap allocation problems, such as memory leaks, dangling pointers, memory bounds violations, and read-before-write operations. The multithreaded nature of most HPC applications makes it especially prone to memory errors, so a tool like MemoryScape can be an invaluable adjunct to traditional source line debugging.

IMSL Numerical Libraries

Numerical libraries are a mainstay of technical computing, especially for scientific applications. Rogue Wave's IMSL libraries offer a broad set of mathematical, statistical, and analytics functions that can be applied to a wide range of disciplines and are built for performance demanding users. The libraries offer basic functionality, such as matrix operations, linear algebra, and random number generation in addition to more complex routines encapsulating genetic algorithms, neural networks, and logistic regression. The libraries are available in standard C, C#, Fortran, and Java.

The use of IMSL functions by the application developer can save weeks or months of development time, eliminating the associated effort of algorithm design, implementation, testing, and documentation. Such use also ensures better application quality since the IMSL functions are well-tested (a 40-year track record), provide robust error handling, and are designed for the kind of scalability inherent in high performance computing environments.

Klocwork

Code analysis tools are used to identify coding defects, reliability issues, and security flaws at the earliest possible point in the development process, namely when the source lines are being written. As a result, such tools can reduce application development time, while assuring high quality, robust code.

Rogue Wave's offering in this area, Klocwork, is designed with such functionality in mind and provides a complete feature set, including static code analysis, code refactoring, detection of security vulnerabilities, code design optimization, and code review support. All of this is provided interactively to the developer, during desktop code development. As a result, many software issues and flaws can be circumvented before testing and integration, saving time and frustration downstream.

Although scientific developers tend not to use code analysis, for large applications built by teams of developers (e.g., climate/weather models, brain simulations, nuclear weapon

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

simulation codes, etc.) such a tool could be invaluable. Since many HPC applications have to be periodically redesigned to incorporate additional levels of scalability (as core and node counts rise) or ported to new architectures, the utility of a code analysis tool could shorten the time needed for recoding and re-testing, easing these inevitable software transitions.

OpenLogic

Much of the application and system software for HPC sites in government and academia is based on open source software. Linux (OS), Open MPI (message passing), and the Rocks Cluster Distribution, GROMACS (molecular dynamics), and OpenFOAM (CFD) are just a few of the examples of open source packages used by HPC developers. Source management of these packages is often haphazard though, costing resources and time.

Although primarily used by enterprise customers, Rogue Wave's OpenLogic is specifically designed to support coherent management of an organization's use of open source packages. It is able to scan source with regard to policy, licensing requirements, security, and compliance standards. Further it provides a certified library of hundreds of open source packages via a free site where users can download the latest releases and patches of these packages.

INTERSECT360 RESEARCH ANALYSIS

Software tools like the TotalView debugger can increase developer productivity by providing a high-level interface for finding application bugs, especially for multithreaded applications on HPC platforms. A variety of debug tools and techniques are used by developers to uncover programming errors, but a well-designed, full-featured tool like TotalView can speed the process considerably. Likewise, the IMSL Numerical Libraries can accelerate HPC application development, in this case, by replacing developer effort with prepackaged software that is well-tested and well-maintained.

Static code analysis and open source management, as provided by Klocwork and OpenLogic respectively, can further speed the development process. Klocwork, in particular, can help find bugs and design errors early in the development cycle, avoiding the more expensive prospect of uncovering these flaws during debugging, testing, or even deployment. OpenLogic offers a way to optimize open source management, adding structure and predictability to the application integration process, avoiding costs related to incompatible or invalid source builds.

While this software does incur extra expense, high quality tools and libraries, such as those offered by Rogue Wave, can cut costs by reducing application development and

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com

maintenance efforts, as well improving the quality of the resulting software, both its performance and its robustness.

The resistance to using such tools and libraries is often rooted in an unwillingness to change established workflows and methods, as well as ignorance of productivity gains that can be realized. Using generic open source tools for debugging (or even employing printf tracing), conducting manual code reviews for static source code analysis, and developing in-house custom mathematical libraries for general-purpose algorithms can alleviate the need for purpose-built tools and libraries, but at the expense of developer productivity and software quality. Any lost productivity from these "good-enough" approaches is hidden within the efforts of developers, who represent a fixed cost to the organization. Realizing the benefits of purpose-built tools and libraries for HPC application development requires a willingness to question development orthodoxy and an analysis of costs and benefits.

Intersect360 Research
P.O. Box 60296 | Sunnyvale, CA 94088 | Tel. [888] 256-0124
www.Intersect360.com | info@Intersect360.com