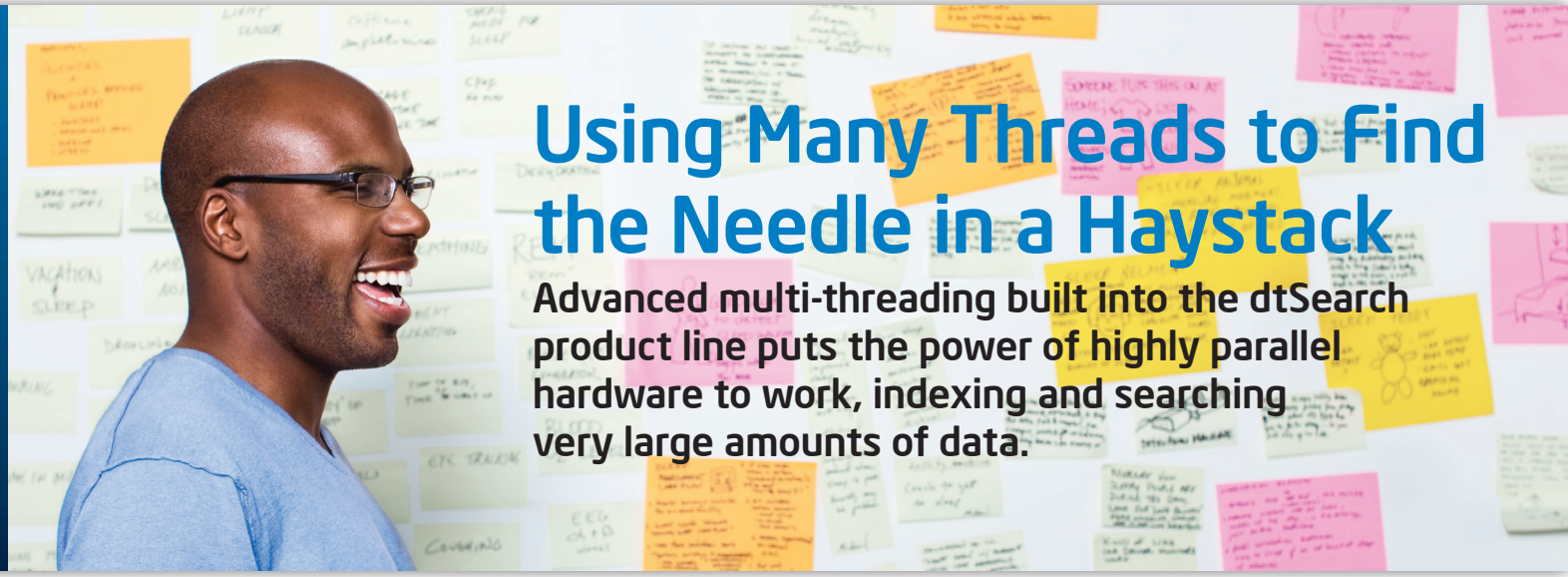
Success Story

Intel® Software Partner Program

dtSearch Corp.

"Performance is essential to our product line. We review performance parameters at every step of data access, data parsing, indexing, searching, and hit-highlighting. Intel® VTune™ Performance Analyzer excels at helping us optimize these processes as part of our development cycle."

- David Thede,
President, dtSearch Corp.



Using Many Threads to Find the Needle in a Haystack

Advanced multi-threading built into the dtSearch product line puts the power of highly parallel hardware to work, indexing and searching very large amounts of data.

CHALLENGE

Deliver flexible search functionality that can be implemented across the broadest possible range of applications and environments. Support advanced functionality, such as fuzzy logic and multiple data classification objects, while maintaining lightning speed over large data domains.

SOLUTION

dtSearch Corp., a member of the Intel® Software Partner Program, has developed a sophisticated line of search products with proprietary file parsing and file conversion capabilities. To support an expanding set of advanced features, dtSearch uses multi-threading to take advantage of the performance headroom provided by multi-core client and server hardware. Using Intel® VTune™ Performance Analyzer and Intel® Concurrency Checker has been central to this effort.

CUSTOMER BENEFIT

Optimizing dtSearch's algorithms enables very high levels of concurrency, as measured using Intel® Concurrency Checker. Enhancing software parallelism using tools and best practices from Intel helps the company focus more intently on its core mission. As a result, dtSearch increased performance levels across many solution domains, improving product quality and achieving a competitive advantage. **More** ›

Learn more: **www.intel.com/partner**

PROOF POINT

Fine-Tuning Search Performance for Intel® Architecture

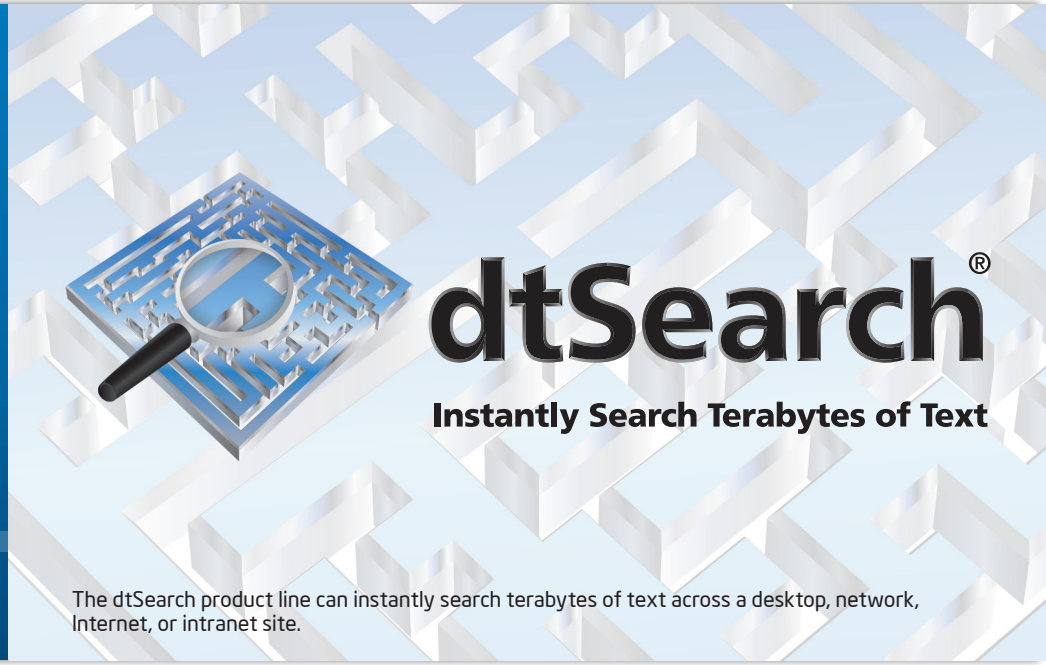
Early in the development of its search functionality, dtSearch identified an opportunity to fine-tune its performance on Intel® platforms.

Today, the company's products and APIs specifically target 32-bit and 64-bit multi-core processors from Intel.

For example, one best practice the company places high on its list of strategies is using Intel® VTune™ Performance Analyzer to locate performance bottlenecks and opportunities for optimization.

Using the Intel® Currency Checker, available through the Intel® Software Partner Program, the company then tested a dtSearch Engine sample application to simulate high-volume concurrent searching of a single shared index, similar to what might occur on a high-traffic web site.

The dtSearch Engine multi-threaded indexed search demo achieved 100 percent parallel time in the Intel Currency Checker test, indicating full optimization for multi-core hardware under that test scenario.



The dtSearch product line can instantly search terabytes of text across a desktop, network, Internet, or intranet site.

Harnessing Multi-core for Advanced Search Functions

Headquartered in Bethesda, Maryland, dtSearch Corp. creates software for finding and retrieving text-based search elements (with highlighted hits) within large bodies of data. To satisfy the company tagline, "Instantly search terabytes of text," dtSearch has constantly refined its algorithms and features since being founded in 1991. The technology is distributed in a portfolio of enterprise-ready applications for both Microsoft Windows* and Linux*:

- **dtSearch Desktop with Spider.** Instantly search desktop content, including popular file types and email formats, along with nested and ZIP attachments; can also spider selected web sites.
- **dtSearch Network with Spider.** Instantly search the many forms of data that exist across a large enterprise network; Spider adds local and remote Internet and Intranet data (including dynamically generated content such as MS SharePoint or CMS data) to a searchable database.
- **dtSearch Publish.** Quickly publish instantly searchable document collections or web site content to portable media, including CDs, DVDs, and external hard drives.
- **dtSearch Web with Spider.** Quickly publish a wide variety of file types to a web site; Spider adds local or remote web data (including dynamically generated content) to a site's searchable database.
- **dtSearch Engine for Linux.** Add dtSearch search features and built-in file format support to your Linux application; for Java* and C++ programmers.
- **dtSearch Engine for Windows and .NET.** Add dtSearch search features and built-in format support to your application; API supports .NET, C++, and Java, including an API for indexing SQL databases along with BLOB data, and a Spider API.

The robustness of these products depends on their ability to take excellent advantage of the processing power of multi-core platforms. Because software parallelism is the key means of delivering that computational muscle, a sophisticated multi-threading model is vital to dtSearch's success. The company has successfully harnessed multi-core processing to provide fast search results, bolstered by innovation that's delivered in the form of advanced software capabilities.

One of the thorniest search problems occurs with hierarchical sorting in cases involving millions of document metadata tags or database records. For example, a user may want to limit retrieved records from a manufacturing database to just parts involved in automobile manufacturing. The user may then want to limit retrieved records to parts used only by a particular automobile manufacturer, and then further refine the search to retrieve records pertaining to a certain type of car or a specific car model.

The relationship between Intel and dtSearch stretches back a number of years, helping dtSearch continue to develop in parallel with the evolution of client and server platforms. As a result, the combination of hardware and software generates synergies that deliver excellent performance and other benefits to end-customers, including internal customers at Intel.

Possible solutions to the challenge of searching for words related to the search string could include automatically incorporating data from an electronic thesaurus or other resources for example. It is also valuable for the user to be able to specify how literally the search results should match the search string he or she enters.

These types of features are necessary to provide the most robust search functionality possible, and because adding operations to the basic search function is processor-intensive, optimizing for multi-core platforms is a necessity.

Targeting Intel® Architecture with Optimized Performance

In its ongoing quest to deliver the highest possible performance on the latest Intel® platforms, dtSearch is always searching for the best tools and techniques for the job. One best practice the company places high on its list of strategies is using Intel® VTune™ Performance Analyzer to locate performance bottlenecks and opportunities for optimization.

"Performance is essential to our product line," explained David Thede, president of dtSearch Corp. "We review performance parameters at every step of data access, data parsing, indexing, searching, and hit-highlighting. Intel VTune Performance Analyzer excels at helping us optimize these processes as part of our development cycle."

Of particular benefit to dtSearch is the integrated support for Intel® Thread Profiler and other benchmarking tools in the VTune environment. dtSearch used this combination of tools in the development of the company's core component, the dtSearch engine, which is incorporated into the full product line. As a comprehensive offering for environments of many shapes and sizes, the dtSearch engine API supports .NET, Java, and C++, including native, thread-safe support for 32-bit and 64-bit Windows and Linux.

Because the dtSearch engine directly targets such a broad range of environments, it must be updated often to ensure full compatibility and high utilization of the resources available from new environments such as updated processors, operating systems, or runtime environments. The VTune analyzer is designed explicitly to be used iteratively throughout the application development life cycle, so it is well suited to ongoing improvements such as those required to keep pace with those updates.

To gauge the success of its optimization efforts, dtSearch reached out to the Intel Software Partner Program to use the Intel® Concurrency Checker, an Intel® Software Assessment Tool developed to ascertain the level of concurrency built into any given piece of software. The results show how well the software uses the parallel resources available from multi-core hardware and allows software makers to anonymously compare their results to those of their peers.

From the results shown with Intel Concurrency Checker, dtSearch was able to conclude that its optimization efforts had indeed borne fruit, taking excellent advantage of multi-core Intel platforms:

- **Simple test scenario.** Designed to be easy to use, Intel Concurrency Checker targets testing and retesting after initial setup to be completed within one hour. This feature encourages periodic retesting as software is refined and updated, an opportunity that dtSearch intends to continue to take advantage of into the future.
- **Realistic runtime implementation.** The tested dtSearch Engine sample application simulated high-volume concurrent searching of a single shared index, similar to what might occur on a high-traffic web site.
- **Excellent results.** Following VTune Performance Analyzer optimization, the dtSearch Engine multi-threaded indexed search demo achieved 100 percent parallel time in the test, indicating full optimization for multi-core hardware under that specific test scenario.[1]

Because the tool is extremely simple to use, the company can also run it against future versions of the software. This ongoing capability will help ensure that dtSearch reaches the goal of advanced parallelism on the latest hardware, leaving nothing to chance in its quest to provide customers with the best of enterprise and online search.

Building Success with the Intel® Software Partner Program

Participation in the Intel Software Partner Program has been an excellent resource for dtSearch as it continues to refine its product line for performance and advanced functionality. The combination of tools and expertise available from Intel gives dtSearch insight into taking advantage of the full range of capabilities of multi-core platforms, and the company has risen to the challenge of advanced multi-threading.

The relationship between Intel and dtSearch stretches back a number of years, helping dtSearch continue to develop in parallel with the evolution of client and server platforms. As a result, the combination of hardware and software generates synergies that deliver excellent performance and other benefits to end-customers, including internal customers at Intel.

Looking forward, dtSearch is building its plans for an increasingly multi-core computing world, and that future promises to hold a universe of new capabilities and opportunities. Ongoing participation in the Intel Software Partner Program is a driver of success for the company, as it seeks to keep pace in its search technology with ever-expanding enterprise data stores.

About the Intel® Software Partner Program

The Intel® Software Partner Program provides a framework for collaborative solution development around Intel® architecture. From business planning and product development to marketing and sales, the program helps to drive increased business success and market opportunities.
Learn more at **www.intel.com/partner.**

Success Story by:

Learn more about dtSearch:
www.dtsearch.com

Visit the Intel® Software Network Parallel Programming Community:
software.intel.com/en-us/parallel

[1] Results reported by the Intel® Software Partner Program Software Assessment Report generated and submitted by dtSearch Corp.

Intel and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
*Other names and brands may be claimed as the property of others.
Copyright © 2010 Intel Corporation. All rights reserved.
1110/BM/MESH/PDF 324578-001US