SOLUTION BRIEF

Data Analytics & AI
Intel® Optimized Software

Bringing New Speed and Intelligence to Enterprise Decision Making

2.37X[1] faster performance for Mphasis DeepInsights* with Intel® Xeon® Scalable processors

Mphasis
The Next Applied

Data analytics is progressing faster than most businesses can evaluate and integrate new technologies. Even many large enterprises with extensive IT resources can't keep pace with breakthroughs that are emerging in the areas of data ingestion, data management, analysis and, perhaps most importantly, artificial intelligence (AI).

Deep learning, machine learning, and other AI technologies offer powerful new ways to cull real-time and contextually relevant insights from the massive and chaotic flow of enterprise transactions and communications. AI algorithms fundamentally shift the paradigm of data usage. By learning over time, they enable progressive improvements in decision making, without the need for complex, hands-on programming.

Of course, AI tools and methodologies are evolving quickly, so applying them to specific business problems can be challenging. Moreover, any analytics solution can only be as good as the data it operates on, so a comprehensive enterprise data strategy is necessary to provide the foundation for success. Mphasis helps large businesses overcome these challenges.

Mphasis: DeepInsights at Enterprise Scale

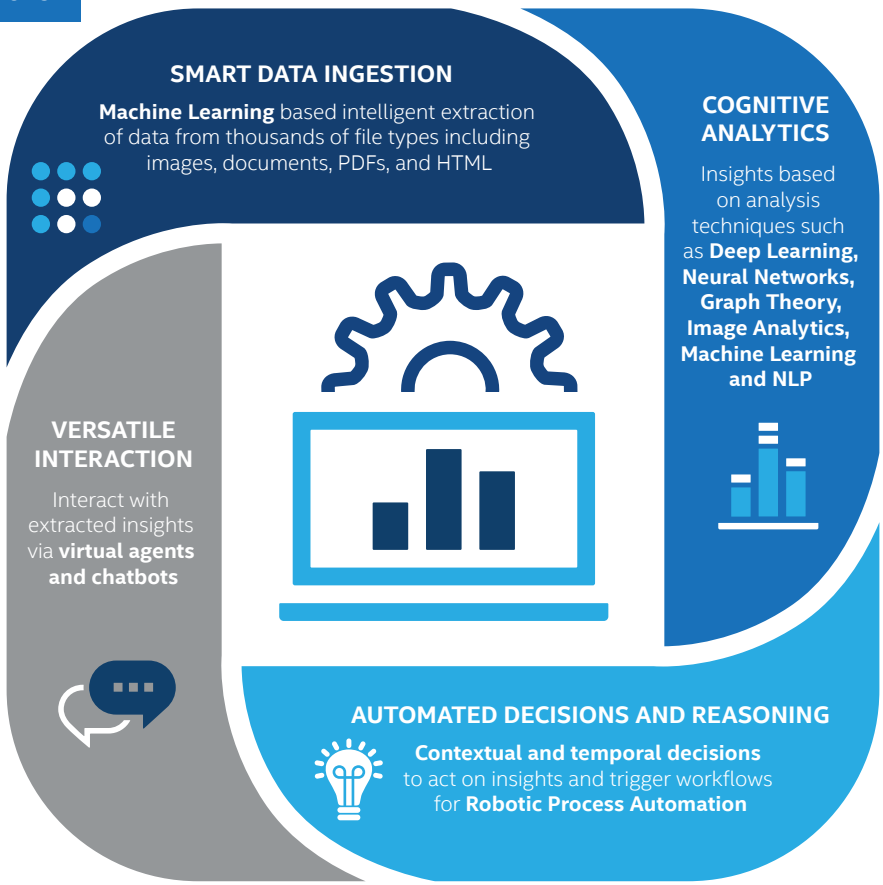
Mphasis, a member of the Intel® AI Builders program, offers solutions and services that help enterprise customers integrate leading-edge data solutions effectively, so they can gain first-mover advantages without excessive cost and risk. The global IT service and solution provider takes an 80/20 approach to solution delivery, providing solutions that are 80 percent compatible with the needs of virtually any large organization, then adding additional modules and customization to target specific business requirements.

Although their solutions and methods are adaptable across almost any industry, Mphasis offers a particularly high level of expertise in the fields of banking, finance, insurance, manufacturing, and logistics. Mphasis is currently working with six global banks, eleven out of the fifteen top mortgage lenders, and three global insurance companies to help those organizations make better and faster decisions based on real-time analysis of enterprise data.

At the heart of the Mphasis approach to enterprise analytics is DeepInsights*, a cognitive intelligence platform that combines state-of-the-art algorithms in machine learning, neural networks, deep learning, semantics, image analytics, graph theory, predictive analysis and natural language processing. With this combination of capabilities in an integrated platform, organizations can process and analyze data of almost any type, both structured and unstructured, without the cost and delays of human-mediated data ingestion. Thousands of file types are supported, including email, call logs, documents, service tickets, website traffic, sensors, and databases. External data sources can also be integrated through APIs, which can help to reinforce findings and weed out false positives.

INTEL®
AI BUILDERS
MEMBER

DeepInsights* - Functions



How it Works

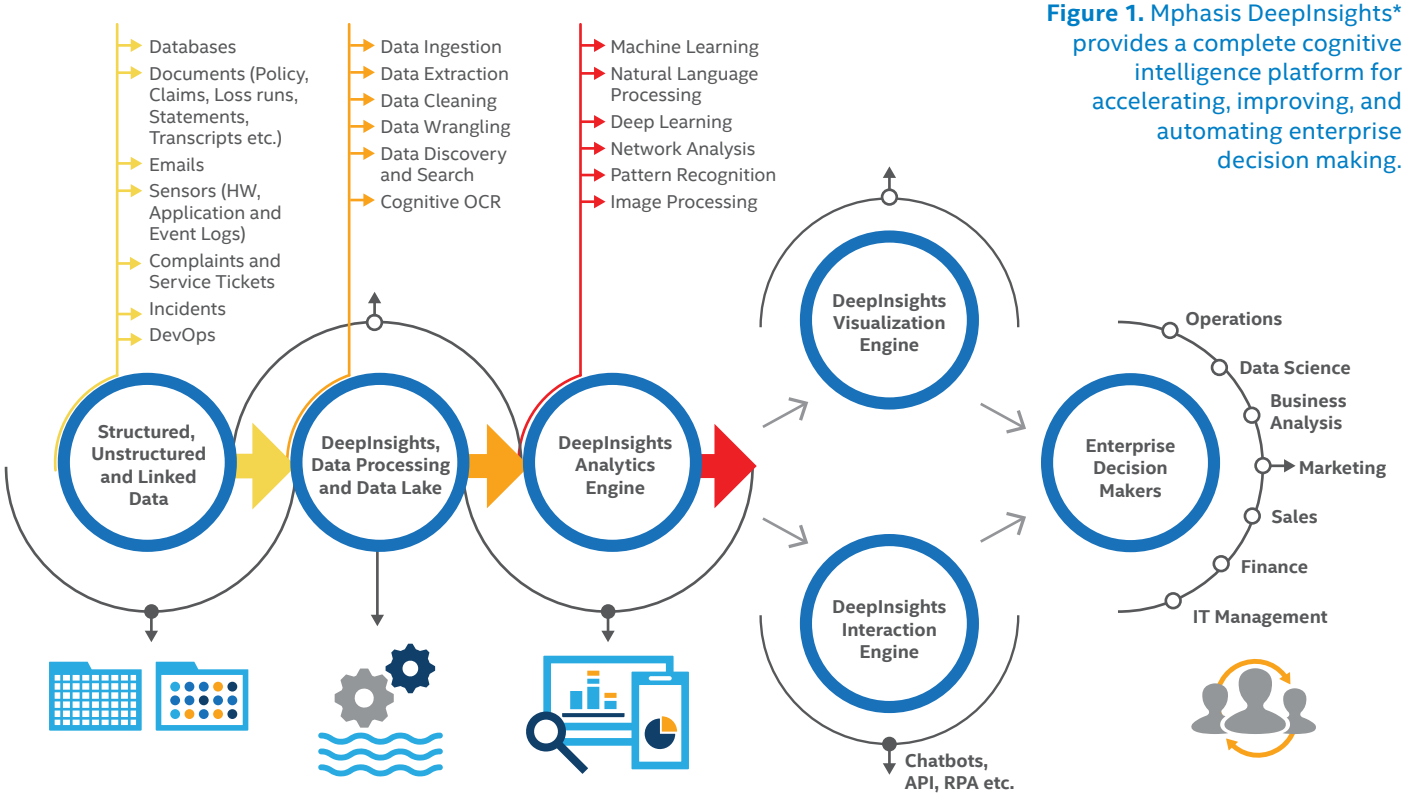


Figure 1. Mphasis DeepInsights* provides a complete cognitive intelligence platform for accelerating, improving, and automating enterprise decision making.

Practical Value for Critical Processes

DeepInsights combines the best of machine learning and deep learning to automatically analyze data within the appropriate business context. It then integrates the results back into business workflows. According to Mphasis, DeepInsights is applicable to any domain in which employees manually review documents and extract data for downstream processing. Cognitive analysis can help to bring greater consistency and reliability to decision making.

DeepInsights extracts actionable information from many sources.

- **Social Networking**. Public and private social networking data can be analyzed in real time for insights into customers and market trends.
- **Email**. Incoming, outgoing, and inter-employee email offers a potential goldmine of insights into communications, operations, and customer experience.
- **Claims and Service Tickets**. Automated sorting and routing, plus contextual insights, can help drive better, faster decisions.
- **Fraud and Abuse**. Sophisticated pattern matching helps to identify real-time threats and reduce false positives.

Hyper-Personalization and Progressive Automation

In these cases and many others, DeepInsights can channel information and insights to human experts, chatbots, or other applications for automated or human-remediated action. The enterprise can also move toward new levels of personalization and consistency across the many channels of customer engagement. Whether customizing an online offer or responding to a phone inquiry or an email complaint, cognitive analysis provides a data-driven, self-learning approach at the scale required for enterprise deployments. It empowers organizations to begin treating more of their customers as the unique individuals that they are—leading to better short-term results and greater, long-term loyalty.

Speed and Flexibility with Intel® Optimized Software Packages

With their 80/20 approach, Mphasis helps to reduce implementation cost and risk and accelerate time to value. Mphasis's cloud-based solutions further reduce upfront costs and make it easier for organizations to expand their solutions over time.

Performance also matters. Whether onsite or in the cloud, time-to-insight is a key factor for real-time cognitive solutions, so fast processing of diverse and fast-moving data is important. Costs are also dependent on required infrastructure, so maximizing performance per server and performance per watt are important both to Mphasis and its customers.

Highly Optimized Distributions of Python* and TensorFlow*

Mphasis and Intel have worked together to optimize DeepInsights for Intel architecture and to make sure that infrastructure solutions are as simple and interoperable as possible. The DeepInsights cognitive platform runs in a Microsoft Windows environment and relies on Intel® Distribution for Python and Intel® Optimization for TensorFlow. Python has become one of the most popular languages for machine learning and deep learning applications, and TensorFlow is one of the most popular deep learning frameworks. Together, these open source applications provide a flexible platform that is advancing quickly to handle a growing range of requirements.

Both Intel Distribution for Python and Intel optimized TensorFlow have been highly optimized by Intel engineers using the Intel® Math Kernel Library (Intel® MKL). Intel MKL provides optimized algorithms for the basic mathematical operations that underpin many number-crunching scientific, technical, analytic, and financial applications. It includes algorithms for many of the most popular and compute-intensive operations used in neural network training and inference solutions.

The software optimizations in Intel MKL help to ensure that each algorithm makes efficient use of the available execution resources in the latest Intel® Xeon® Scalable processors, including multiple cores and threads (up to 28 cores and 56 threads per processor). Memory and cache usage are also optimized, which helps to ensure that required data gets to the cores quickly and efficiently to avoid lags in processing.

The algorithms in Intel MKL are optimized to take advantage of Intel® Advanced Vector Extensions (Intel® AVX), which allow a single instruction to be executed simultaneously across multiple data points. Intel has enhanced this technology in successive processor generations to provide increasing levels of parallelism. The latest Intel Xeon Scalable processors support Intel® AVX-512, which enables simultaneous processing of all the data elements stored in 512-bit vector registers. Optimizing software for this strategy is known as vectorization and can dramatically increase performance for operations that can be parallelized in this way.

Better Together: Up to 2.37X[1] Faster Performance with Intel® Xeon® Scalable Processors

A key advantage of using Intel software tools and libraries is that the optimized code tends to perform better not only on current server platforms, but also on future platforms. That's because Intel integrates new capabilities and execution resources into each new processor generation and optimizes Intel MKL and other tools and applications to utilize these advances. This typically translates into significant performance gains.

To verify the performance improvements of Mphasis DeepInsights across successive Intel® Xeon® processor generations, Mphasis and Intel benchmarked the optimized solution on two servers. The system under test was a two-socket server configured with the latest Intel® Xeon® Platinum 8180 processor; the baseline server was a one-socket public cloud instance configured with the previous generation Intel® Xeon® processor E5-2686 v4. Both servers were running Microsoft Windows Server*.

Intel engineers from the Intel AI Builders program set up and configured the server and provided access to Mphasis engineers so they could run the benchmark. The tests compared performance for DeepInsights as it processed data on a pre-trained neural network based on a custom Natural Language Processing (NLP) topology.

The benchmark results showed that the server configured with the Intel Xeon Scalable processors delivered 2.37X[1] the performance of the server configured with the Intel Xeon processor E5-2686 v4. Mphasis can use these gains to help its customers analyze larger data volumes and to deliver more and better insights using a smaller hardware footprint.

Inference Performance with Intel® Xeon® Scalable processors

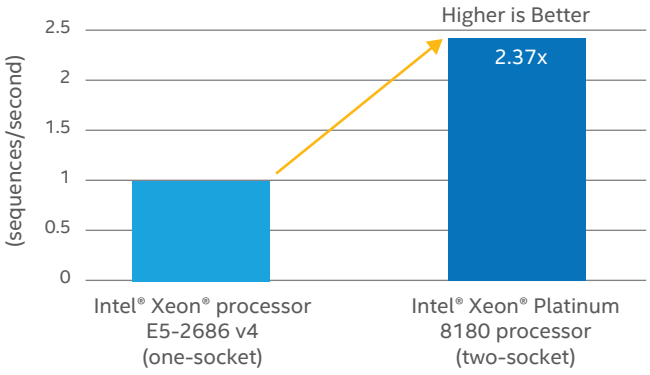


Figure 2. Running Mphasis DeepInsights* on a two-socket server based on the latest Intel® Xeon® Platinum 8180 processor delivered 2.37X1 the performance of the same application running on a single-socket server based on the previous-generation Intel® Xeon® processor E5-2686 v4.

Better Solutions for the Broader AI Community

Intel works with open source and commercial software developers to optimize a wide range of popular platforms and applications for each new Intel processor generation. In the case of open source software, such as Python and TensorFlow, Intel typically up-streams the Intel® optimized code into the main distribution of the application. With this approach, users can obtain highly optimized performance out of the box on Intel Xeon Scalable processors, as well as many other Intel® processors.

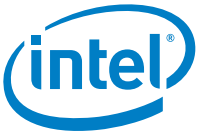
This software optimization work is helping to fuel the AI revolution by allowing companies such as Mphasis to run demanding machine learning and deep learning workloads on industry-standard servers, without the need for costly, specialized hardware platforms. Implementation is simpler, and AI applications can be integrated more easily with the full range of enterprise workloads.

Conclusion

Mphasis and other members of Intel AI Builders are integrating leading-edge AI technologies into enterprise-ready solutions that are already beginning to transform select industries. We believe these solutions mark the beginning of a broad move toward mainstream AI adoption that will drive ever-deeper improvements in efficiency, automation, responsiveness and personalization. As this transformation unfolds, Intel hardware and software technologies will continue to play a foundational role, helping vendors and customers speed innovation, while reducing cost and complexity.

To learn more, visit Intel AI Builders at:
https://builders.intel.com/ai

To learn more about Mphasis, visit:
https://www.mphasis.com/



[1] Tests performed by Intel, October 2018.
Workload: Natural Language Processing (NLP) using a custom topology.
Baseline system configuration: public cloud instance (one-socket server configured with 1 x Intel® Xeon® processor E5-2686 v4 (2.30 GHz, 8 cores), 31GB memory, 62 GB Intel® Solid State Drive (Intel® SSD), Microsoft Windows Server.
New system configuration: System Configuration: Two-socket server configured with 2 x Intel® Xeon® Platinum processor 8180 (2.50 GHz, 28 cores), 192 GB DDR4@2666MHz memory (12 x 16 GB DIMMS), 1.5 TB Intel® SSD (SC2BX01), Windows Server 2016 Std. Software application for both servers: TensorFlow-Serving r1.9 (https://github.com/tensorflow/serving); Intel Optimized software application: TensorFlow-Serving r1.9 + Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN) (https://mirror.bazel.build/github.com/intel/mkl-dnn/archive/0c1cf54b63732e5a723c5670f66f6dfb19b64d20.tar.gz) + optimizations (availability of optimizations expected in TensorFlow-Serving release 1.10).

© 2019 Intel Corporation. Intel, Xeon, AVX, the Intel logo, and Xeon logos are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

Intel processors of the same SKU may vary in frequency or power as a result of natural variability in the production process.

For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Performance results are based on testing and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/performance/datacenter.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804.

Intel, Xeon, AVX, the Intel logo, and Xeon logos are trademarks of Intel Corporation in the U.S. and/or other countries.

Microsoft, Windows, and the Windows logo are trademarks, or registered trademarks of Microsoft Corporation in the United States and/or other countries.

*Other names and brands may be claimed as the property of others.

© 2019 Intel Corporation. 0119/RA/MESH/PDF 338646-001US