

Solution: Deep Insights

Mphasis



RESULTS



INTEL®
AI BUILDERS
MEMBER

2.37X INCREASE

In inference performance



Partner: Mphasis is a \$1 billion global service provider, delivering technology based solutions to clients across the world. Their product, Deep Insights, generates actionable insights quickly from physical and digital data sources such as annual reports, bank statements, emails, customer complaints, in real time.

Challenge: Manual analysis is non-real time, time consuming, costly and low on accuracy. The challenge is connecting the data to the enterprise workflow across applications and processes. A critical need exists for an industry standard server to analyze larger data volume, run demanding machine learning and deep learning workloads.

Solution: Mphasis' DeepInsights™ is a cognitive intelligence platform that enables enterprises to have faster and more effective access to insights from data. Optimizing on IA helped deliver better accuracy in detection and extraction of key fields and reduced time taken for data extraction. Achieved 2.37x speedup in inference performance using Intel distribution for Python on Intel Xeon Scalable processors, compared to previous generation.

Solution Brief: <https://builders.intel.com/docs/aibuilders/bringing-new-speed-and-intelligence-to-enterprise-decision-making.pdf>

Configuration: 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).

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. For more complete information visit <http://www.intel.com/performance>. Performance results are based on testing as of August 2018 and may not reflect all publicly available security updates. See configuration disclosure for details.

