

Redefining Cognitive Automation For Future-Ready Enterprise IT

Intel® AI collaboration with ignio, a cognitive automation software, allows enterprises to drive agile and proactive IT operations and infrastructure



INTEL®
AI BUILDERS
MEMBER



TATA
CONSULTANCY
SERVICES



ignio



"ignio is an AI-based cognitive automation software for the enterprise and more than 100 large customers have already embraced ignio as their strategic AIOps platform. We are delighted to collaborate with Intel and their cutting-edge architecture in the journey to transform enterprise IT and business operations."

Akhilesh Tripathi

Global Head, Digitate

Executive Summary

Digital transformation is a critical factor for enterprises to remain competitive in a fast-evolving economy and keep pace with skyrocketing consumer expectations. This is possible with an IT infrastructure that is modern, agile and scalable enough to support new digital and blended business models.

However, exploding data volumes and the growing scale and complexity of enterprise data centers result in unplanned IT downtime. This disrupts mission-critical operations, causes loss of data, and impairs application services, not to mention the inconvenience caused to consumers. In fact, 98% organizations surveyed in a research study stated that a single hour of downtime cost them over USD 100,000.¹

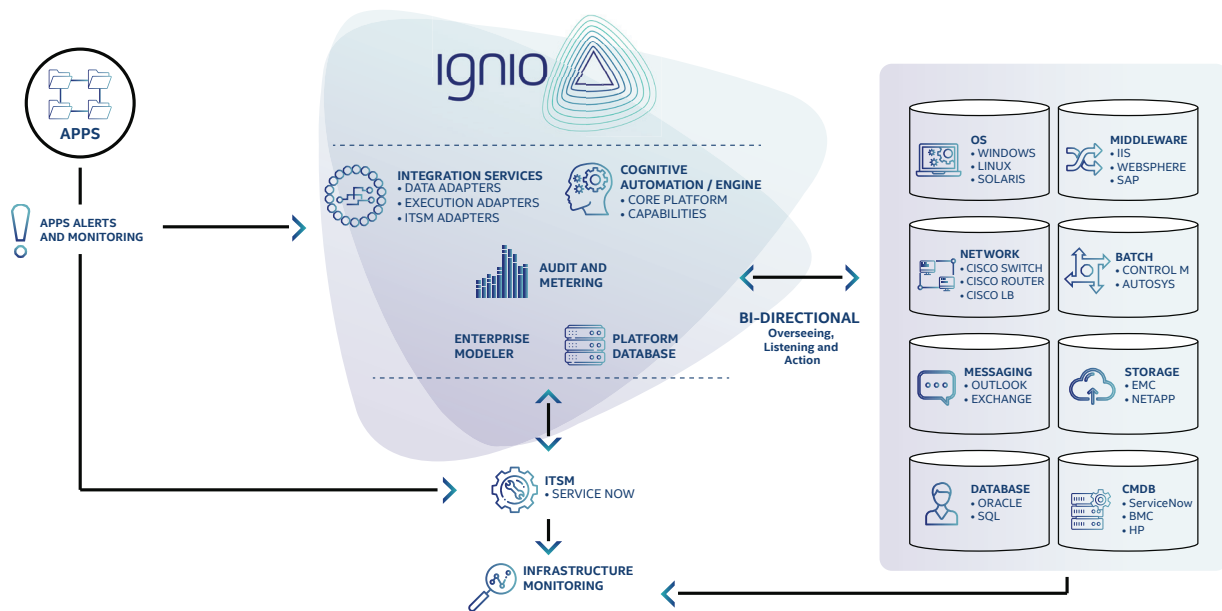
Digitate, a software venture of Tata Consultancy Services (TCS), offers ignio, a cognitive automation product that combines AI and advanced software engineering to deliver enterprise-wide benefits to IT infrastructure and application operations. Combining TCS' contextual knowledge of various industries and deep AI expertise, with Intel's state-of-the-art AI product portfolio, ignio helps organizations run enterprise IT more efficiently, improve customer experiences and increase the agility and stability of IT.

Agile, Reliable IT Secures the Foundation for Enterprise Success

ignio is a cognitive automation solution that spans across enterprise IT (AIOps), workload management, ERP operations and business operations with a common core.

- Maximizes machine learning (supervised, unsupervised and reinforcement) to build contextual blueprint of the enterprise, profile normal behavior characteristics and detect anomalies in real-time.
- Uses explainable AI to bridge the data-related gaps, inadequacies and quality issues identified in blueprint.
- Uniquely combines AI, ML and the domain knowledge to perform case, rule and model-based reasoning for prediction and fast, accurate root cause analysis.
- Provides superior experience by integrating with context-rich collaboration channels such as chat, search based conversational AI and VTT (Voice to Text).
- A one-of-a-kind software product that closes the loop between real-time predictions and actions to prevent or resolve associated conditions in IT and business operations.

ignio Architecture



Speed and Flexibility with Intel Optimized Platforms

For AI-based solutions like ignio to be implemented effectively and seamlessly, performance matters. Time-to-insight is a key factor for real-time cognitive solutions, so fast processing of diverse and fast-moving data is important.

Digitate and Intel have worked together to optimize ignio for Intel architecture and to make sure that infrastructure solutions are as simple and interoperable as possible. The ignio platform runs in a Linux environment and relies on the Intel Distribution of OpenVINO™ Toolkit and Intel Optimization for TensorFlow. Based on convolutional neural networks (CNN), the Intel Distribution of OpenVINO Toolkit extends workloads across Intel hardware and maximizes performance. It supports heterogeneous execution across Intel Computer Vision (CV) accelerators, and speeds time-to-market through an easy-to-use library of CV functions and pre-optimized kernels. The Intel Optimization for TensorFlow has been highly optimized using the Intel Math Kernel Library (Intel MKL).

The software optimizations 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 for these platforms 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 (vectorization) can dramatically increase performance for operations that can be parallelized in this way.

Better Together: Up to 275X² Faster Performance with Intel Xeon Scalable Processors

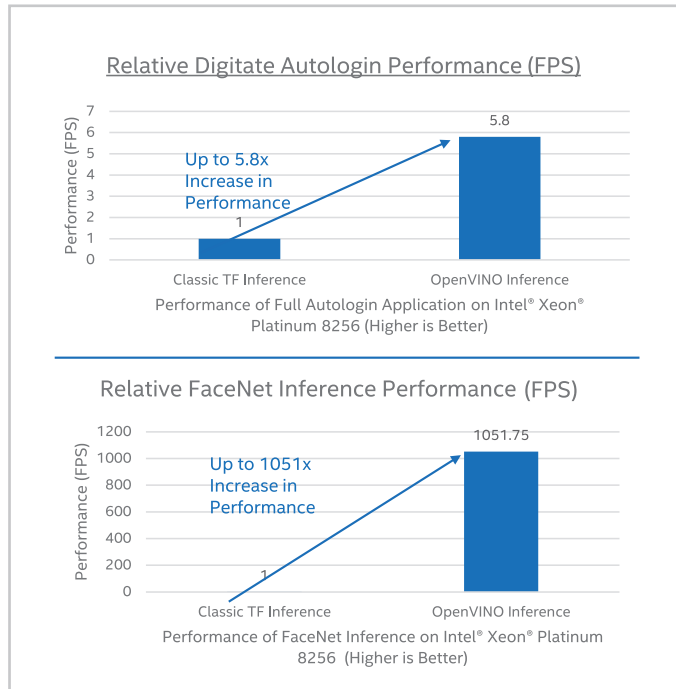
A key advantage of using Intel software tools and libraries is that the optimized code tends to perform better not only on current server platforms, but also on future platforms. That's because Intel integrates new capabilities and execution resources into each new processor generation and optimizes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

The 2nd Generation Intel Xeon Scalable processors deliver AI acceleration through new Intel Deep Learning Boost, which includes embedded instructions (Vector Neural Network Instructions, or VNNIs) that speed up dense computations characteristic of convolutional neural networks (CNNs) and other deep neural networks (DNNs). The result is more efficient inference acceleration for deep learning use cases such as image classification, speech recognition, language translation and object detection. On the software side, the performance difference between baseline open source deep learning software, and Intel-optimized software can be up to 275X² on the same Intel Xeon Scalable processor.

The impact of Intel AI products and technology optimization can be seen from one of the use cases for ignio. The objective of the solution was to reduce the time it took for a deep learning model to perform face recognition. This process involves various stages of which, one of the key stages generates a 512-dimension vector for the face detected using FaceNet, and compares the vector generated with an existing set of vectors to find the best match.

Working with the 2nd Gen Intel Xeon Scalable processors and Intel Distribution of OpenVINO Toolkit improved FaceNet inference times by up to 1051X, while delivering a 5.8X overall performance gain on ignio's auto-login workload.³

Inference Performance with Intel Xeon Scalable Processors



The benchmark results showed that the server configured with the Intel Xeon Scalable processors delivered better performance for ignio – enabling faster and better insights using a smaller hardware footprint.

Delivering Maximum Impact for Enterprises

Building on the performance advantages provided by Intel optimized platforms, ignio delivers the capabilities and features needed by enterprises to drive agile and proactive

IT operations and infrastructure. The advantages that this solution delivers can enable enterprises to:

- Vastly simplify IT operations
- Drive cost and time saving
- Increase IT infrastructure utilization
- Reduce data flow overhead

Reimagining the Future of AI

Intel works with open source and commercial developers to optimize a wide range of popular platforms and applications for each new Intel processor generation. This optimization work is helping to fuel the AI revolution by allowing companies such as Digitate 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.



"Intel and Digitate, a software unit of TCS, have been working closely to accelerate business transformation and reshape the enterprise intelligence landscape. Intel's AI product portfolio provides a strong foundation to augment the power of ignio AIOps. Our collaboration will help enterprises realize maximum value out of their IT deployment."



Prakash Mallya

Vice President & Managing Director,
Sales & Marketing Group, Intel India



Digitate is a member of the Intel **AI Builders** Program, an ecosystem of industry-leading independent software vendors (ISVs), system integrators (SIs), original equipment manufacturers (OEMs), and enterprise end users, which have a shared mission to accelerate the adoption of artificial intelligence across Intel platforms.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No product or component can be absolutely secure.

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

¹Source: <https://itc-corp.com/blog/2016/08/cost-of-hourly-downtime-soars-81-of-enterprises-say-it-exceeds-300k-on-average/>

²275X Inference throughput performance improvement with Intel® Optimization for Caffe* compared to BVLC-Caffe*. Intel measured on 12/11/2018. 2S Intel® Xeon® Platinum 8180 CPU @ 2.50GHz (28 cores), HT ON, turbo ON, 192GB total memory (12 slots * 16 GB, Micron 2666MHz), Intel® SSD SSD5C2KF5, Ubuntu 16.04 Kernel 4.15.0-42-generic; BIOS: SE5C620.86B.00.01.0009.101920170742 (microcode: 0x0200004d); Topology: Resnet-50 Baseline: FP32, BVLC Caffe* (<https://github.com/BVLC/caffe.git>) commit 99bd99795dcd0b1d3086a8d67ab1782a8a08383 Current Performance: INT8, Intel® Optimizations for Caffe* (<https://github.com/Intel/caffe.git>) commit: Caffe* commit: e94b3ff41012668ac77afea7eda89f07fa360adf, MKLDNN commit: 4e333787e0d66a1dca1218e99a891d493dbc8ef1

³Testing conducted in May 2019, comparing the face recognition time with and without OpenVINO™ on Intel® Xeon® Scalable processors.

Configurations:

NEW: Tested by Intel as of 05/19/2019. 2 socket Intel® Xeon® Platinum 8256 Processor, 4 cores per socket, HT On, Turbo On, Total Memory 374 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.02.01.0008.031920191559, Deep Learning Framework: Intel® OpenVINO® 1.1.133 using Intel® MKL 2019.3. FaceNet topology customized by Digitate*, custom test data, tested using batches of 1 (optimized for latency).

BASELINE: Tested by Intel as of 05/19/2019. 2 socket Intel® Xeon® Platinum 8256 Processor, 4 cores per socket, HT On, Turbo On, Total Memory 374 GB (12 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.02.01.0008.031920191559, Deep Learning Framework: TensorFlow 1.13.1 (Anaconda repo 'tensorflow') using Intel® MKL 2019.3. FaceNet topology customized by Digitate*, custom test data, tested using batches of 1 (optimized for latency).

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

