WHITE PAPER
MARCH 2018

IT@INTEL

Increasing Product Quality and Yield Using Machine Learning

We are now using machine learning to predict issues with tool and relay forecasts into an intuitive, visual format, using customized front-end applications for large-scale activities and web-based solutions for operational and summary activities.

Executive Overview

Sensor data and analytics, powered by a robust Industrial Internet of Things (IIoT) solution can help manufacturers gain actionable insights into the behavior and performance of their tools in real time. With earlier and more accurate information, maintaining tool health becomes easier, which leads to better product yield, higher quality, and lower maintenance costs.

Intel has been investing in factory automation for over three decades, and specifically in IIoT for the last 10 years. We began with collecting data at various factories and applying advanced analytics to solve problems, but growing process complexity required more advanced tools and methods. As compute performance increases and data storage costs decrease, we have continually advanced our analytic solutions to create better insights into our manufacturing process.
We are now using machine learning to predict issues with tool and relay forecasts in an intuitive, visual format, using customized front-end applications for large-scale activities and web-based solutions for operational and summary activities.

Our predictive metrology solution connects process tools and collects rich data to provide:

- Better process equipment control and performance
- A high degree of certainty of product quality and conformity to specifications
- Verifiable engineering lead improvements with process diagnostics

Connecting existing equipment and developing machine learning models, is required for predictive analytics and manufacturing leadership.

Shai Monzon
Industrial Manufacturing Engagement Manager, Intel IT

Stephen Gray
Manufacturing Compute Service Owner, Intel IT

Contents

Contributors

Sarah Kalicin
Data Scientist, Data Center Group, Intel IT

Elaine Rainbolt
Advanced Analytics & AI Engagement Manager, Intel IT

Acronyms

GUI graphical user interface

IIoT Industrial Internet of Things

Background

Manufacturers across industries strive to improve throughput, yield, and product quality for better forecasting, cost reduction, and a competitive advantage. But, inconsistency in equipment performance and difficulty in predicting maintenance requirements often lead to quality issues and longer time-to-market (TTM). For example, industrial bakers expanding distribution and product offerings are faced with maintaining even consistency across multiple ovens. Temperature fluctuations in a single oven can result in products that do not meet consumer expectations and result in higher production costs. Intel has experienced similar consistency challenges as printed patterns on chips have decreased in size while performance demands have skyrocketed. Decreasing dimensions require increased precision, and the slightest variation in tool performance can result in wasted materials, higher costs, and time delays. This requirement for better process measurements affects all manufacturers, regardless of industry. The difference is simply in the margin of fluctuation allowed in a specific product.

Actionable information from tools and processes is necessary to maintain consistency in manufacturing. But collecting the data is not the only issue. IT organizations must store data, clean it, integrate it with other data, and then analyze it for meaningful insights. Like most manufacturers, the challenges we faced include:

- **Shorter process windows.** With a growing market comes increased pressure to deliver products to market faster.
- **Unconnected tools.** Many manufacturing tools are decades old, and while they continue to serve our business needs, they often lack connectivity and data-collecting capabilities. Process-sensitive data is missing from these legacy tools.
- **Summary-level statistics.** Traditional summary statistics do not encompass the richness of available data, making it more difficult to diagnose specific tools and processes.

Sensor data can help engineers verify the behavior and performance of equipment in near real time, providing the opportunity to intervene earlier and with more accuracy to remain within required specifications. Intel's journey to predictive metrology through machine learning began more than a decade ago through capturing data for meaningful insights. The goal then, and now, was near-real-time process control to reduce or eliminate variations.

An IIoT Solution

Smart factories—those making the transition to Industry 4.0—develop a rapid learning cycle, sometimes called "fail fast," to quickly adjust and optimize technology and processes. An Industrial Internet of Things (IIoT) solution can help enable this practice by using big data. But beyond equipment-related cost savings, the visibility gained through data can reveal previously unseen issues that lead to new opportunities.

At Intel, our end-to-end IIoT solution began with academic collaboration to determine scientific measures specific to the wafer production process and how to visually interpret data. For example, process gases and byproducts produced during the process emit different colors of light when energized in a vacuum chamber. When measured with the appropriate instrument and analyzed, these gasses and byproducts provide a unique fingerprint of the processing of that wafer. Our goal was to correlate this fingerprint to the physical properties of the process on the wafer, for example, the depth of a trench. We then developed offline machine-learning models based on the statistics from the fingerprint to forecast wafer quality. In order to implement the prediction models in manufacturing, we developed a near-real-time prototype to collect sensor and contextual data, standardize it, and merge the data into a distributed database environment. We also designed a workflow to manage the flow of the data and score sensor data against the data model. The output from the model provided a near-real-time prediction of product quality. We then designed a second workflow to manage the models. When wafer quality metrics were available from downstream physical measurements of the wafer's properties, we used this data to evaluate the accuracy of the model's prediction and retrain or modify it. The autonomous self-learning and modified routine remove human efforts necessary to maintain the models.

The journey from prototype to a robust IIoT solution included a multi-disciplinary team, standardizing the framework, acquiring and managing the data, and tailoring our analytics and machine-learning tools.

Building the Team

An IIoT project, enhanced with machine-learning models, can be disruptive and can potentially challenge normal operations. Our project involved academia, research and theory, data scientists, and an array of IT specialists to manage the proof of concept (PoC), deployment, and scaling across multiple factories. Stakeholders across the organization participated in developing a process that represents the business model, the usage, the functionality, and the implementation.

Gaining the support of the entire organization was important to our success. We needed participation from a variety of disciplines, including sponsorship at the highest levels. Investing in an IIoT solution was an opportunity to solve many of the daily challenges people experienced. It also required measurable return on investment (ROI). We conducted a series of outreach activities to help senior managers understand the project and its potential value across the organization. To promote participation and investment in the project, we identified the pain points an IIoT solution could address for each stakeholder, proving the project value to management.

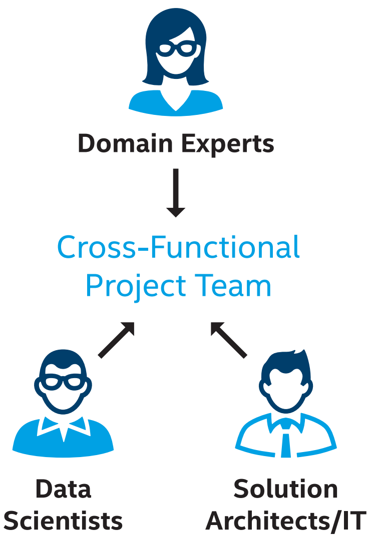
We created a cross-functional project team with the following roles:

Domain Experts

- **Business owners**—or stakeholders—provided project funding and helped us prioritize solution features and feedback.
- **Equipment owners** provided feedback and insights into performance and behavior of equipment on the factory floor. This allowed us to incorporate additional data requirements that increased data richness from the first iteration of raw data—ending equipment isolation—to standardized sources in data-mining and machine-learning stages.

Data Scientists

- **Visual analysts** created mock-ups and rapid prototypes to express user requests, ideas, and concepts into basic technical requirements. Presenting vast amounts of visual data generated by IIoT systems that can be understood and processed by humans is a specialized software development skill that requires human factors knowledge, as well as technical development.
- **Data analysts** helped to create standardized export-orientated data structures for deployment in the models. This helped reduce how long analysts spent on gathering, cleaning, and aligning data, allowing them to focus on higher value activities such as development, selection, evaluation, and deployment of the best model(s) for anomaly detection, machine learning, and product quality prediction. Data analysts, working with visual analysts and software developers, facilitated model deployment and the presentation of the results.



Solution Architects/IT

- **Enterprise architects** created a framework defined by data that ensured future changes were confined to specific modules or workflows within the collection process. The enterprise architects also adopted or defined standard message structures by source and type, and identified existing capabilities to handle storage and additional data types.
- **Software developers** identified raw data structures from existing and new sensor sources and converted them into standard structures for processing during Extract, Transform, and Load (ETL). The structures were divided into usable components and assembled for visualization and analytics to provide just-in-time data with flexible, logical blocks with links for exploitation.
- **IT application and DB engineers** built install packs and patch release kits to ensure uniformity across all environments. To reduce operational complexity and maintain uniformity, the IT application and DB engineers used remote monitoring and verification and reduced manual interactions. This role also identified and codified automated shutdown, start-up, and recovery procedures to ensure data quality and to reduce disruptions.
- **Project managers** managed the competing requirements, various participants and different phases, and helped us incorporate feedback and expand the baseline without negatively impacting production.
- **IT infrastructure engineers** delivered the server, storage/backup, and recovery platforms based on the architectural design. This required the team to deliver a cost-effective high-availability environment that would provide the customer with a secure, scalable, mission-critical solution.

Standard IIoT Framework

We focused on developing a service-oriented architecture (SOA), based on Intel® technology, which separated data from logic. Our primary goals were performance, scalability, availability, capacity, and redundancy. Existing tools were integrated with sensors to provide real-time monitoring and optimization. Data mining provided new patterns and machine learning built predictive models for the business processes.

As we created the framework, we included a feedback loop to prior steps, as well as progressing to the next step. For example, as we completed the data visualization step, not only did we progress to the data-mining layer, but we also went back to the data-integration and data-collection layers to add extra capabilities that the visualization activity highlighted as deficiencies (see Figure 1).

The framework included:

- **Data collection.** The sensor framework collected raw data from existing tools and processes. Data structures varied across sensor types, and we used an SOA for the application framework to ensure future changes could be confined to specific modules or workflows. Initially, we concentrated on identifying the raw data structures provided by existing and new sensor sources and converting them into standard structures for processing during ETL.
- **Data storage.** Data was stored in an analytics-ready structure. The databases were logically distributed with recent data stored on faster hardware, and historical data on slower hardware.
- **Data integration.** To simplify data integration, we adopted standard message structures by source and type. We also extended existing functionality for handling and storage rules for other data types, such as product information. Raw data from tools and process were integrated into the data storage, while retaining the unique origins, time stamps, and other identifying factors for downstream analytics.

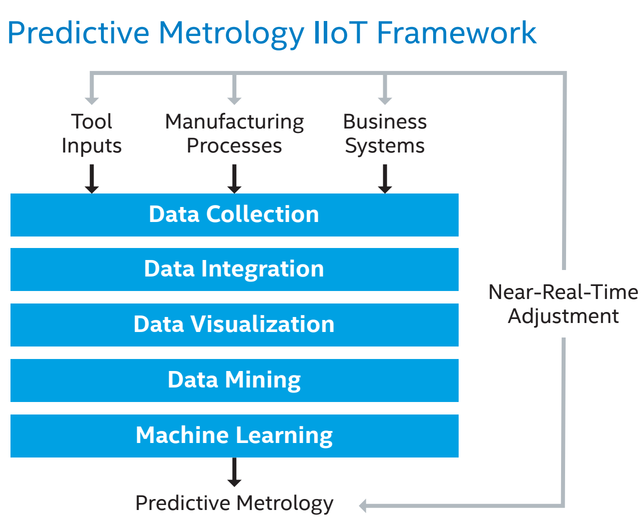


Figure 1. As we developed each stage of the architecture, we progressed toward machine learning and revised the prior stage with additional capabilities. The machine-learning models include feedback to the tools for near-real-time adjustment.

- **Data visualization.** The graphical user interface (GUI) used the standardized data, divided into usable components, and assembled for visualization and analytics. We developed drill-down capabilities to summarize large volumes of data into visual representations that provided meaning to the user. All available data could be represented, viewed, and verified to provide users with confidence in its accuracy.
- **Data mining.** Data mining correlated statistical patterns in large datasets and established relationships between those patterns and the tools, processes, and products.
- **Machine learning.** Machine-learning models predicted product throughput, yield, quality, and tool health. They also created a feedback loop that allowed near-real-time adjustments of tools and processes, and continuously updated the models based on new and changing information.

IoT Sensors and Data Acquisition

For any data-centric project, we need to understand exactly what measurements in terms of process characteristics (in this case, the physics and chemistry) to collect. Machine learning and analytics can only be successful if the underlying data contains the right information supporting known laws of physics and chemistry. Manufacturing equipment typically uses actuators, sensors, and controllers to control input. For example, flow controllers control the quantity of gas defined by the recipe. The process tool reports the flow and what was achieved compared to the baseline. But it does not contain measurements of how the gas chemically reacted in the process chamber or how the byproducts interacted. This chemical reaction measurement is not typically included in the process chamber design, and it can be difficult to measure. It is important to carefully examine what sensor data is available, whether it contains enough information to provide meaningful information on the process, and if it is sensitive to process variabilities. When the chemical and physical nature of a process is too complex to model based on first principles, we correlate the process variability to those observed in wafer quality, using a machine-learning model.

When the chemical and physical nature of a process is too complex to model based on first principles, we correlate the process variability to those observed in wafer quality, using a machine-learning model.

We conducted a pilot project to demonstrate that if we collected the data produced by the tool and integrated it with associated externally generated data, we could predict the quality of the wafer produced by that tool. We also added localized sensors to each tool to collect additional data that augmented what was collected from the equipment. We deployed Intel® Next Unit of Computing (Intel® NUCs) based on the Intel® Core™ processor to acquire and transform these data sources into standardized data structures using an alignment algorithm. This algorithm prepared the data using contextual synchronizing, such as identifying the workload, tool, its state of operation, the time of day, and so on. The data was source-agnostic, which allowed it to be consolidated and blended accurately.

The preprocessed data at the edge also provided lossless data reduction. The data was transmitted to a centralized data store through message queues, and then to a data merge store for analytics.

Based on results from the pilot, we performed a number of framework iterations and learned from each one. For example, we found a more efficient way to send data from the sensors to storage, by converting from binary form to structured form and focusing on the parameters most likely to cause variations that resulted in yield or performance loss.

It is not complex to deploy this platform from the initial pilot into production, as scaling and replicating the environment are supported by a variety of off-the-shelf products, such as third-party cloud solutions for computing and storage. We used standard Intel NUC hardware at the edge with an Intel Core processor in virtual machines (VMs). We used the Intel® Xeon® processor E5 family for data acquisition and storage, which provided excellent compute power. VMs allowed us to easily deploy and manage these processors across the Intel® Architecture (see Figure 2). Using heterogeneous hardware allowed us to best match the analytics needs of the factory to the machine-learning tools' capabilities and connectivity constraints. The most important aspect is that the factory must be networked to collect and transmit the data, especially to achieve near-real-time insights. Reliable and robust connectivity is the foundation on which all IIoT platforms are built; without it, projects can fail to meet expectations and achieve business requirements.

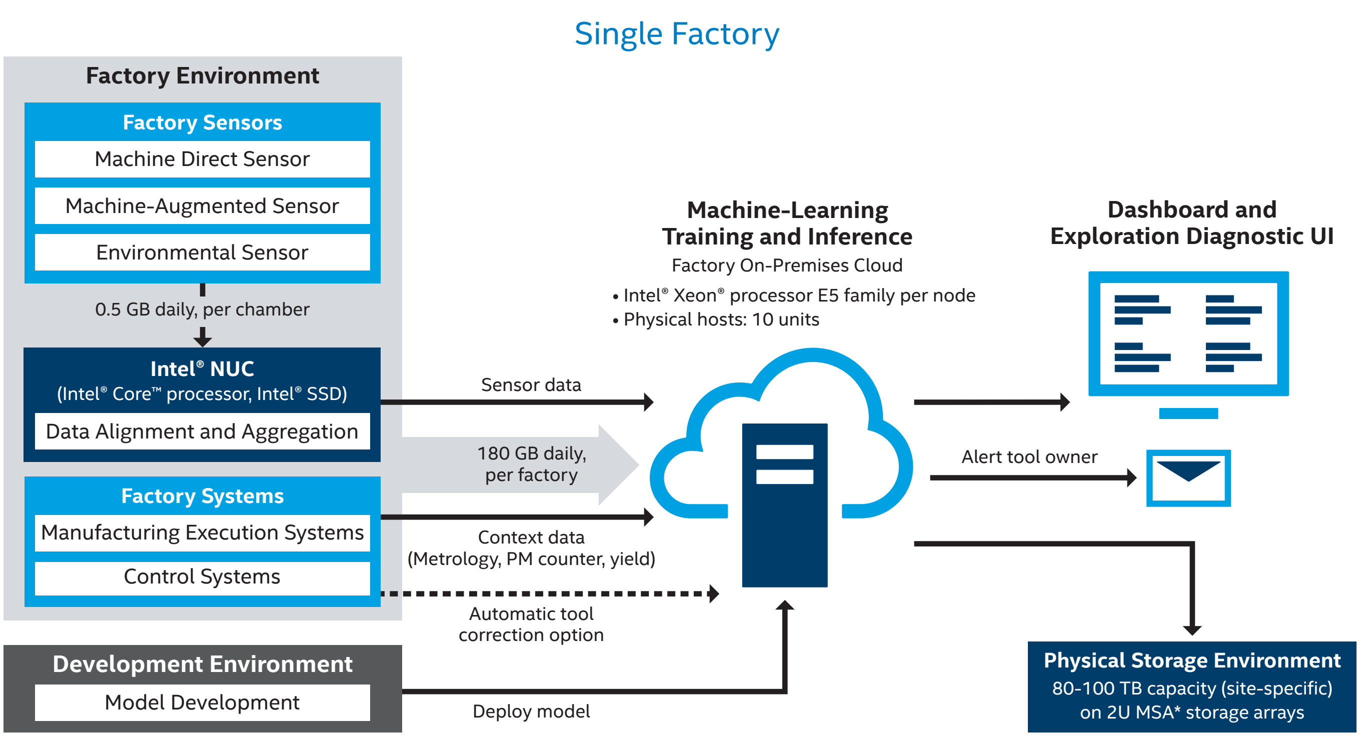


Figure 2. Data was preprocessed at the edge, then merged with standardized data from multiple sources to be analyzed with machine learning.

Data Analytics with Machine Learning

We use the sensor data, actual metrology, and predictive metrology to train machine-learning models to predict the product condition on completion of the manufacturing activity. We evaluate the model predictions based on how well they predicted wafer quality weeks before the wafers were measured at end of the line. By comparing the prediction to actual results, as well as known bad wafers, we can establish a baseline and then further evaluate different modelling techniques.

We use the Random Forest algorithm because it allows us to avoid overfitting the model and potentially decreasing the accuracy of results. The classifier can handle missing values and be modelled for categorical values. We also use principal component analysis (PCA) to reduce the dimensionality of datasets with multiple correlated variables. Using PCA, we summarize data, while being sensitive to relative scaling. The model runs on an Intel Xeon processor E5 family-based high-performance server.

Our current machine-learning model automatically updates itself based on a defined threshold that triggers additional learning. Presently, updates can occur as often as every six hours up to as infrequently as once every two-and-half months.

We also developed a web-based graphical user interface (GUI) to provide intuitive access to the model and data (see Figure 3). Using data visualization to monitor the results of the model and verify its accuracy is very useful. The interactivity of graphs allows human experts to dive deeper to investigate process issues.

Data, such as time-series, is linked for analytics, allowing users to drill into details of lots, wafers, and tools. Naturally, this created a challenge because we wanted to avoid data duplication; ultimately, we decided that the data structure was necessary for analytics, which took precedence over storage constraints.

The continuously iterating models will improve and evolve as we understand how to use the data, identify additional data needs, and refine the platform.

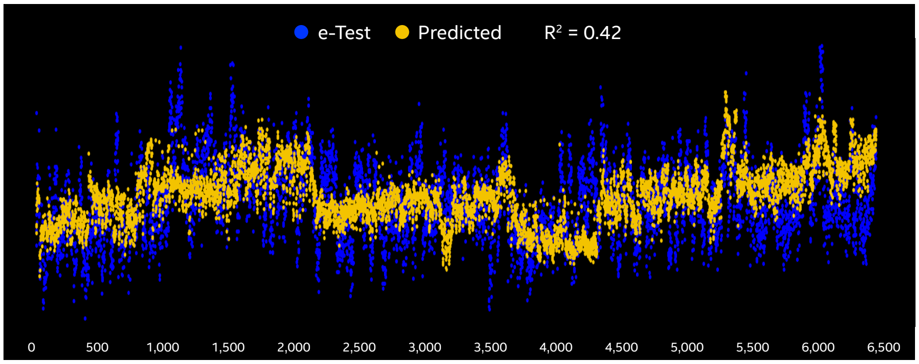


Figure 3. The graphical user interface (GUI) provides an intuitive view of the prediction model against the actual wafer measures at the end of the line.

Anticipating and Adjusting the Model

It is critical to anticipate and adjust the model to achieve ongoing success and deeper insights. We initially underestimated the amount of time it would take to transition from the prototype to the production process. It was necessary to redefine code, and anticipate scenarios outside of the prototype environment, to achieve the level of scalability we needed.

We also discovered that we needed to allow for adequate time to explore new process discoveries, which are a natural outcome of Industrial Internet of Things (IIoT) solutions. We refresh the machine-learning model based on new training data regularly, and often the results of these adjustments come months after they have been made. Model retraining—a very resource-intensive and time-consuming activity—is tailored to the purpose of the model and can be scheduled based on time or events. The schedule is determined on a model-by-model basis and optimized to balance the refresh need against the time and cost required. Future plans include developing a model-for-model refresh function that further automates this activity.

Results

With the combination of a collaborative team environment, strong executive sponsorship, and determination to create an impactful advanced analytics solution, analysis that once took days can now be performed in minutes. The benefits of our IIoT solution include (see Figure 4):

- **Tool control and matching.** We can now determine mismatches in tools with near-real-time adjustments of recipes. We can also predict maintenance requirements, reducing the need for preventive intervention and increasing equipment availability and utilization. Cross-tool comparisons allow us to make equipment fleet improvements by identifying deltas in performance between functioning and malfunctioning equipment.
- **Predictive metrology.** We attained prediction with a high degree of certainty within targeted limits, reducing the need for dynamic sampling. This allowed for the reduction in physical measurements of the product, increasing throughput without sacrificing quality.
- **Process diagnostics.** We gained better insights into equipment behavior and condition with verifiable engineering lead improvements.
- **Fault detection.** With known mathematical relationships between sensor data and metrology, we can more accurately predict faults in materials created by specific tools. Also, time-series data takes advantage of feature richness that traditional summary statistics often do not encompass.

Many tools' sensors did not provide data that was granular enough; therefore, instead of adding sensors, we collected the additional data from log files already created, but not currently being used to their full potential. By deploying new sensors to these existing tools, we can help equipment owners move from uni-variant to multi-variant analysis, gaining better insights.

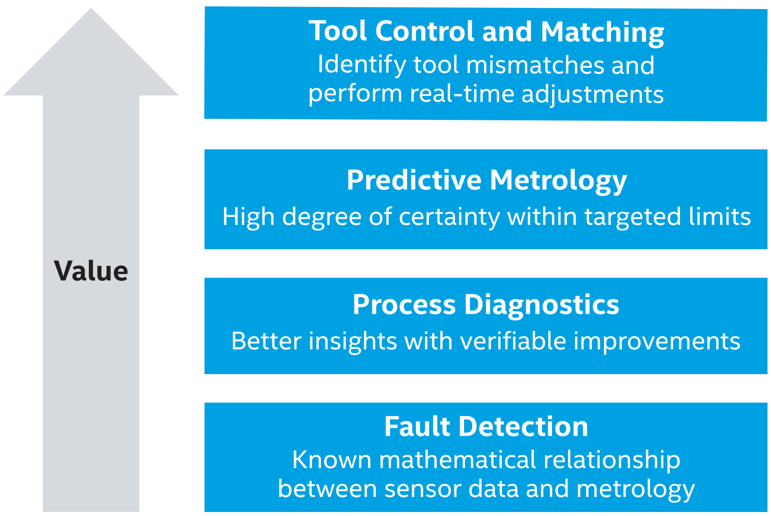


Figure 4. Our predictive metrology process with machine learning has delivered better tool control, higher certainty, better tool matching, and the ability to make near-real-time adjustments to recipes.

Conclusion

Many manufacturers across a wide variety of industries face similar challenges: decreased process windows, unconnected tools, a lack of meaningful data, or too much data. But connecting the unconnected with IIoT can bring new insights that improve product quality and yield, as well as reduce the cost of tool maintenance and unexpected downtime.

We combined our IIoT solution with a machine-learning model to predict wafer quality and yield, using detailed data that also allows us to track tool performance. Our multi-year journey to predictive metrology included a multi-disciplinary team, connecting existing tools, developing a standard framework, and using models that self-manage and self-learn. With management's continuous support and a dedicated team, the project has delivered meaningful results.

The results have provided us with better tool control, higher quality products, better tool matching, and near-real-time adjustments on the factory floor. The solution can be expanded to additional tools and industries, and we have started work in these areas. We will continue to invest in this solution as business needs change, and to match advances in machine learning, storage technologies, and compute performance.

For more information on Intel IT best practices, visit **intel.com/IT**.

IT@Intel

We connect IT professionals with their IT peers inside Intel. Our IT department solves some of today's most demanding and complex technology issues, and we want to share these lessons directly with our fellow IT professionals in an open peer-to-peer forum.

Our goal is simple: improve efficiency throughout the organization and enhance the business value of IT investments.

Follow us and join the conversation:

- Twitter
- #IntelIT
- LinkedIn
- IT Center Community

Visit us today at **intel.com/IT** or contact your local Intel representative if you would like to learn more.

Related Content

If you liked this paper, you may also be interested in these related stories:

- Connecting the Unconnected video
- Joining IoT with Advanced Data Analytics to Improve Manufacturing Results paper
- Horizontal IoT Platform Paves the Way to Enterprise Success paper