

Optimizing Manufacturing with the Internet of Things

Intel manufacturing advances operational efficiencies and boosts the bottom line with an IoT and big data analytics pilot

Intel proves IoT success by using big data analytics to bring cost savings, predictive maintenance, and higher product yields to its own manufacturing processes.

Introduction

The volume, variety, and velocity of data¹ being produced in manufacturing is growing exponentially, creating opportunities for those diligently analyzing this data to gain a competitive edge, respond to changing market dynamics, and increase manufacturing margins, productivity, and efficiency. Equipment across the factory floor is generating thousands of different data types, such as unit production data at multiple levels, equipment operation data, process data, and human operator data, which can be stored to be made available for analysis.

Although large manufacturers have been using statistical process control and statistical data analysis to optimize production for years, the extreme composition of today's data provide opportunities to deploy new approaches, infrastructure, and tools. Manufacturing industries are ready to embrace the use of big data, supported by higher compute performance, open standards, the availability of industry know-how, and the vast availability of highly skilled, data statisticians fueled by academics. With access to new manufacturing intelligence, manufacturers will improve quality, increase manufacturing throughput, have better insights into root cause of manufacturing issues, and reduce machine failure and downtime. With these new business values and technology capabilities, manufacturers will be able to change business models and practices in order to optimize

designs for manufacturability, thereby improving supply chain management and introducing the use of customized manufacturing services to shorten time to market for products customized for smart consumers across various geographies.

This paper outlines an Internet of Things (IoT) pilot in one of Intel's manufacturing facilities to show how data analytics applied to factory equipment and sensors can bring operational efficiencies and cost savings to manufacturing processes. With industry collaboration from Cloudera, Dell, Mitsubishi Electric, and Revolution Analytics, this IoT big data analytics project is forecasted to save millions of dollars annually along with additional return on investment business value.

Challenge: How to extract value from all of the data in manufacturing?

Big data is characterized by huge data sets with varied data types, which can be classified as structured, semi-structured, or unstructured, as shown in Table 1. Structured data fits nicely into neatly formatted tables, making it relatively easy to manage and process. Structured data has the advantage of being easily entered, stored, queried, and analyzed. Examples include manufacturing data stored in relational databases, and

data from manufacturing execution systems and enterprise systems. On the other hand, unstructured data such as images, text, machine log files, human-operator-generated shift reports, and manufacturing social collaboration platform texts may be in a raw format that requires decoding before data values can be extracted. Semi-structured data is a form of structured data that does not conform to the formal structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements

and enforce hierarchies of records and fields within the data. In manufacturing, the power of big data technology stems from the ability to merge and correlate these data set types to create business value through newfound insights. The other value proposition of big data technology is it allows manufacturers to aggregate and centralize various types of data in a cost-effective, scalable manner.

In manufacturing, process variability stemming from various factors, like material, process recipes and methods, and equipment differences, drives a real business need for manufacturers to turn to a big data solution based on a scalable platform that can grow with their businesses and manufacturing requirements. Machine data is strongly correlated to yield, quality, and output, thereby providing valuable information to proactively detect processes that are getting out of control. However, some types of manufacturing generate massive data files (gigabytes in a few days per tool type, as shown in Table 2), limiting the ability to store, analyze, and extract useful information from them using conventional methods. Without the use of big data technologies, it is extremely hard to even visualize the information in large data sets from various sources.

MANUFACTURING INDUSTRIES EXAMPLES	REAL-TIME, SEMI-STRUCTURED DATA	UNSTRUCTURED DATA	STRUCTURED DATA
Semiconductor	Machine builder standards like	Operator shift reports	RDBMS databases
Electronics	SECS/ GEM, EDA or custom-based on	Machine logs	NoSQL*
Solar	COM, XML	Error logs	Enterprise data warehouse
Machinery	Sensors (vibration, pressure, valve, and acoustics), Relays	Texts	Files stored in manufacturing PCs
Energy	RFID	Vision images	Spreadsheets
Automobiles	Direct from PLCs, Motor and drives,	Audio/Video	
Aerospace	Direct from motion controllers, robotic arm	Manufacturing collaboration social platforms	
Chemical and Pharma			
Metalworking			
Food and Beverage			
Pulp and Paper			
Clothings and Textiles	Manufacturing historians (time series data structures)		
Furniture			

Table 1. Manufacturing data examples

DATA TYPES	DATA SIZE (per week)	EXAMPLES
Machine parameters and error logs	~5 GB per machine	Used to monitor machine performance: dispense height, placement (x, y, z), belt speed, flow rate, oven temperature, laser power, etc.
Machine events	~10 GB per machine	Used to measure process time: start dispense, end dispense, start setup, and end setup
Defect images from vision equipment	~50 MB per unit or 750 GB per lot	Used to identify root cause of failure modes, defect commonality, defect mapping

Table 2. Data size examples

Addressing the challenge: IoT pilot using big data analytics server and IoT gateway in Intel manufacturing

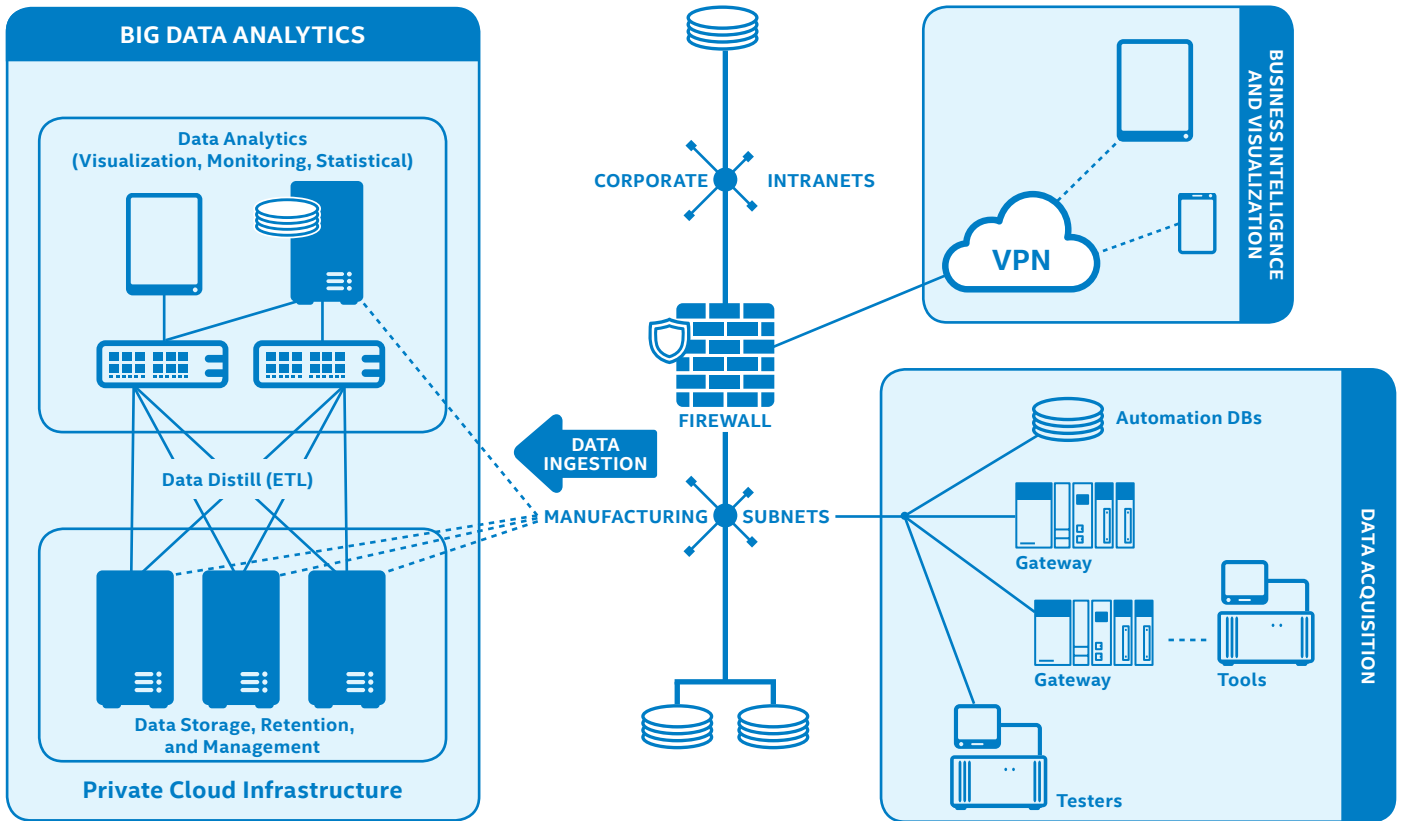


Figure 1. Building blocks for end-to-end infrastructure enabling manufacturing intelligence from the factory floor to the datacenter

Figure 1 shows a high-level IoT manufacturing architecture for small to large data sets that constitute data acquisition and aggregation of various types of data from the manufacturing shop floor and the manufacturing network, which opens up the possibilities of visualizing, monitoring, and data mining for new business intelligence.

As an example, the architecture can: clean, extract, transform, and consolidate structured data from existing databases and unstructured data from tool sensors and log files from legacy equipment in a data store platform (i.e., Hadoop*). The data can then be visualized and analyzed by high-level factory applications running in different virtual machines on the same on-premise server as the data store server. Alternatively, the data can

be accessed with other analytical or monitoring applications on the network. Other enhanced capabilities could include running analytics in Hadoop or other types of file systems, or running analytics in memory for faster performance. The results of the analysis can be presented to users via intuitive visualization capabilities in the business intelligence layer of the network.

Big data analytics IoT server setup in Intel manufacturing pilot

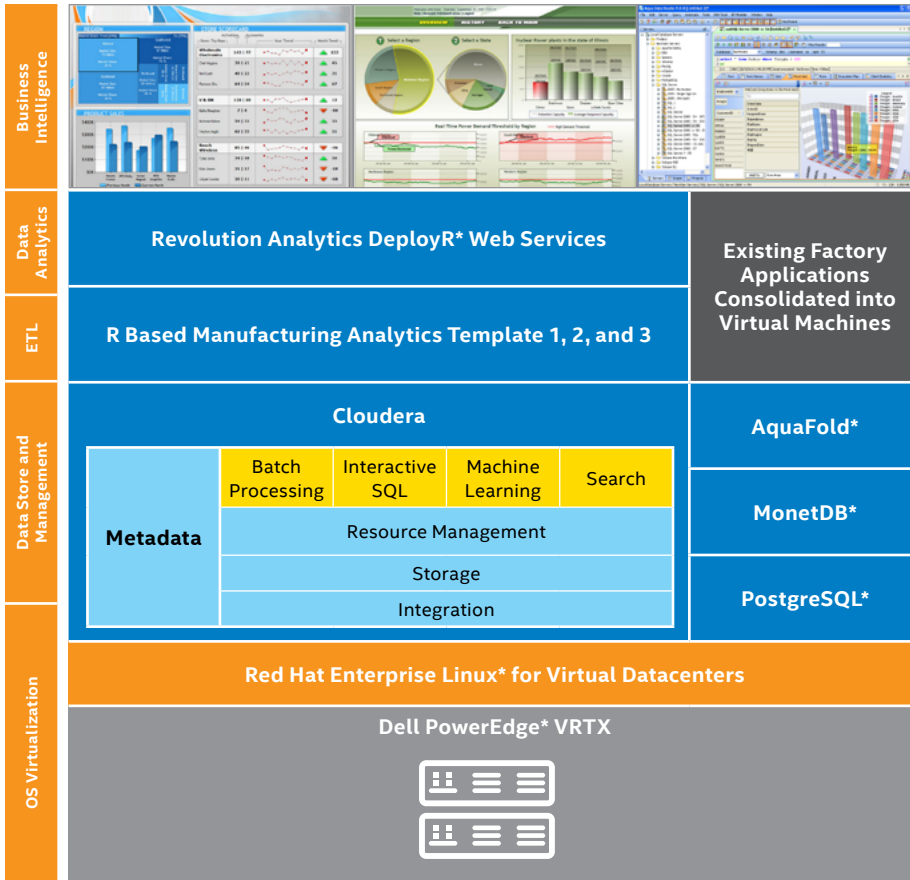


Figure 2. Big data analytics server

The big data analytics server used by Intel manufacturing in its pilot deployment is shown in Figure 2. A compact system from Dell, the PowerEdge* VRTX,² was selected to host the big data and analytics software in a private cloud setting as an on-premise server. The system consists of a Dell PowerEdge VRTX chassis with twenty-five 900GB hard disks and two Dell PowerEdge M820 blade servers, each equipped with four Intel® Xeon® processors from the E5-4600 product family on each blade. The Intel Xeon processor E5-4600 product family provides a dense, cost-optimized, four-socket processor solution with up to eight cores, up to 20MB of

last-level (L3) cache, and up to 1.5TB maximum memory capacity, along with communication pathways to move data more quickly.

The two M820 servers host analytics and application software, and the Hadoop nodes, which run in multiple virtual machines. Red Hat Enterprise Linux* for Virtual Datacenters operating system provides a complete virtualization software solution for servers designed for a scalable and fully virtualized datacenter.

Analytics and Application node

Figure 3 shows how the software is allocated to different virtual machines (VMs). The node hosts five VMs that run the various analytics and application workloads. This includes:

Revolution R Enterprise* from Revolution Analytics

By linking with high-performance Intel® MKL multithreaded math libraries, Revolution R Enterprise uses the power of multiple processors to accelerate common matrix calculations. In addition, Revolution R Enterprise parallel, distributed algorithms enable statistical analysis of Big Data to scale linearly with multi-core servers, high-performance computing clusters, big data appliances, and Hadoop*.

- **Revolution R Enterprise* from Revolution Analytics** is analytics software built upon the powerful open source R statistics language. The software provides a seamless, secure data bridge between analytics solutions and enterprise software, thereby solving a key integration problem faced by businesses adopting R-based analytics alongside existing IT infrastructure.³
- **MonetDB***: an open source column-oriented database management system designed to provide high performance on complex queries against large databases, such as combining tables with hundreds of columns and multimillion rows. MonetDB has been applied in high-performance applications for data mining, online analytical processing (OLAP), geographic information systems, and streaming data processing.⁴
- **PostgreSQL***: a powerful, open source object-relational database system used for online transaction processing (OLTP).⁵

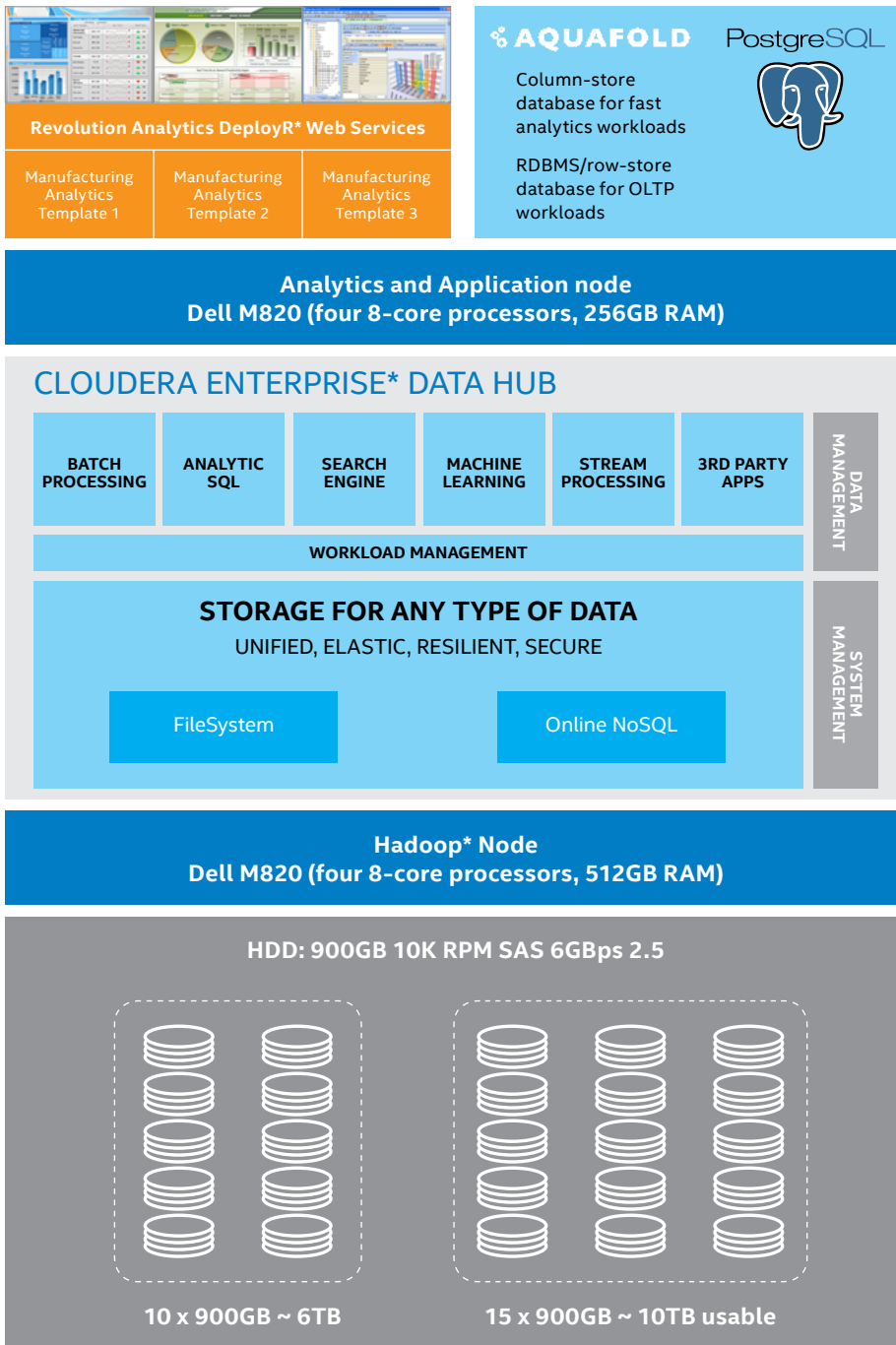


Figure 3. Software allocations to virtual machines on the big data analytics server

- **AquaFold***: an application server used to build and deploy production quality database and reporting applications quickly. It includes capabilities like multiple sources data synchronization, cross-database data migration, data warehousing extract transform and load, scheduling data exports/ imports, and custom business intelligence dashboards. Supports communication to various production databases in the network like PostgreSQL, Microsoft SQL*, Oracle*, and IBM DB2* with ODBC* and JDBC* connections.⁶

Hadoop* node

Four virtual machines are provisioned to run a basic, four-node Cloudera Hadoop cluster, consisting of one head node and three worker nodes.

- **Apache Hadoop** is an open source distributed software platform for scalable distributed computing. Written in Java, it runs on a cluster of industry-standard servers configured with direct-attached storage and cost-effectively scales performance by adding economical nodes to the cluster.
- **Cloudera Enterprise Data Hub* (CDH)** offers a unified platform for big data by providing one place to store, process, and analyze all of their data resulting in an extension of existing investment value and enabling fundamental new ways to derive value from their data. CDH is 100 percent Apache licensed open source and is unique in offering unified batch processing, interactive SQL, and interactive search and role-based access controls.⁷

A closer look at the Cloudera stack

Cloudera Enterprise Data Hub* (CDH) delivers the core elements of Hadoop*—scalable storage and distributed computing—along with additional components such as a user interface, plus necessary enterprise capabilities such as security, and integration with a broad range of hardware and software solutions.

CDH can help businesses transform and accelerate the way big data is used. When combined with datacenter architecture based on Intel® Xeon® processors, Cloudera and Intel provide a comprehensive approach for industrial and manufacturing businesses seeking to analyze data to make factories run more efficiently.

Internet of Things (IoT) Gateway

An Intel® Atom™ processor-based gateway by Mitsubishi Electric, called the Mitsubishi Electric C Language Controller of MELSEC-Q series*,⁸ is used to aggregate and securely ingest data into the big data analytics server. Data ingestion is the process of validating, filtering, and reformatting the data to make it easier for the big data analytics software to work on it.

The Mitsubishi Electric C language controllers of MELSEC-Q series are embedded solutions equipped with numerous features characteristic of intelligent systems, including robust network connectivity and the high computational performance needed to process large amounts of data

collected from sensors or via the network when supporting sophisticated system control and operations. At the core of this controller is a hardware platform based on Intel® architecture and the Wind River VxWorks* real-time operating system.

Mitsubishi Electric developed C language controllers of MELSEC-Q Series to satisfy the diverse requirements of factory automation, including excellent reliability, tolerance of harsh environments, and long-term availability. These features make it a robust and reliable product that requires little maintenance for IoT manufacturing applications.

In place of ladder logic used in conventional programmable logic controllers, the C language controllers of MELSEC-Q series use the international standard C languages (C and C++) for greater programming flexibility. This allows users to take full advantage of their existing C language software and development.

CIMSNIPER* is a data acquisition and processing software package for Mitsubishi Electric C Language Controller of MELSEC-Q series. It can collect process data (including SECS messages) and manufacturing equipment errors without modifications of existing systems.

Big data analytics cases at Intel manufacturing

Over the past two years, Intel has developed more than a dozen big data projects that have bolstered both operational efficiency and the bottom line. Here are a couple of examples:

Reduced product test times

Every Intel® chip produced undergoes a thorough quality check involving a complex series of tests. Intel found that by using historical information gathered during manufacturing, the number

of tests required could be reduced, resulting in decreased test time. Implemented as a proof of concept, this solution avoided test costs of \$3 million in 2012 for one series of Intel® Core™ processors. Extending this solution to more products, Intel expects to realize an additional \$30 million in cost avoidance.⁹

Improved manufacturing monitoring

Data-intensive processes also help Intel detect failures in its manufacturing line, which is a highly automated environment. Intel pulls log files out of manufacturing tools and testers across the entire factory network, which can be up to 5 Terabytes an hour. By capturing and analyzing this information, Intel can determine when a specific step in one of its manufacturing processes starts to deviate from normal tolerances.¹⁰

Specific to the current end-to-end platform pilot deployment discussed here, Intel, in collaboration with Mitsubishi Electric, Cloudera, Revolution Analytics, and Dell, successfully pioneered capabilities that have made tremendous headway in the use of data mining sciences to solve practical manufacturing issues, thus saving Intel millions of dollars through cost avoidance and improved decision making. The major goal of this project is to extract the value propositions of data and data analytics to obtain better insights in predictive manufacturing and reduce manufacturing costs without sacrificing throughput or quality.

The following details some of the groundbreaking work and discoveries Intel accomplished through the integration of big data analytics and technologies for the IoT in manufacturing.

Pilot Results:

Use Case 1: *Reducing non-genuine production yield loss through the monitoring and analysis of machine parametric values and timely replacement of parts before they fail*

Background

Automated Test Equipment (ATE) is a machine that is designed to perform tests on different devices referred to as a devices under test (DUT). An ATE uses control systems and automated information technology to rapidly perform tests that measure and evaluate a DUT.¹¹ The ATE systems interface with an automated placement tool, called a handler, that physically places the DUT on a Test Interface Unit (TIU) so that it can be measured by the equipment.

Problem Statement

Defective TIUs will wrongly categorize good units as bad, which negatively impacts Intel manufacturing operation costs. Defective TIUs can cause DUTs to be wrongly categorized, including the rejection of good units. Intel manufacturing's objective was to detect TIU defects before they happen so they could repair or replace them before units are wrongly categorized. When a faulty TIU wrongly categorizes good units as bad, the units are scrapped. Some components were replaced with spare parts during regular preventive maintenance, even if they were still operating properly to avoid such problems.

Results and Benefits Yield

The analytics capability predicted up to 90 percent of potential TIU failures before being triggered by the existing factory's online process control system. This helped, in this case, to replace defective TIUs before causing over rejection of good units, thus reducing

yield losses by up to 25 percent.¹² In addition, Intel was able to save spare cost by reducing the need to replace spare parts before they fail during preventive maintenance resulting in an estimated 20 percent reduction in spare cost.

Use Case 2: *Reducing yield losses by eliminating and minimizing incorrect ball assembly in ball attach equipment*

Background

The ball attach module is where solder paste is printed onto the lands on the substrates. Solder balls are placed into the ball attach lands, and the paste holds them in place. The entire package goes through a reflow oven, which melts the paste and balls to the substrate lands.

Solder balls are vacuumed onto tiny holes of the placement head. The head is inspected for excess or missing balls. Once the head is aligned to the substrates, the balls are placed in the solder paste on the substrates. After releasing the balls, the placement head is inspected for any remaining balls. Finally, the substrates are inspected by a camera vision system for any missing or shifted balls.

Problem Statement

Units with missing balls are faulty material and a yield loss. There are numerous scenarios in which balls are missing from the units, including inadequate vacuum pressure.

Results and Benefits

By visualizing and correlating sensor readings with various machine data and execution system data, Intel was able to reduce yield losses, optimize maintenance cost, and avoid sudden equipment downtime.¹² This enabled the technicians to proactively fix the problem, toward the journey of having a predictive maintenance capability.

Use Case 3: *Using image analytics to identify good or defect units*

Background

A machine vision equipment is a module that screens units and categorizes them into good and marginal units. The good units are sent forward to be processed while the marginal ones are inspected and determined by a manufacturing specialist to be either good or defective. This manual process takes time.

Problem Statement

The manual process to inspect and categorize marginal units is cumbersome and sometimes takes about 8 hours to successfully segregate a host of true reject units from marginal ones.

This is caused by the time it takes for the units to reach the operator, flow to a segregation module, and to ultimately be segregated. Image analytics enables identification of reject units moments after being inspected by the inspection module.

Results and Benefits

The marginal images recorded at the machine vision equipment module are preprocessed. Each image, which is unstructured data, is resized, cropped, converted into grayscale before converting each pixel to binary. The next stage of the process involves feature selection, where the unstructured image is defined by a set of distinct values. These values are then fed into various machine learning algorithms to determine true rejects and marginal rejects.

The image analytics shortens the time to save true rejects from a pool of marginal units. Image analytics identifies defects roughly 10 times faster than the manual method.¹²

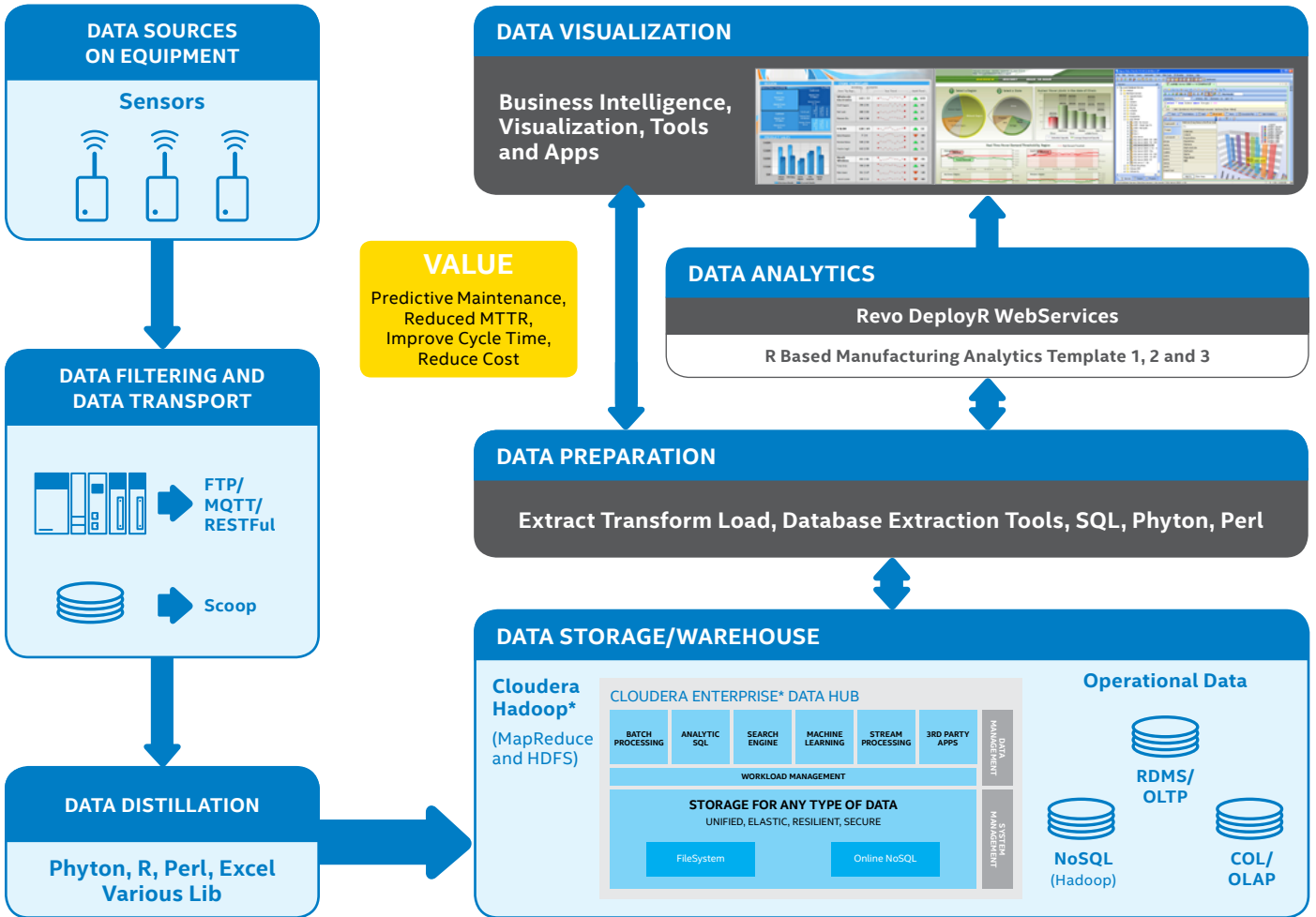


Figure 4. High-level data flow

Data Flow

Figure 4 shows how data flows from the previously discussed use cases.

- Machine data and sensors data are sent by the Intel Atom processor-based gateway to the Big Data Analytics Server (BDAS) as they are acquired in real time. For instance, machine data from the ball attach module acquired over machine interface ports and from analog sensors are streamed at a velocity of a few MBs per a few seconds.

- All incoming factory data is stored in Hadoop*. The following, non-exhaustive list shows some of the possibilities that can be achieved using readily available capabilities in Hadoop:
 - HTTP: The Big Data Analytics Server exposes an authenticated REST HTTP endpoint that supports operations into HDFS.
 - Apache Sqoop* provides the connectivity tool for moving data from non-Hadoop data stores—such as relational databases and data warehouses—into Hadoop.
 - Flume* can receive a stream of continuous log data.
- The data specific to these use cases described in the section above are in the form of Comma-Separated Value (CSV) files or raw images. While the Hadoop ecosystem includes plenty of ways to ingest data (as described in the previous point), some of the machines in the factory have limited network transfer capabilities, which required custom engineering to deliver data into HDFS:

- FTP: The IOT gateway has an FTP client that periodically connects to the Big Data Analytics Server and transfers over the latest acquired data directly into HDFS. Other streaming protocols like MQTT and REST can be used depending on the real time streaming and analytic requirements.
- CIFS share (Windows share): The Big Data Analytics Server provides a Windows/CIFS share directory that the gateway can copy files into.
- The CSV files are directly imported into HBase* using Pig, while the raw images first go through a preprocess using a map-reduce job using computer vision techniques to produce a text data representation of the image.
- Depending on the operational requirements, the data is stored in one of three databases: NoSQL (Hadoop), RDMS/SQL, or Coli/OLAP.
- The databases are accessed and processed using various tools, including AquaFold, ad hoc reporting, workflow scheduling, and ETL and database integration. At the same time, the Cloudera Distribution for Hadoop performs various operations on the data.
- The processed data is further analyzed using Revolution Analytics tools designed for factory applications.
- The data is presented to users in easy to understand dashboards.

Summary of key learnings and conclusions

Intel integrated and validated a big data analytics on-premise server solution with data extracted from Intel's own manufacturing network and from equipment and sensors through the use of Internet of Things gateway to validate the business value of Internet of Things in manufacturing. The Mitsubishi Electric C language controllers of MELSEC-Q series were used. The pilot involved close collaborations between the factory engineers, the IT department, and industry experts from Cloudera, Dell, Mitsubishi Electric, and Revolution Analytics. The team started leveraging existing machine performance and monitoring data, then proceeded to leverage big data analytics and modelling to ingest the data to predict potential excursions and failures. Being able to predict the machine component failures allows engineers to repair and prevent the excursion, hence reaping tremendous savings from wasting production units, time to repair, and machine components.

An integrated suitcase of various software building blocks on the Big Data Analytics Server and IOT gateway were used. This framework can be applied and implemented for manufacturers which have not started leveraging on the intelligence contained in manufacturing data. Manufacturers that have already been using data to improve their efficiency can incrementally add to their existing capabilities to evolve their data mining and analytics capability to the next level.

Big data analytics and the Internet of Things in manufacturing as an end-to-end platform is the critical backbone to enable the vision of smart manufacturing. This platform is scalable and available in various configurations using currently available industry-standard building blocks.

For Intel, this pilot is forecasted to save millions of dollars (USD) annually with additional return on investment business values that Intel is still realizing. Benefits include improving equipment component uptime, minimizing wrong classification of good units as bad (hence increasing yield and productivity), enabling predictive maintenance, and reducing component failures. There are many other types of parametric, metrology, product, and equipment data—both structured and unstructured within Intel manufacturing environment and machines—which could be mined and analyzed to extract new business values. Capitalizing on this opportunity will enable further efficiency and productivity enhancements to the factory and ultimately create a competitive advantage.

1. Gartner analyst Doug Laney introduced the 3Vs concept in a 2001 MetaGroup research publication, 3D data management: Controlling data volume, variety and velocity. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
2. www.dell.com/us/business/p/poweredge-vrtx/pd
3. www.revolutionanalytics.com/revolution-r-enterprise
4. en.wikipedia.org/wiki/MonetDB
5. www.postgresql.org/about
6. www.aquafold.com/
7. www.cloudera.com/content/cloudera/en/products-and-services/cloudera-enterprise.html
8. www.mitsubishielectric.com/fa/products/cnt/plcq/items/index.html
9. Source: "2012-13 Intel IT Annual Report," www.intel.com/content/www/us/en/it-management/intel-it-best-practices/intel-it-annual-performance-report-2012-13.html.
10. Source: "Intel Cuts Manufacturing Costs With Big Data," InformationWeek, March 18, 2013, www.informationweek.com/software/information-management/intel-cuts-manufacturing-costs-with-big-data/d/d-id/1109111.

The TCO or other cost reduction scenarios described in this document are intended to enable you to get a better understanding of how the purchase of a given Intel product, combined with a number of situation-specific variables, might affect your future cost and savings. Nothing in this document should be interpreted as either a promise of or contract for a given level of costs.

11. Source: <http://www.techopedia.com/definition/2148/automatic-test-equipment-ate>

12. Results might vary depending on package size, process, and equipment used in the manufacturing process.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order. Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web site at www.intel.com.

Copyright © 2014 Intel Corporation. All rights reserved. Intel, the Intel logo, Intel Atom and Xeon 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.

