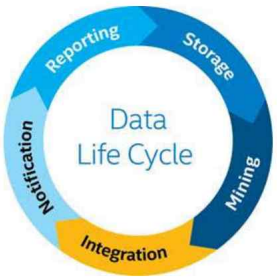
Internet of Things and Large Scale Data Analysis in Intel's Manufacturing Environment
Steve Chadwick – principal engineer, Intel



Motivation

Thoroughly understanding the data life cycle and managing it more efficiently allows Intel IT to protect information, solve complex problems, and manage the IT environment effectively.

Approach

Data Life Cycle is Key to Harvesting Wisdom from Data

Each stage of the data life cycle includes activities and requirements to optimize the value of the data. The data life cycle consists of the following stages:

- Data storage. We use storage methods that can handle tens to hundreds of data points per second, terabytes of summary data, rapid read/write capabilities, and archiving. We've carefully designed our storage system with an eye toward future data growth.
- Data mining. Mining the data in accordance with Intel Privacy Principles (see the Intel white paper, "Applying Privacy Principles in a Rapidly Changing World") helps discover correlations in data from a single source and reveals new insights. We choose CPUs and networks that can meet high-performance computing needs. We think of it as the following equation:

Raw transactions (data) + Business logic (information) = Reports (wisdom)

- Data integration. To solve complex problems and create meaningful insights, we perform data integration across disparate sources. This integration provides multiple, distinct viewpoints that, when combined, reveal a holistic view.
- Notification. To make notifications as effective as possible, we prioritize them, deliver them to appropriate devices, highlight what is most important, and when appropriate, prompt a response.
- Reporting. We deliver actionable information tailored to the device and user receiving it, using HTML5-based content that is clickable and easily consumable.

Organized by:

Results

Manufacturing Analytics in Action

The following two examples illustrate how much business value can be generated by concentrating on the data life cycle, using it to power advanced analytics and targeted reporting. You can read about more examples in the IT@Intel paper, "Joining IoT with Advanced Data Analytics to Improve Manufacturing Results."

- Example #1: Evaluating Manufacturing Tool Health
  Outcome: Analysis that used to take 4 hours now takes only 30 seconds

Engineers at Intel's factories are responsible for tool health. Using sensor data and investigatory interaction, these engineers can now run analyses that determine what sorts of tool health data is meaningful, what decisions need to be made in which timeframes, and under what conditions should notifications be sent.

- Example #2: Consuming Large-Scale Sensor Data (aka fault detection classification)
  Outcome: Ability to process over 5 billion data points per day, resulting in measurable improvement in equipment availability and yield

Many of our factory sensors on equipment collect hundreds of data points per second. We extract this data and use it for both real-time detection and end-of-line correlation. Engineers and manufacturing technicians use the meaningful results to refine equipment behavior.

Because sensor data can often be large and complex, we consider the following when designing our storage systems and algorithms:

- How will the data be consumed?
- How is the data stored (schema)?
- Where does the data reside? How fast can it be retrieved (affects hardware choices)?

It is important to start with something simple and build from there. The time series analysis we needed to do is quite complex, but our first attempts began with only a single variable. (Remember, Rome wasn't built in a day.)

References

IT@Intel paper, "Joining IoT with Advanced Data Analytics to Improve Manufacturing Results."

Intel Privacy Principals: http://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/applying-privacy-principles-in-rapidly-a-changing-world.html

Organized by: