Implementing a Global Machine Data Collection System Across Many EMS Factories

Myckel Haghnazari (Flex Ltd),
Timothy M Burke, PhD (Arch Systems Inc.)
Milpitas, California and Palo Alto, California

Originally presented at IPC-APEX Expo March 6-12, 2021

Abstract

As EMS Providers such as Contract Manufacturers look forward to Industry 4.0, their need for complex data analysis to inform manufacturing decisions takes on even more significance. Managing manufacturing data has become as important to operational excellence as managing the solder paste and electronic components placed on each circuit board. However, while there are many examples of small-scale, operational data collection systems that inform real-time factory operations, there are relatively few examples of large-scale historical machine data collection systems. In this paper we describe one such system and discuss both the advantages it brings as well as the practical implementation challenges to build and maintain it at scale. It is expected that in the years to come many EMS factories will find use for such systems as increasingly proven use cases involving complex machine learning, predictive maintenance and even artificial intelligence become more common in the industry. Each of these advanced Industry 4.0 applications requires as a prerequisite the kind of large, centralized, historical machine dataset that a system like what we describe collects and manages.

Managing manufacturing data has become as important to operational excellence as managing the solder paste and electronic components placed on each circuit board.

Introduction

As EMS Providers such as Contract Manufacturers look forward to Industry 4.0, their need for complex data analysis to inform manufacturing decisions takes on even more significance. Managing manufacturing data has become as important to operational excellence as managing the solder paste and electronic components placed on each circuit board.

It is important, though, to distinguish two distinct and very different data uses inside of an EMS factory:

1. **Tactical Data:** This is data that local line and building supervisors need to effectively manage their day to day operations. Its characteristics are that it supports quick, real time decisions and its value diminishes once the decision is made. Examples of tactical data are warning lights on machines when they stop working so that a technician knows to diagnose and clear a fault. The tactical value of the data in this case is that it reduced the amount of time between when a fault occurred and when it was cleared. For that value to be realized, it is not important that any information about the fault necessarily be communicated or stored, just that it existed so that a technician assigned to monitor the machines knows which one to approach and inspect.
2. **Strategic Data:** This is historical data that is expected to be aggregated and analyzed long after a given shift or work order has been completed. Strategic data is used to support larger decisions inside of a factory such as the number of machines and lines that are needed to support customer demand and whether it is worthwhile to invest in new, advanced production machines. It can also be used to detect and resolve issues before they rise to the level where production is impacted, and tactical data is needed to quickly resolve them.

The time has long past when it was possible to run an EMS factory without the benefit of tactical data. Labor efficiency considerations over the years have resulted in a small number of technicians required to manage a larger set of machines, hence the need to notify technicians with specific machine needs has become crucial to run the operations efficiently. This also means that there are a wide variety of commercial systems that provide tactical operational data, since all EMS factories need it.
For example, most if not all, SMT machine vendors provide built-in ways for their machines to indicate faults or even allow servicing to be performed remotely.

Given the ubiquity of tactical data collection systems throughout EMS factories, it is tempting to assume that there are also many strategic data collection systems as well. However, experience shows that this is not typically the case for several factors:

1. Historically the cost of collecting and storing large volumes of data such as what is generated inside EMS factories was prohibitively expensive. This cost led to a parsimonious approach to data collection where only the minimum number of data fields for any given purpose were collected. For example, a Manufacturing Execution System (MES) might collect the pass/fail result of an inspection test but not collect any information on the specific defects detected. The pass/fail was the minimum data needed and hence all that was collected, even though defects and other rich information were present inside the inspection machine and could have been collected at the same time. As explained later, collecting just a few data fields deprives the information of necessary context that allows it to be usefully analyzed retrospectively.

> A Manufacturing Execution System (MES) might collect the pass/fail result of an inspection test but not collect any information on the specific defects detected.

2. Until relatively recently, the data analysis techniques capable of pulling insights from large historical datasets were not commercially available. This led to a situation where data was initially not collected because it was not clear how to analyze it but then once the analysis techniques were developed, there were no large historical datasets on which to apply them.
3. Tactical data often does not contain enough detail to support strategic analysis. Taking the previous example of warning lights indicating which machines currently need inspection, that single binary indicator is enough for tactical use. However, it is not enough to support meaningful historical analysis. For example, it lacks information on what the specific fault was and how it was resolved. You could store the total number of faults per machine and look for patterns, but this can have a high rate of false positives since it does not distinguish between normal issues and exceptional problems. A machine that ran out of consumable parts is not distinguished from one with a broken motor.

This, then, is the situation that many EMS factories find themselves in: they generate large amounts of transient operational data that is used locally to notify technicians and engineers about lines and machines with faults that currently need attention. However, that data is not stored historically or analyzed to find opportunities to prevent the occurrence of so many machine faults in the first place. Furthermore, the specific operational data that is generated often lacks associated details that would be necessary to support historical analysis even if it were saved and aggregated.

For example, consider a concrete use case. Pick-and-Place machines (PNP) pick electronic components at high speed from component feeders and place them onto circuit boards as a key step in the manufacturing process of printed circuit board assemblies. Many PNP machines provide a real time fault alert when the error rate of picking parts from a single feeder exceeds a defined threshold. For example, if three parts in a row from a single feeder are lost then the machine will stop and alert a technician to inspect it.

> The problem was never acute enough to trigger a tactical alert but, nonetheless, contributed a significant source of component attrition that would have been easily solvable had the problematic feeder been identified. This latter problem is what strategic data analysis addresses.

What if, instead of three parts, only two parts in a row are ever lost, but this happens consistently: day after day, month after month, even year after year? The problem was never acute enough to trigger a tactical alert but, nonetheless, contributed a significant source of component attrition that would have been easily solvable had the problematic feeder been identified.

This latter problem is what strategic data analysis addresses. It looks at collected data over long periods of time to identify trends that, ultimately, result in acute tactical problems but could be resolved before that stage if identified sufficiently in advance. However, as noted above, just aggregating the same operational alert signals does not solve the problem because those alerts lack context and details needed to support large scale analysis. In the case of the feeder example above, there is no information on which feeder had the problem, only the machine it was attached to so there is no way to calculate per-feeder error rates over time from just the operational data alone.

What this all means is that there is a need for a new kind of strategic data collection system to complement the operational data collection systems already in place in most, if not all, EMS factories. The characteristics of this new system are:

1. It should collect rich, contextualized data that can be unambiguously interpreted when combined into a large historical dataset. For example, rather than just collecting "feeder fault on machine M", it needs to collect "feeder F had specific error E at time T while placing component C on machine M running process P". This latter dataset allows modern "big data" analysis tools, including machine learning or artificial intelligence algorithms, to identify commonalities between F, E, T, C, M and P to predict whether feeder F has an abnormally high mispick rate that should be investigated or not. It also allows tracking feeder F across multiple machines and components to see if its error rate is elevated uniformly or only when attached to a specific machine or placing a specific component.
2. It should combine data from as many machines, lines and factories as possible into a single place for analysis together. Much of the value of historical analysis comes from comparing different observations against each other to establish baseline and abnormal conditions automatically. The more lines and machines that are observed of a given type, the more examples of behavior that can be analyzed and classified, permitting more powerful conclusions.

It's important to note that just as almost all modern EMS equipment produces operational alerts, it also generates and makes available rich data about its actions. However, the specific method of retrieving that information varies widely from machine vendor to machine vendor even if the data itself is fairly consistent for machines of a given type.

The result of the preceding analysis is that a successful centralized data collection system should look like the diagram in Figure 1. The important characteristic of Figure 1 is that the integration point with each machine's data happens directly with the machine's monitoring system that contains rich, contextualized information, not via its operational alerting system that typically only reports occurrences of problems without rich details.

This paper surveys the primary challenges to implementing a centralized data collection system like what is shown in Figure 1 for aggregating and making use of strategic analytical data. It is based on the experiences of the authors in implementing such a system on hundreds of electronics assembly lines across a network of more than two dozen factories.

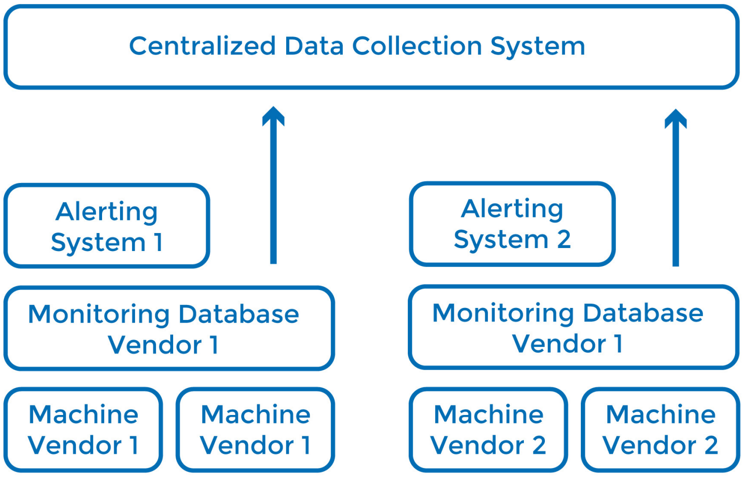

Figure 1 - Centralized data collection system collecting rich machine data directly from machine vendor monitoring databases, not from alerting systems in order to capture contextualized information necessary for aggregation and historical analysis.

Collected Best Practices for Centralized Data Collection

Since centralized strategic data collection systems are less common in EMS factories than tactical alerting systems, it is useful to review the unique challenges that occur in a centralized data collection system and how they differ from local systems and require special solutions.

Impact of System Reliability on Design of Analysis Algorithms

Operational data collection systems are typically small scale. Their structure mirrors the structure of the factory so there is often a separate, disconnected system for each factory line or at most several lines grouped together. Since the goal is to provide alerts to the operators of each line, there is no need for information to cross between lines. This small-scale nature of operational collection systems means their reliability does not need to be exceptionally high for them to function adequately.

For example, if an operational alerting system works 99% of the time, there is little additional value in getting it to work 99.9% of the time. This relaxed reliability need is important because it means such systems are comparatively simple to design and implement.

The constraints on a centralized data collection system are more stringent. Consider a system connected to 100 factory lines, each of which has a 99% chance of reporting data correctly. Assuming data connections to each line fail independently, the joint probability that all 100 lines are successfully reporting data at any given time is just 37% (0.99^100). Thus, the reliability of a centralized data collection system must either be significantly higher than 99% or the analysis performed on its data must be highly tolerant to temporary problems on a small number of lines at any given time. The first option for dealing with reliability is self-explanatory though practically challenging, so let us focus on the second option.

Take, for example, a centralized data collection system designed to calculate a Line Utilization (LU) score for each of 100 factory lines and then report the average LU for all lines as a single global Key Performance Indicator (KPI) for the entire factory. LU, for the purposes of this discussion, is defined as the actual number of component placements of an SMT line in a period of time divided by the installed placement capacity of the pick-and-place equipment on that line.

The direct approach to this problem would be to collect the total placements reported from each line, sum them and then divide by the fixed installed placement capacity of each line to get the LU value. This calculation is simple and clear, however it produces a biased underestimate in the face of unreliable machine data collection. Whenever a connection to a machine or line is lost, there are no placements reported, so it is never the case that losing a connection results in more than the actual number of placements being recorded, only ever less. It is a standard result in this case that the calculated LU will follow a binomial distribution, which means that a 99% reliable connection to each line would produce an LU value just 1% lower than the true value. If the reliability were 90%, then the LU value would be 10% lower than the actual value. It could be that a 1% LU error is tolerable and no special action is required since the LU calculation is relatively stable in the face of temporary connection losses.

However, there are other kinds of analysis that are much more sensitive. Consider a machine fault analysis system that calculates the amount of manufacturing time lost due to each kind of machine fault received. A basic design for such a system could look like the following. A signal is received when a machine fault begins. Another signal is received when the fault is cleared. The system computes the time between these two signals as the duration of the fault and then groups these durations based on the fault code to get the aggregate impact of each fault. Unlike the LU calculation, this simple approach to fault impact ranking is highly sensitive to unreliable machine connections. What if you miss the single fault cleared signal? When do you clear the fault? Without more complex analysis techniques, the results are likely to be highly skewed by a small number of faults that appear to "never be cleared" and are abnormally long.

There are other kinds of analysis that are much more sensitive. Consider a machine fault analysis system that calculates the amount of manufacturing time lost due to each kind of machine fault received.

There are many interesting kinds of analysis that have error tolerance characteristics similar to the fault example above. While there are case-by-case methods to make each more stable, there is also a general principle that can be followed to make these calculations inherently stable: include a qualification step on all potentially unstable analysis routines that does not attempt to analyze all occurrences of a problem but rather only those for which complete information is available. This qualification could use a heartbeat or other signal indicating that data collection was reliable during the entire event or it could be based on some feature of the event data itself that inherently distinguishes real events from data collection errors. By always analyzing only a subset of qualified events and then generalizing, the results can remain robust even in the face of substantial gaps or errors during data collection.

Passively Obtaining Rich Data from Machines

Many SMT machines and their included or optional support software are designed to report data to external systems. However, the most common use of these reporting systems is to enable integration of the SMT machine with Manufacturing Execution Systems (MES). MES systems are considered mission critical tools for running a factory, so it is expected that if they are non-operational, then the factory will shut down. Even more so, many machines that support MES integration will actively refuse to run if configured to report data to an MES system which is not available to receive it. This self-stop feature prevents situations where a product could be built without, for example, collecting necessary traceability data on the parts used to produce it.

Given the difficulty of building a global data collection system with the same reliability of a local factory MES tool, it is preferable to continue running the factory and just not collect centralized data whenever the data collection system is not operational. This means that whenever possible, a passive connection to each machine should be chosen. Passive, in this case, means that the machine chooses to discard data whenever the collection system is not available rather than halt operation. Choosing a passive integration strategy removes the critical failure more where a fault in the centralized data collection system connected to many production lines halts them all simultaneously.

Handling Expected Data Volumes

One common concern with centralized data collection is whether the date rate or overall volume will be too much for any one system to handle. On this point it is important to note that while EMS factories generate a substantial amount of data, it does not generate data on the same scale as many other modern computer systems. For example, data from an SMT machine is typically reported every time a panel passes through a machine. There are typically on the order of ten messages for each panel and five different SMT machines per line. A standard cycle time is on the order of 30 seconds so this means that an SMT line is expected to produce about 10*5*2*60 = 6,000 messages per hour. This means that a large factory with 50 SMT lines could be expected to produce about 300k messages / hour when fully connected. An entire factory network with 20 factories would produce about 6 million messages hour hour. Six million messages per hour is substantial but easily handled by modern message processing systems.

Six million messages per hour is substantial but easily handled by modern message processing systems.

It is also important to note that data from different factory lines are independent of each other so such a data collection system is easily horizontally scalable by assigning different factory lines to different data collection instances that do not need to coordinate with each other.

In terms of overall data storage sizes, it is important to note that modern, cloud-based data management software is designed to handle petabytes of data without issue. The data volumes generated inside of EMS factories, even large ones, do not exceed these capabilities.

Receiving Data in Multiple Formats and Protocols

At the time of writing this article, there is not a single dominant protocol or standard by which SMT machines report performance and error data. There have been past standards that achieved partial adoption and there are newer standards that are currently in the process of achieving adoption. The result is that machines from different vendors and ages will produce data in different formats and make it available via different protocols.

There are two choices for handling this heterogeneity: have the machine standardize data before reporting it or accept data in whatever format the machine can produce and then standardize it as part of the data collection system. Both approaches have merit but practical experience teaches that the second approach is the better one. The primary reason is that machine vendor reporting software is a software product like any other. This means that it both has bugs and requires active support and maintenance by its creators in order to remain functional over long periods of time. By using the most common and supported reporting software available for each machine, one stays within the mainstream support offered by the machine vendor and maximizes the odds that the chosen protocol will remain supported in the future while minimizing the odds that your system will be the first to discover a critical bug in how the reporting software works. In contrast, if your system uses a special or one-off integration to the machine, there is a higher chance that it will uncover latent bugs in the reporting protocol as well as that the ongoing burden of support will eventually exceed the capacity or desire of the machine vendor and they will stop producing updates.

Avoiding Manual Data Entry

Many local operational monitoring systems rely on manual data entry for enriching machine data with contextual information such as external failure reasons or factory conditions relevant to the reported data. The local monitoring systems are able to rely on manual data entry because the users entering the data tend to be the same as the users interacting with the system. For large-scale centralized systems, this is no longer the case and the unreliability of manually entered data can have a much larger impact. Experience has shown that for centralized data collection systems, avoiding any kind of manual data entry step is best whenever possible. Oftentimes this is possible by using rich machine data as a proxy for the information otherwise entered by the user. For example, some operational monitoring systems rely on manual entry of a work order number in order to contextualize the process running on a line when interpreting machine data. While it may not be possible to avoid manual entry if the work order is truly needed, since most machines do not know the work order number, it may be possible to replace the work order with a machine recipe or program identifier that has a similar ability to uniquely identify the running process on the line but does not require manual entry.

Experience has shown that for centralized data collection systems, avoiding any kind of manual data entry step is best whenever possible.

One other important reason to avoid relying on manual data entry is the timescale difference between when data is available from the machine (nearly instantaneously) and when a human is able to manually annotate it (several minutes to hours later and highly variable). If your system requires manual data enrichment as part of processing, that means that most data, when received will be a not-processable state because it is waiting for manual data to be entered before it can be interpreted. In practice this can be challenging to manage. On the other hand if no manual data entry is required then machine data can be processed and stored immediately as soon as it is available. As message rates scale to several million messages per hour, being able to always process machine messages on the fly is an important characteristic.

System Design

The combined best practices and lessons learned presented above leads to an overall system architecture as shown below in Figure 2. The key aspects of the design are:

- Storing historical data in a low cost data lake environment designed for efficient querying of large amounts of data. There are many standard implementations of this kind of system that are available either as open-source software or commercial offerings.
- Allowing machines to report data in the most supported way for each machine vendor to ensure that the data collection protocol has the greatest chance of remaining supported in future machine software versions.

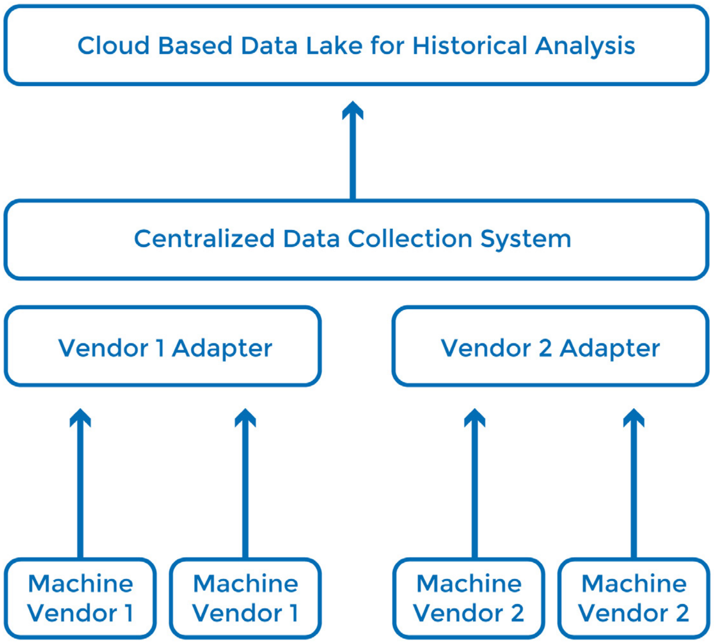

Figure 2 - Block diagram of centralized data collection system showing vendor specific data ingestion systems and cloud based long-term data storage.

Conclusion

We have presented the overall design and lessons learned from implementing a global factory data collection system across a large network of EMS factories. The key benefit of such a system is that it enables large scale strategic analysis of machine data in a cost-effective, centralized location. We see such data collection systems as important components in enabling strategic, Industry 4.0 programs across the EMS industry as they permit the creation and use of large-scale machine data sets necessary for advanced machine learning or AI applications.