An Approach to Highly Available and Extensible Data Management Systems for Large Scale Factory Floor Data

Jaehui Park and Su-young Chi
Knowledge Convergence Service Research Team
Electronics and Telecommunications Research Institute
Daejeon, Korea
email: {jaehui,chisy}@etri.re.kr

***Abstract*—In recent years, Big data prevailed in various domains such as marketing, manufacturing and finance. Although many analytical models and practical systems have been studied, it is not a typical task to adopt them to domain problems. There exists a gap between understanding the problems in the field and adapting the Big data techniques. This work proposes a data-driven framework bridging the gap between existing data management issues and problems on the shop floors in manufacturing industries. In this paper, we propose a highly available and extensible data management system to integrate manufacturing data with regard to four factors: man, machine, material and method. This data management system supports a large scale data in terms of reliability and efficiency for data collectors and analytical systems. Furthermore, as an ongoing study, we have investigated a set of requirements and practical problems from field-workers rather than executive managers. We formulated a comprehensive system structure towards a predictive manufacturing system that can capture, in advance, potential risk factors, such as, machine worn out progress and production time loss tendency.**

Keywords; Big data; data-driven framework; high availibility; extensibility; predictive manufacturing.

I. INTRODUCTION

With the rapid growth of sensors and communications, various data from the real world comes into the digital world with large volume and heterogeneity and at high speed. The words ‘Big data’ have been attracting much interest from various domains such as marketing, manufacturing and finance. Big data concerns many aspects of recent data issues, for example, gathering data from scattered sources, storing data in persistent storages with efficiency and reliability, and analyzing data to bring business insights. Although many analytical models and practical systems in Big data environment have been studied, it is not a typical task to adopt them directly to field problems. There exists a gap between understanding problems in the field and adapting the Big data techniques. For example, industrial enterprises, which are in relatively conservative domains, are seeking business insights within their ERP/SCM information while they lack practices of associating web data, social networks and even with the their shop floor data.

Manufacturing execution systems (MES) manage product life-cycle, resource schedules, order execution and dispatch and so on. MES is known as a production information system for manufacturing decision makers. A MES provides useful insights for the managers in industrial enterprises to make decisions to optimize the production performance. Many MES vendors, for example, SAP or Rockwell Automation, try to adopt recent big data techniques to improve the performance of their solutions. However, we note that utilizing MES capabilities does not always improve manufacturing performance. This is due to the nature that MES is a system for the executive managers, and is not coping with the requirements of field workers. However, field workers concerns on more specific problems on the shop floors, such as, optimal settings of machine equipment. These are a variety of practical problems issued by technical workers on the factory floor, which cannot be solved by existing solutions, such as MES/ERP/SCM. This work aims at proposing a data-driven framework bridging the gap between recent data management issues and actual problems on the shop floors in manufacturing industries.

In this paper, we propose a highly available and extensible data management system for manufacturing data. This paper is a work-in-progress report of the ongoing project “Development of a predictive manufacturing system using data analytics in small and medium enterprises (SME)”. This report mainly describes a system architecture for industrial data management systems and integrated data structure design. Our contributions summarize as follows: 1) a highly available data management server architecture and 2) an extensible manufacturing database model for integrating data with regard to four factors: *man*, *machine*, *material* and *method*. Furthermore, we have investigated a set of requirements and practical problems from workers on the shop floors rather than executive managers. We try to formulate a prototype for predictive manufacturing system that can capture, in advance, potential risk factors, such as, machine worn out progress and production time loss tendency. The system is expected to enhance the existing manufacturing enterprises complementing the vision of smart factories [1], factory of the future [2], industry 4.0 [3]. We plan to build a test-bed factory to evaluate our proposed

 ISBN: 978-1-61208-423-7

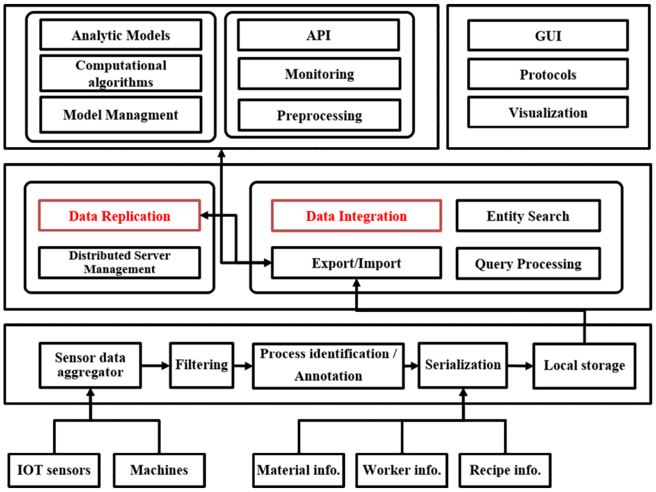

Figure 1. An architecture of a predictive manufacturing system.

system in the domain of manufacturing automobile components, especially, motor shafts. Discussions including the current issues and plans will be presented in the last sections.

II. ARCHITECTURE

To introduce the integrated data model, we first introduce an architecture of *predictive manufacturing systems* (PMS). PMS provide four sub-modules to 1) gather data from different sources and serialize them, 2) integrate the data and replicate them in the distributed server systems, 3) build analytics models learned from the data, and compute future events 4) visualize data. Figure 1 shows the sub-modules. In the bottom level, available data sources are identified, which can be extended. *Data Collector* module aggregates the raw data and annotates/serializes into meaningful inputs. This module can be regarded as supervisory control and data acquisition (SCADA) system. The difference of our system include ad-hoc sensors, so called, internet-of-things (IOT) sensors to acquire the data that cannot be obtained from machine internals. Collectors can be easily extended by simply adding IOT sensors based on the problem solving requirements gathered directly from the field workers. For a trial, we plan to set up a video system to acquire vision data of material form changes in milling machines. *Analytics System* leverages the data gathered from the factory floor. It builds analytical models using machine learning algorithms. Based on the model, it predicts events to provide the workers get an insights to solve the problems. For example, based on the vision data gathered from the sensors, the system learns the patterns of anomalies to machine faults or product defectives. In our test-bed, we try to detect the occurrence of rolled chips, which are residue of milling process, in milling machines. Although the problems are well-known in the fields of milling processes, there is not a systematic solutions. The solution has been relied on experience of the field workers. The benefit of our systems is that the faced-problems raised from the factory shop floors can be solved. Existing solutions, i.e., MES solutions, cannot be suitable to

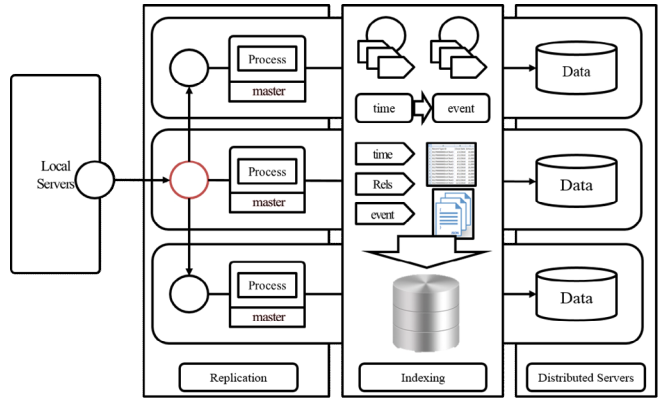

Figure 2. A processing framework.

this kind of field problem solving because they focus on the upper levels of application and the lower levels of data. Hence, we believe our systems complements the existing systems.

In this section, we report the sub-modules for 1) Data Replication and 2) Data Integration (colored in red in Figure 1.) First, we build replicated server system to achieve the availability of aggregated industrial data. Every data element is a type of codes and numbers with temporal information. Therefore, it can be easily maintained in a distributed system. Figure 2 shows the conceptual front-end architecture of data replication module. This module gets input from local servers and replicate them. In our implementation, we use wsrep APIs for synchronous write-set replication. Because industrial data should be process in a transactional manner, we should guarantee that each data nodes in the replication cluster process input in the same order and uninterrupted. Replicated data can be indexed in distributed manner. Distributed server maintains the data using uniform structures we defined the next section. High availability is a distributed server system characteristic to maximize up time to running time. To tolerate down time, the system show eliminate of single points of failure. This means adding redundancy to the system both at data level and at system level. Also, failure detection should be provided. Maintenance activities should be provided. However, server maintenance is out of scope. Manufacturing environment always has scheduled downtime, for example, equipment maintenance. In our implementation, we avoid the single point of failure by introducing multi master replication. This enables high availability with no slave lag and no lost transactions. Data Replication server implementation is illustrated in Figure 3. We construct a physical system using three nodes with Intel Xeon E5 processors, 16G DDR3 memories, and 10TB HDDs. The physical system can be virtualized in clouds. Mater-master replication and distributed database management systems are organized using MariaDB galera clusters. We developed user interface using MariaDB connectors. To communicate with external modules, we define protocols using XML messages and SQL parameters. This implementation is still in progress.

Second, we define a database structure for the Data Integration module. This sub-module identifies the semantics

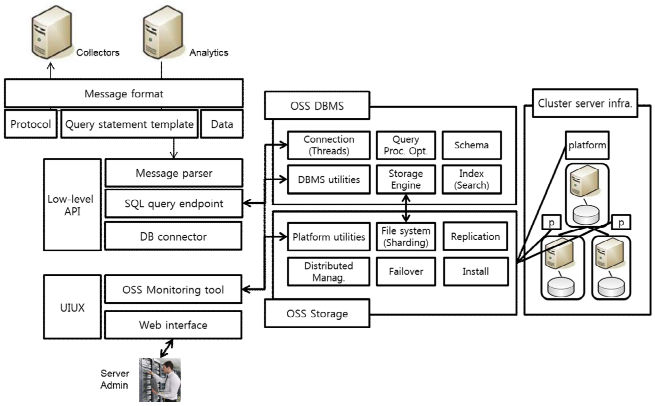

Figure 3. Data Replication server implementation.

from the shop floor data, for example, control codes in the numerical control machines and its continuous changes. The control codes implement the behaviors of the machines to obtain products from computer-aided design programs. Therefore, to obtain a predictive model by understanding the underlying data characteristics, the semantics should be designed in the database structure. Data integration task involves dealing with flexible adaptation of a variety of data. We note manufacturing domains share some common factors to identify shop floor data. We extract to factors in four categories: *man, machine, material and method*. Man denotes the category for personnel information, shift information, performance, and so on. Human factors in the manufacturing process can be utilized. Machine denotes the category for machine internal information, such as machine status, its logs, equipment detail, axis of machining tools, fault/alarm history and so on. Material denotes the category for the specifications of input materials and the intermediate results from previous process. Method denotes the category for machine parameters to yield an end product. We refer to the parameter settings as *recipe*. This information can be used to improve the production performance of the factory if we can produce a gold recipe by optimizing the parameter settings. The theoretical study parts of our project try to optimize the milling machine parameter to reduce initial setting time. To support the study, we provide integrated data structure to gather the features to build effective models. Based on our data model, all of those factors can be visible and utilized in a uniform way.

III. Data Model of Data Integration Module

To understand the data with four factors in shop floor, we first analyze the data sheets, which are created manually. This data is maintained with sloppy identifies and different types of documents. We refine and categorize the semantics to integrate them with man, machine, material and methods. Figure 4 illustrates the conceptual image to categorize the manual data. Identifier parts comprise id, location, time, and man. Based on the identifier information we can break data into several pieces of similar characteristics into a single information unit. First, quality control data contains inspection item, allowance, inspection frequency, defective

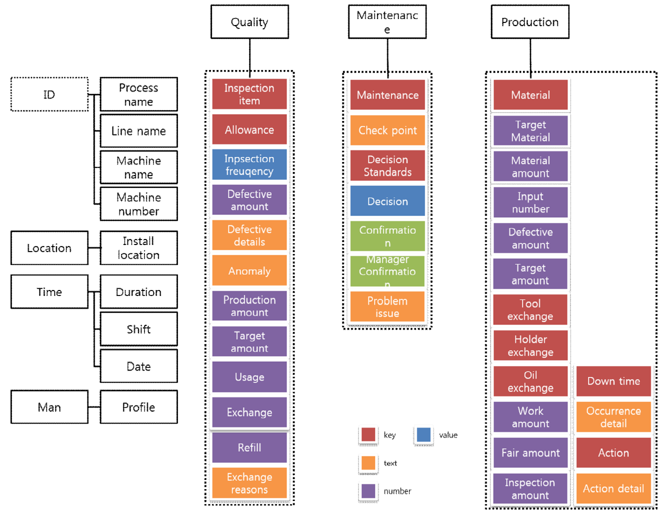

Figure 4. Manual data (production documents) mapping to the integrated database structure.

amount, tool usage/exchange/refills and so on. This information can be utilized to labels for machine learning algorithms in the analytics system. Also, production data contains amount data of material, fair/defective products, up/down time, occurrence/action time. This information explains the status of entire production processes. Maintenance information some temporal data for maintenance event for machine. This can be used for relationships to machine data. Figure 5 illustrate our first draft of integrated data structure. There are two main identifier p_code and m_code, which represents process identifier and machine identifier, respectively. Every event associated to machine, such as inspection, maintenance, machine errors are pre-defined by technical workers. Based on the definition, which is denoted by 'standards', the system records the tuple of 1) items in the standards, 2) temporal information and 3) values. This structure is flexible because every event can be extended by adding standard/log tables to the four factors. Equipment, material, process and others associate above three information. Our data structure can log every transactional and non-transactional information. In the context of server system (presented in Section 2), the vertical row-based data structure can be appropriate to multi-master replication setting. Our data model design cannot be a single solution to maintain the manufacturing data. However, our model is flexible enough to extend to various data sources.

IV. Conclusion

We are currently working on surveying a set of requirements and practical problems from field-workers rather than executive managers. Traditionally, in manufacturing information systems (MIS), data of production, costs, labor, warehousing and so on has been important because entrepreneurs and managers are more interested in it rather than the shop floor environment. As the number of machines and facilities increases and they are automated, field workers can be faced to real world problems

 ISBN: 978-1-61208-423-7

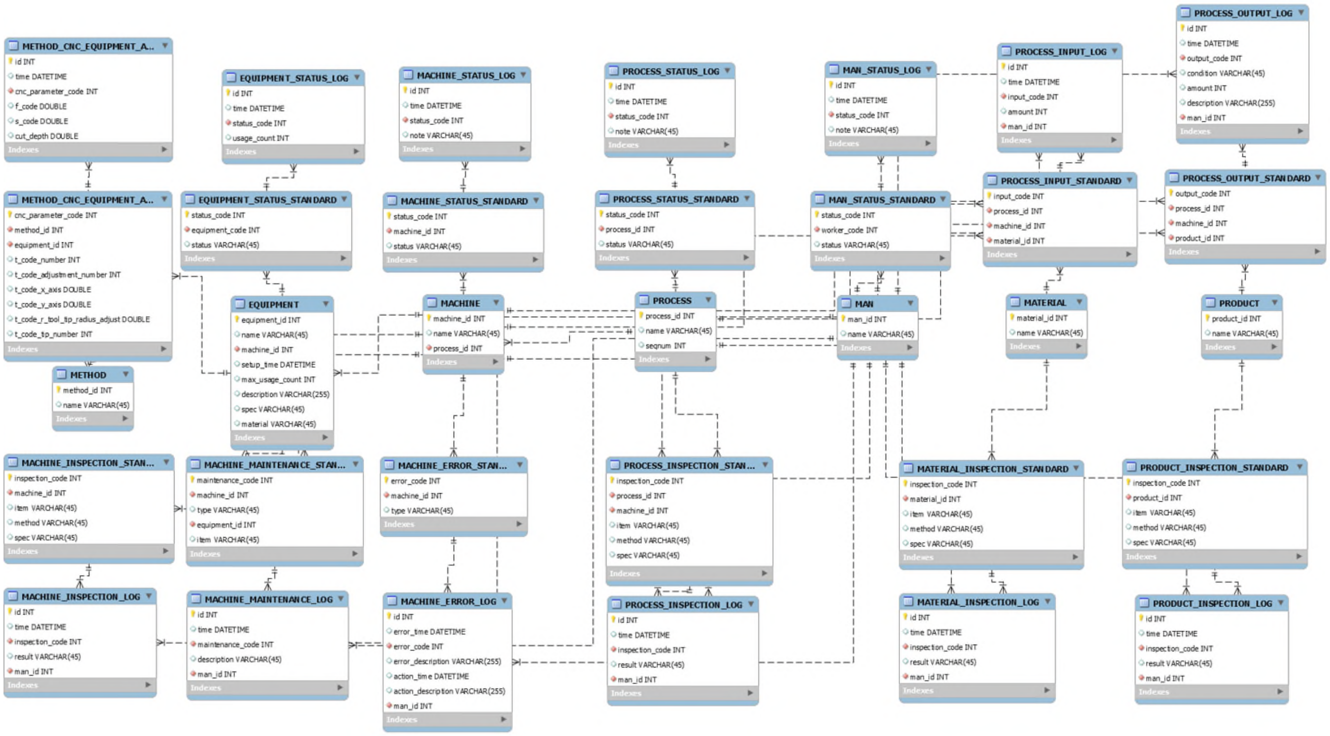

Figure 5. A data model for integrating manufacturing data.

not stretching to the upper levels. Therefore, there should be know-hows and experience that should be maintained by human intervention. However, we are currently formulating a comprehensive system structure towards a predictive manufacturing system that can capture, in advance, potential risk factors, such as, machine worn out progress and production time loss tendency. Our system reduces the human intervention in the process of manufacturing. Also, prediction results may reduce the down time of production to improve performance.

As a further work, we attach streaming data processing part at the input port of the replication modules. Such parts could be useful for computing descriptive statistics in real time manner, which can reduce the burden of analytics system. There are several well-known tools, for example, Apache Storms and Spark. We plan to utilize those tools for processing streaming data.

ACKNOWLEDGMENT

This work was supported by the Knowledge Services Industry Core Technology Development Program (10051028, Development of Predictive Manufacturing System using Data Analysis of 4M Data in Small and Medium Enterprises) funded By the Ministry of Trade, Industry & Energy (MI, Korea).

REFERENCES

[1] F. Shrouf, J. Ordieres, and G. Miragliotta. "Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm," Industrial IEEE International Conference on Engineering and Engineering Management (IEEM), 9-12., Dec. 2014. pp. 697-701

[2] http://www.economist.com/blogs/schumpeter/2013/10/manufacturing [retrieved: May, 2015]

[3] H. Kagermann, W. Wahlster, and J. Helbig. Recommen-dations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group. 2013

 ISBN: 978-1-61208-423-7