

Design and Implementation of a Hybrid Equipment Data Acquisition System(HEDAS) for Equipment Engineering System(EES) Framework

Gyoung-Bae Kim*

EES 프레임워크를 위한 하이브리드 생산설비 데이터 습득 시스템(HEDAS)의 설계 및 구현

김 경 배*

Abstract

In this paper we design and implement a new Hybrid Equipment Data Acquisition System (HEDAS) for data collection of semiconductor and optoelectronic manufacturing equipments in the equipment engineering system(EES) framework. The amount of the data collected from equipments have increased rapidly in equipment engineering system. The proposed HEDAS efficiently handles a large amount of real-time equipment data generated from EES framework. It also can support the real-time ESS applications as well as non real-time ESS applications. For the real-time EES applications, it performs high-speed real-time processing that uses continuous query and filtering techniques based on memory buffers. The HEDAS can optionally store non real-time equipment data using a HEDAS-based database or a traditional DBMS-based database. In particular, The proposed HEDAS offers the compression indexing based on the timestamp of data and query processing technique saving the cost of disks storage against extremely increasing equipment data. The HEDAS is efficient system to collect huge real-time and non real-time equipment data and transmit the collected equipment data to several EES applications in EES framework.

▶ Keyword : Equipment Engineering System, Equipment Data Acquisitions System, e-Manufacturing Execution System

• 제1저자, 교신저자 : 김경배

• 투고일 : 2011. 11. 14, 심사일 : 2012. 12. 04, 게재확정일 : 2011. 12. 27.

* 서원대학교 컴퓨터교육과 부교수 (Dept. of Computer Education, Seowon University)

요약

본 논문은 장비엔지니어링 시스템(EES) 프레임워크에서 반도체와 광전자 제조장비를 위한 새로운 하이브리드 생산설비데이터 습득 시스템을 설계하고 구현한다. 장비엔지니어링 분야에서 장비로부터 수집되는 데이터 량이 급격히 증가하고 있다. 제안된 HEDAS(Hybrid Equipment Data Acquisition System)는 EES 프레임워크에서 발생하는 대용량의 실시간 데이터를 효율적으로 처리한다. 또한, 제안된 시스템은 실시간 EES 응용 뿐만 아니라 비실시간 EES 응용을 지원할 수 있다. 실시간 EES 응용을 위해서 HEDAS는 메모리 기반의 연속질의와 필터링 기술을 이용하여 고속의 실시간 처리를 수행한다. HEADS는 비 실시간 장비 데이터를 HEADS 기반의 데이터베이스 또는 기존의 데이터베이스에 선택적으로 저장할 수 있다. 특히, 급격하게 증가하는 장비 데이터에 대해 디스크 저장 비용을 절감하기 위해 타임스탬프 기반의 압축 인덱싱과 질의처리 기법을 제공한다. HEDAS는 EES 프레임워크에서 대용량의 실시간 및 비 실시간 장비 데이터를 수집하여 다양한 EES 응용에 수집된 데이터를 전송할 수 있는 효율적인 시스템이다.

▶ Keyword : 생산성비공정시스템, 생산설비 데이터습득시스템, e-제조실행시스템

I. Introduction

The development of Information Technology has led to increased global competition and rapidly changing customer requirements in the manufacturing environment. The amount of the data that was able to be collected from equipment increased rapidly in equipment engineering system (EES)[1]. Moreover, the equipment data use applications have been expanded and the data necessity has been diversified. So the efficiency of the applications decreases, the cost and time has become more necessary. Manufacturers must reduce production cost and raise products quality. Especially the investment costs of the semiconductor and TFT-LCD factories are very enormous, so the manufacture system needs to improve the overall equipment effectiveness (OEE)[2, 3]. In view of this, international SEMATECH manufacturing initiative (ISMI)[4] proposed an EES to improve OEE of the whole factory by collecting equipment engineering data to monitor and predict health status of equipment and products quality.

According to the equipment engineering capability (EEC) guideline[1] made by ISMI, equipment data are divided into equipment manufacturing data and equipment engineering data and transmitted via two different ways separately.

The equipment manufacturing data are transmitted to manufacturing execution system(MES) host for equipment control and work in process (WIP). At the same time the equipment engineering data are delivered to EES functions such as monitor, diagnostics, predictive maintenance, etc. These two kinds of data are transmitted through different transmission ports separately to prevent and reduce network burden[5].

In the EES framework environments of semiconductor industry, the equipments continuously generate the equipment acquisition data which is event, status, and trace data[6]. The data gathered from the sensors in the factory equipment must be synchronized so that ordering of information is preserved at remote nodes analyzing the information. Realizing the importance of high-speed data acquisition, the SEMI has developed the equipment data acquisition (EDA) standard to govern the communication of real-time equipment and metrology diagnostics information in the semiconductor factories[7-10]. The equipment data gathered by EDA standard also can be stored in EES DB or customer DB in order to support historical EES applications such as fault detection and classification(FDC).

In EES framework, the insert transactions of EES occur from 50,000 to 500,000 per second in semiconductor industries. Each application may request the equipment data to EES in on-line (real-time) or off-line (historically). The

EES framework needs more an intelligent EDA system that can store and search large volume of equipment data in real-time[10, 11]. The most EES are adapted to a traditional disk-based DBMS such as Oracle[12], IBM DB2[13], and MySQL[14] to handle large amounts of equipment data.

The main memory DBMS that uses main memory instead of slower disk in order to access database quickly has been developed. The main memory DBMS over a disk-based DBMS has greatly improved the speed of transaction processing. Altibase has been announced an EES solution for semiconductor productivity management using a hybrid MMDB[11]. Because of the limitations of the capacity of main memory, it is impossible to reside all EDA data in the memory database.

The equipment data of EES is in form of data stream. It arrives continuously, rapidly, unboundedly, and in real-time from sensors in EES[15]. It is impossible to control the order in which elements arrive and to store the entire data elements. Therefore, the data stream management system(DSMS) has appeared to handle large amounts of equipment stream data. While DSMS can efficiently process the EDA events that occur in EES framework, it is difficult to store the large volume of equipment data into database efficiently in EES framework.

In this paper we design and implement a hybrid equipment data acquisition system(HEDAS) which efficiently handles the equipment data for real-time processing and huge stored data for statical and analysis and statistical processing in EES framework. It supports real-time and non real-time (historical) EES applications. For the real-time EES applications, the HEDAS gives the efficient real-time events processing that handles large amounts of equipment data. It performs high-speed real-time processing that uses continuous query and filtering techniques based on memory buffers.

In particular, the HEDAS offers a compression method saving the cost of disks storage against extremely increasing equipment data. If the real-time requirements of equipment data are removed over time, we can compress and store the non real-time equipment data into the database of HEDAS. The HEDAS also uses the compressed block indexing methods for the efficient query processing of

the compressed equipment data. When the HEDAS compresses the non real-time equipment data, it generates the local index of compressed data based on the data's timestamp. This method improves the searching speed through the partial decompression of compressed data.

The rest of this paper is organized as follows. Section II surveys some related work and techniques. Section III presents the EES framework. Section IV explains the proposed HEDAS of design and implementation in detail. Section V evaluates the performance of the HEDAS. Finally, Section VI concludes this paper.

II. Related Works

Recently, some commercial prognostics system solutions that implement EES framework have become available. MES links plan management and workshop control in an enterprise, which is an integrative management and control system of workshop production oriented to manufacturing process. The traditional MES software causes great development difficulty, long developing period, high cost, low reliability as well as bad reconstructing and integration ability[16]. Chua, Liu, Wang et al.[17] conducted a study on the determination of the size of the lot release system for the backend assembly of semiconductors. The system has been implemented in actual semiconductor plants by integrating it with the upper systems, including MES and ERP. Sandell and Srinivasan[18] did similar simulation study on the lot release polices particularly for the semiconductor fabrication. Using a distributed object-oriented technique, Cheng, Shen, Deng, Nguyen[19] presented a systematic approach to developing a computer-integrated MES framework which is open, modularized, distributed, configurable, interoperable, and maintainable. Their MES framework is designed by the process of constructing an abstract object model based on domain knowledge, partitioning the application domain into components, identifying generic parts among components, defining framework messages, and developing design patterns for generic parts. S.W.Lee and H.K. Lee[20] proposed a data acquisition system which is suitable for the production process of compound semiconductors was

designed and built by analyzing the characteristics of the compound semiconductor fabrication process.

A data stream is a real-time, continuous, ordered sequence of items. Designing an effective data stream management system requires extensive modifications of nearly every part of a traditional database, creating many interesting database problems such as adding time, order, and windowing to data models and query languages, implementing approximate operators, combining push-based and pull-based operators. Aurora[21] is a workflow-oriented system that allows users to build query plans by arranging boxes and arrows. NiagaraCQ[22] is a continuous query system that allows continuous XML-QL queries to be posed over dynamic Web content. STREAM[23] is an all-purpose relation-based system with an emphasis on memory management and approximates query answering. The u-GIS DSMS[24] can be applied to ubiquitous GIS applications. The u-GIS DSMS consists of GES(GeoEdge Server) and GSS(GeoStream Server). The GES coupled with GSNs(GeoSensor Network) collects data in GeoSensor and manages them. The GSS processes data stream inputted from locally distributed GES.

III. Equipment Engineering System Framework

1. EES Framework

E-Manufacturing is advanced manufacturing that takes advantage of Internet and information technologies to efficiently integrate the MES and the EES. Fig. 1 shows the overall architecture of EES that includes three type interfaces and EES applications. MES serves as the intermediary between a business system such as ERP and a manufacturer's plant floor control equipment, helping to manage production scheduling and sequencing, creating an audit trail for track and trace, and delivering work instructions to shop floor workers. The goal of MES is to increase productivity and yield. EES framework provides the gathered equipment data into upper layer applications that require the real-time data or historical data.

2. EDA System

The EDA standards are a collection of SEMI standards for the semiconductor industry to improve and facilitate communication between IC Maker's data gathering software applications and the factory equipment. When implemented together, these standards provide a convenient interface for EDA using SOAP/XML messages over an HTTP or HTTPS connection. There are three kinds of equipment engineering data (event data including detailed equipment event(DEE), analog trace data and dynamic/static context data) that is classified according to the equipment type in EES.

Adding appropriate information such as timestamp to DEE data, it makes DEE data to be possible to completely reproduce what the equipment behave and happen outside of the equipment. Everything from job at upper tier to device's response at the end are DEE data, therefore considering analysis for the future, not only ID and status, but also timestamp will be important. From the timestamp attached DEE data, it is possible to analyze responses for orders and for extended use such as behavior analysis. The trace data is the monitor data used for grasping equipment processing status and equipment capability status. It can be categorized into two kinds of data. One is the trace data for grasping equipment basic functions and equipment capability status, such as vacuum characteristic data, heat transition data, etc. This trace data is always collected from equipment. Another one is the trace data in order to grasp processing related data, such as process pressure, process temperature, etc.

The main EDA SEMI standards include E120, E125, E132, and E134[1]. Solutions must comply with the specific SOAP/XML implementations of these standards; E120.1, E125.1, E132.1, and E134.1.

Most events in the factory occur at frequencies or durations of 10 ms or greater. Therefore a general requirement for time synchronization and time-stamping is about 1 ms accuracy, since it is ideal to have a measurement process that is one order of magnitude better than the desired resolution[1]. However, certain fault events have durations as brief as 1 to 100 microseconds[2].

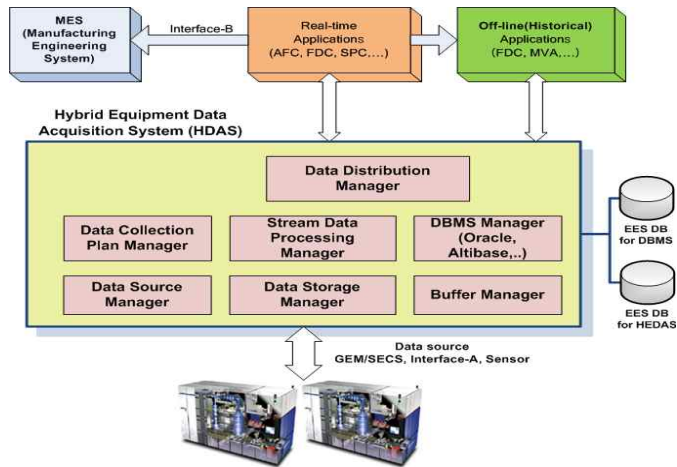


Fig. 1. Architecture of HEDAS

IV. Equipment Engineering System Framework

The HEDAS is composed of seven major managers. In this section, we explain the architecture and features of HEDAS in detail.

1. Architecture of HEDAS

The HEDAS for EES framework can be divided into four parts: application parts, the HEDAS part, EDD databases part, and data sources parts. Fig. 1 shows the architecture of the HEDAS. The HEDAS consists of seven managers: data distribution manager (DDM), data collection plan manager (DCPM), data source manager (DSM), data stream processing manager (DSPM), data storage manager (DSM), buffer manager (BM), and DBMS manager DM. DSM has data schema of input sources and acquires the equipment data from data source according to the Interface A standard. This manager should be controlled by DCPM. A administrator or user can give the rules of data collection from data sources to this manager. According given rules, DSM acquires the data from the data source. DSPM handles the stream data that is provided continuously from DSM in

real-time. This manager provides real-time stream data that is defined by administrator to the real-time applications through the DDM. The HEDAS especially has their own database in order to store compressed form equipment data for saving the disk storage space. If programmers want to use the traditional DBMS for off-line applications, they should store the equipment data into the database by DBMS manager. So, administrators or programmers can optionally choose to store the equipment data in two kinds of database. Application part is divided into three types: MES applications, real-time applications, and off-line applications.

2. Data collection plans for equipment data acquisition

Recently, many intelligent applications of semiconductor factories require variety and complex data types from EES framework. The HEDAS should provide the following features for EDA system: data gathering from various data resources, insertion timestamp into a equipment data, data conversion, data filtering, and data conditions.

The DCPM can gather the data from various data resources such as sensors, GEM/SECS, and etc. by defining equipment data schema. According to SIMI standard, the

equipment data have to include the data timestamp and convert to the XML format. An advanced application may require an original equipment data as well as mixed or combined equipment data according to data collection rules. Also they want to give the conditions of data gathering and filtering to process events and traces. For those purpose, the DCPM of HEDAS is cooperated with the DSM to acquire various types of equipment data in intelligent control methods. According to data sources, it provides diverse protocols because there are a lot of different ways to get data. Data integration is totally different from the structure and type of data taken. As a result, it has to change to the proper data storage architecture and type to be processed in HEDAS. Filtering eliminates unnecessary data in real-time applications by using pre-registered specific conditions.

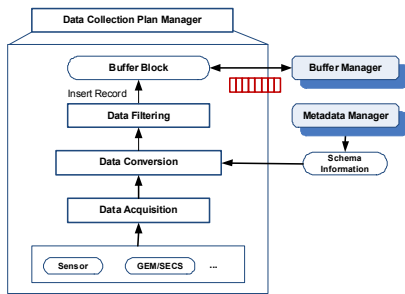


Fig. 2. Processing steps of data collection plan manager

Fig. 2 illustrates the components of DCPM and the process to get data. Data acquisition module gets data, serving various connection methods such as TCP/IP, and JDBC. Data conversion module converts obtained data into compatible data type and structure with predefined schema information and appends timestamp. A large amount of equipment data will be filtered by DCPM according to defined conditions. The filtered and converted data in buffer blocks will be saved on the disk in forms of compressed data by BM.

3. Data compression method using compressed indexing

In order to save disks storage, the HEDAS gives the data compression method using the compressed indexing. The HEDAS temporarily stores large-capacity equipment data in the memory. The HEDAS provides the data compression method to save very expensive disk spaces. While real-time equipment data for real-time applications is stored in the memory buffer blocks by DSPM, non real-time data for historical applications is compressed and stored on a disk by BM and DSM. BM manages blocks of memory storage spaces to save and compress them with high speed. It organizes a lot of buffer blocks in the buffer pool, and one of buffer blocks is allocated in one of data sources connected with.

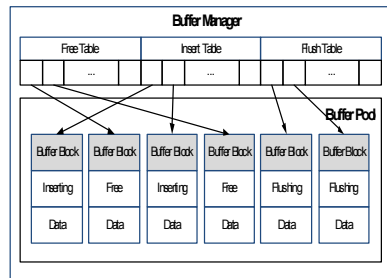


Fig. 3. Data structure of buffer manager

Fig. 3 illustrates the data structure of BM. It has three tables (free table, insert table, and flush table) and provides many buffer block in advance. To gain equipment data, the buffer pool changes a buffer block in random free status to inserting status, and serves when it connects to the data source. This buffer block loads gained equipment data, if the space for storage is exceeded. It changes the status to flushing, and flushes it on the disk.

The compressed buffer blocks are stored on disk using compressed indexing methods by SM. Fig. 4 shows the storage structure of SM. It is formed with a header, an index part, and a data part. The header has the information about the schemata saved in files and the most recent and oldest time for creation. The index part creates a time index for each buffer block. The index node pointing to one buffer block has the minimum and maximum timestamp of

equipment data for the key. The data part manages compressed data set for each buffer block.

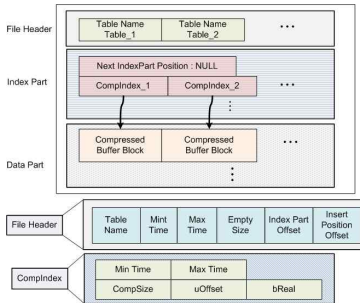


Fig. 4. Storage structure of storage manager

4. Query processing method using data stream processing

The HEDAS extends the SQL query in order to support the stream query in EES framework.

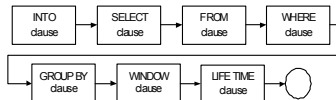


Fig. 5. SQL Syntax of continuous query language

Fig. 5 shows the extended SQL syntax for continuous query language(CQL). The extended SQL syntax has INTO-clause, WINDOW-clause, and LIFETIME-clause. INTO-clause is similar to FROM-clause of SQL. A programmer gives the name of data sources or other resources such as device name of semiconductor factories. INDOWS-clause specifies the scope and interval of CQL. LIFETIME-clause describes the execution lifetime of the CQL and query. An example of CQL is as follows.

CQL Example)

```

INTO      OutSource
SELECT   f1, AVG(f2)
FROM     InSource
WHERE    f1 =40
WINDOW [RAGNE 100 SLIDE 50]
    
```

GROUP BY f2
LIFETIME 3600;

The HEDAS is able to handle both a real-time query and a historical query. There are three types query in HEDAS: real-time query, non real-time query, and mixed query.

The real-time query is generated by real-time applications such as the processing events. It quickly accesses the data stored in memory buffer block. The stream processing technique is adapted to continuously generated equipment data for the on-line real-time query. The non real-time query that is generated by historical applications accesses the data stored in disk database of EES. The mixed query contains both of them.

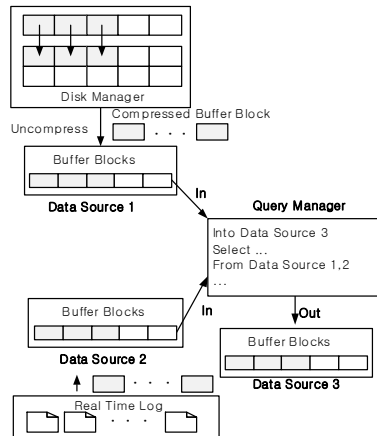


Fig. 6. Query processing steps for the mixed query

Fig. 6 illustrates the query processing step of the mixed query. In this example, the DSPM should execute the CQL query to both the real-time data in memory and the compressed data in disk database. The results of the query should be stored into the output data source. Data source 1 is generated into buffer blocks from the compressed disk database according to the given conditions of the query. Data source 2 is generated into buffer blocks from the equipment data in memory. After making Data source 1 and

data. For real-time data query, the combined system that can be used in existing data acquisition system is similar to the HEDAS. For historical data query, however, the performance of the combined system is rapidly degraded because of using ORACLE and MySQL.

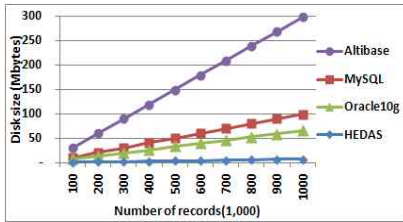


Fig. 9 The Disk size of DB using compression method and the traditional DBMS

Fig. 9 compares the disk size of the HEADS database using compression method and the traditional DBMS database using normal one when the system is stored data in database for historical equipment data. The performance of HEDAS is excellent compared with other systems because the HEDAS uses efficient compression algorithm for text type of equipment data. Therefore, the HEDAS can reduce the size of database dramatically.

VI. Conclusion

In this paper we designed and implemented a new hybrid equipment data acquisition system which efficiently handle a large amount of real-time equipment data in EES framework. The HEDAS combines data stream management system with database management system in order to support real-time and historical EES applications. For the real-time EES applications, the DSMS of HEDAS gives the efficient real-time events processing that handles large amounts of equipment data. For historical EES applications, the HEDAS provides HEDAS-based database and a traditional DBMS-based database optionally to store non real-time equipment data. The HEDAS offers a compression method saving the cost of disks storage against extremely increasing equipment data. HEDAS also uses the

compressed block indexing methods for the query processing of the compressed equipment data. It improves searching speed through the partial decompression of compressed data.

The HEDAS plays a role of high performance real-time equipment data collection system in EES framework and transmits the equipment data to several EES applications. The proposed system can store very large data generated in semiconductor system and reduce storage cost.

참고문헌

- [1] International SEMATECH, Equipment Engineering Capabilities (EEC) Guidelines, Version 2.5, 2002.
- [2] Rockwell Automation, "Industry White Paper - Making Sense of e-Manufacturing: A Roadmap for Manufacturers," 2000.
<http://www.rockwellautomation.com/>
- [3] ISMI, International SEMATECH Manufacturing Initiative.
<http://www.sematech.org/>
- [4] SEMI, Semiconductor Equipment and Material International.
<http://www.semi.org/>
- [5] Yu-Chuan Su, "Embedded System Framework Design for Data Collection and Analysis in the Semiconductor and Optoelectronic Industries," ICIEA 2009, pp.76-81, 2009.
- [6] Bistel, EES Framework.pdf,
<http://www.bistel-inc.com>
- [7] Kalappa, N., Moyn e, J., Parrott, J. and Li, Y. "Practical Aspects Impacting Time Synchronization Data Quality in Semiconductor Manufacturing," Proceedings of AEC/APC Symposium, October 2006.
- [8] Parker, J., Reath, M., Krauss, A.F., Campbell, W.J. "Monitoring and Preventing Arc-Induced Wafer Damage in 300nm Manufacturing," 2004 International Conference on Integrated Circuit Design and Technology, p.131-134, 2004.
- [9] Yasutsugu Usami, Isao Kawata, Hideyuki Yamamoto, Hiroyoshi Mori, and Motoya Taniguchi, "e-Manufacturing System for Next-generation Semiconductor Production," HITACHI Review