Journal of Engineering and Applied Sciences 12 (10): 2662-2666, 2017
ISSN: 1816-949X

An Autonomic Computing based on Big Data Platform for High-Reliable Smart Factory

[1]Jeong-Joon Kim, [2]Deuk-Woo Lee, [1]Dong-Beom Ko, [1]Sung-Il Jeong and [1]Jeong-Min Park
[1]Department of Computer Engineering, Korea Polytechnic University,
237 Sangideahak-ro, Gyeonggi-do, Siheung-si, South Korea
[2]Ubist, 153-719 Seoul, Geumcheongui, Korea

Abstract: This study will design a big data platform for the high reliability smart factory. With the advent of the fourth industrial revolution, the interest and demand for autonomic control systems are increasing. Autonomic control system in order to establish an accurate diagnosis and error response for the target system, the machine requires a machine learning through big data. But now big data platform cannot guarantee the reliability of the data because there is not enough data verification. In addition, existing facilities need a uniform standard because the format for data collection differs depending on the manufacturer. Therefore, this study designs a reliable big data platform for autonomic control engine using manufacturing facility data standard. Through this, pattern recognition of autonomic control engine using reliable data, machine learning, accurate diagnosis of system and countermeasure strategy against error can be established.

Key words: Big data, trust system, autonomic computing, analytics, system architecture

INTRODUCTION

With the advent of the fourth industrial revolution in recent years, interest and demand for intelligent systems such as industry 4.0 and smart factory are increasing. Therefore, autonomic control systems for productivity improvement and cost reduction are becoming important issues. At present, autonomic control systems have the advantage of abstraction of massive system objectives, easy modeling and strategies for errors. However, a big data-based machine learning process is needed to establish a correct system diagnosis and a response strategy for errors. Many big data open source based on Apache's Hadoop have recently been developed. As a result, companies have used big data platform which combines a variety of software such as storage, search, analysis and visualization in a large parallel processing platform based on Hadoop which enables them to use social network services such as Facebook, Twitter was developed (Shin, 2012). However, there is a disadvantage that the big data platform cannot guarantee reliability because there is insufficient data verification. In this study, we design a big data platform to provide reliable data to autonomic control engine of smart factory. By providing this data, reliable data is provided to the autonomic control engine to enable accurate diagnosis through machine learning, pattern recognition and countermeasure strategy against errors.

MATERIALS AND METHODS

Background: This study explains the need for a big data platform in an autonomic control system, introduces the advantages and disadvantages of the existing big data platform and the communication middleware used in the smart factory.

Autonomic control system (Ko *et al.*, 2016): With the advent of the 4th industrial revolution projects such as industry 4.0 and smart-factory are underway and autonomic control systems are becoming an important issue to improve productivity and reduce costs of intelligent systems. The autonomic control system basically has four stages: target system monitoring, analysis, planning and execution. Figure 1 shows an autonomic control system architecture.

The goal-achievement evaluator calculates the achievement of the entire system based on the goals of the system. 'Error analysis and inference engine' analyzes the present errors and deduce the cause based on the data received from the 'Goal-achievement

Corresponding Author: Jeong-Joon Kim, Department of Computer Engineering, Korea Polytechnic University, 237 Sangideahak-Ro, Siheung-si, Gyeonggi-do, South Korea

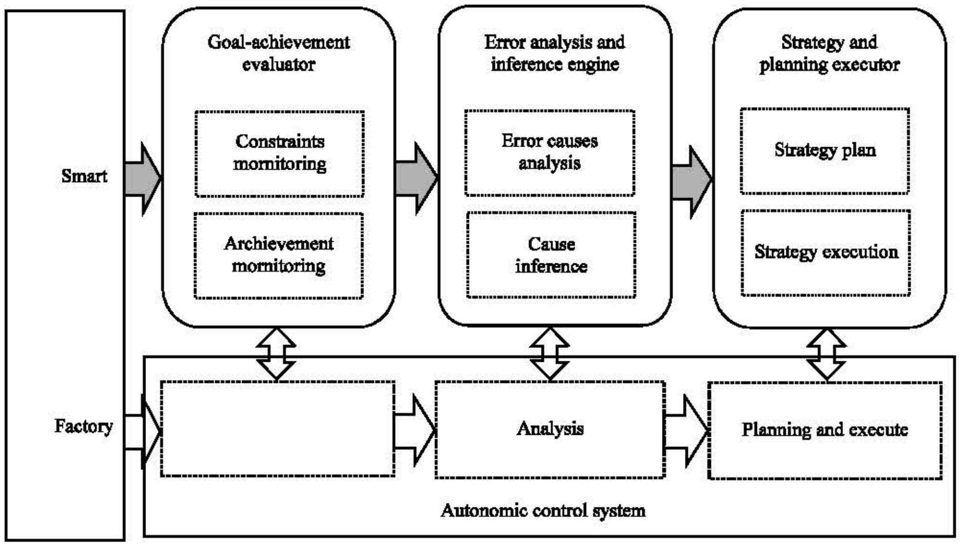

Fig. 1: Autonomic control system architecture

evaluator' when the overall system performance is below the standard. 'Strategy and planning executor' implements the problem-solving strategy established by the 'Goal-achievement evaluator'.

The autonomic control engine has the advantage of being able to easily model and monitor goals by modeling a large and complex system with overall and detailed goals. However, in order to establish an accurate system diagnosis and problem solving strategy, it is necessary to learn machine learning and pattern recognition. Therefore, reliable data for machine learning and pattern recognition are needed.

Hadoop ECO system: Hadoop is a software framework that allows the Apache Foundation's projects to process large amounts of data. The Hadoop framework of the Apache Foundation and the Hadoop ECO system (Landset and Khoshgoftaar, 2015) can be built through Hadoop subprojects that assist in collection, storage, processing, analysis and visualization. Figure 2 shows the structure of the Hadoop ECO-system. ZooKeeper is a distributed administrator that ensures load balancing and guartees availabled and Oozie is a workflow manager that manages the work in Hadoop.

HBase is a column based database based on HDFS, Pig and Hive are similar to SQL but control data stored in HDFS. Mahout is an analytical tool that provides a variety of analytic functions such as classification, referral, filtering, clustering, regression analysis and pattern mining. HCatalog is a data manager that allows data generated by specific modules to be used in other modules. Avro supports remote procedure call and data

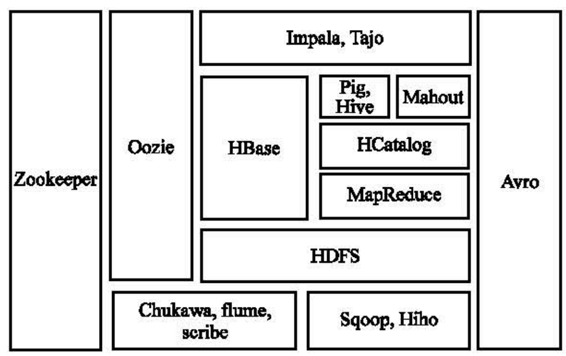

Fig. 2: Hadoop ECO system

serialization and Chukwa and flume allow stable storage of generated data in HDFS. Scribe is Unlike Chukwa and Flume, Scribe transfers data to the main server not HDFS and can be stored in HDFS using the Java Native Interface (JNI). Sqoop is responsible for transferring data to various repositories such as HDFS, RDBMS and DW, while Impala and Tajo are responsible for retrieving and storing data through SQL such as RDBMS.

Hadoop has the advantage of low cost and fast data processing. However, there is a disadvantage that reliability cannot be guaranteed because there is insufficient data verification. Therefore, a reliable big data platform is needed.

MTConnect: With the evolution to the computer environment, it became possible to interwork with various manufacturing facilities in the manufacturing process. However, these services are specialized in accordance

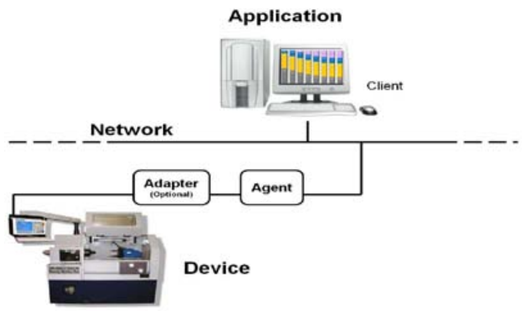

Fig. 3: MTConnect architecture

with the characteristics of the manufacturer and there is a problem in using a manufacturer-dependent protocol. To solve these problems, standardized communication middleware technologies that can be applied to various industries have been developed. The main core technology of communication middleware is data-driven high-reliability real-time data distribution technology. That is a standardized industrial communication protocol must be supported for the facilities made by various manufacturers to be controlled/managed in an integrated environment. This study introduces MTConnect, one of the standardized communication protocols.

MTConnect is an extensible lightweight protocol developed for data exchange between manufacturing facilities and applications. MTConnect is primarily used for monitoring and data analysis in industrial network environments.

MTConnect provides manufacturing facility data in XML format (based on HTTP protocol). Interoperability is greatly enhanced by exchanging data of different kinds. In addition, using the RESTful interface approach has increased the scalability and versatility of the interaction. Figure 3 shows MTConnect Architecture. MTConnect consists of ‘Adapter’ which changes raw data to MTConnect standard format, ‘Agent’ which collects, stores and arranges data and processes client’s request.

In addition to MTConnect, communication middleware and protocols such as DDS (Data Distribution Service) (OMG, 1997), OPC (OLE Process Control) (OPCF, 2017) and OPC-UA (Kim and Bahn, 2016; AI, 2007) are being developed to exchange data between these devices.

Interaction with heterogeneous devices and systems, however, requires a higher level of interoperability as well as simple data exchange/conversion. In accordance with this demand, MTConnect is a communication/analysis based industrial communication protocol standard using ‘Agent/API’. Therefore, in this study we use MTConnect standard to collect the data of the device.

In related studies, autonomic control engine requires reliable data and existing big data platform does not provide reliable data. In this study, we design a reliable big data platform for autonomic computing and collect data of the facility using MTConnect standard.

RESULTS AND DISCUSSION

System design: This study introduces the big data platform for smart factories and explains each module.

System architecture: The big data platform proposed in this study consists of the following system architecture as shown in Fig. 4. As shown in Fig. 4, the big data platform provides data collection, refinement, analysis and visualization procedures for reliable data requiring the autonomic control engine.

Data collection for the big data platform is done through the MTConnect standard. The MTConnect module consists of a data receiver for reliable data collection, a data loader for storing and managing the received data and a data provider for providing the requested data on demand.

Data receiver: The data receiver consists of a trust filter and a data adapter. The trust filter is used to refine real-time data transmitted from the facility to the data required by the autonomic control engine. The data adapter is responsible for converting the data transmitted in a format that does not conform to the MTConnect standard in the facility to the MTConnect standard.

Data loader: Data loader is a module for managing facility. Data loader consists of queue manager and Data Manager. The data transmitted from the facility is stored in a queue type storage. The queue manager is responsible for sizing and initializing the repository. Data manager stores the transferred data in the repository and provides the data when requested.

Data provider: The data provider is a module for providing stored data to the outside. The data request to MTConnect is made by the HTTP get method and the data to be returned is converted into XML and provided. The data provider consists of an HTTP parser that parses requests made in the HTTP get method a data searcher that retrieves data from the repository in response to parsed requests and an XML parser to export the data in XML format (Fig. 5).

Data refiner: The data refiner consists of a data preprocessing module a semantic link module and a data indexer module (Fig. 6). The data preprocessing module acts as morpheme management, sentence separation and semantic generation.

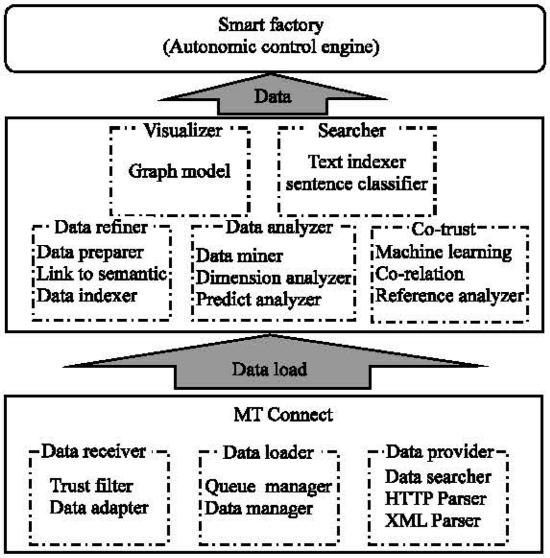

Fig. 4: Big data platform system architecture for smart factory

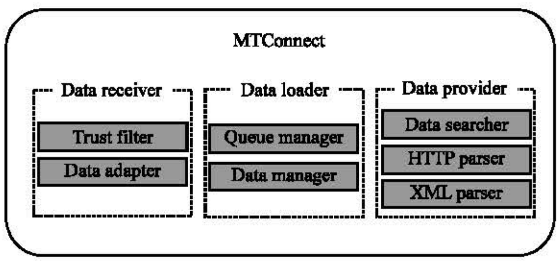

Fig. 5: MTConncet module

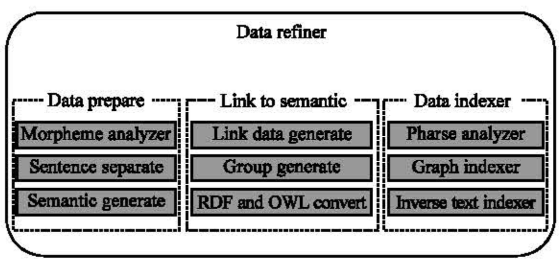

Fig. 6: Data refiner

The semantic link module is responsible for connection creation and group creation of semantic data, data conversion of RDF and OWL format for semantic data connection. The data indexer module constructs the generated semantic data into a graphical index and a reverse text index to help the user quickly find the data. Data analysis and processing consists of data refiner, data analyzer, co-truster, visualizer and searcher.

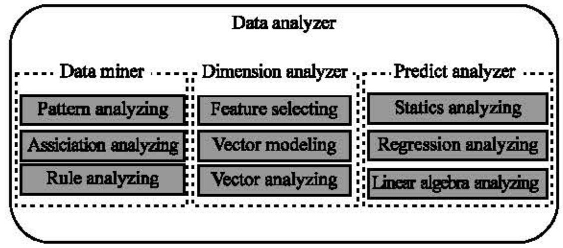

Fig. 7: Data analyzer

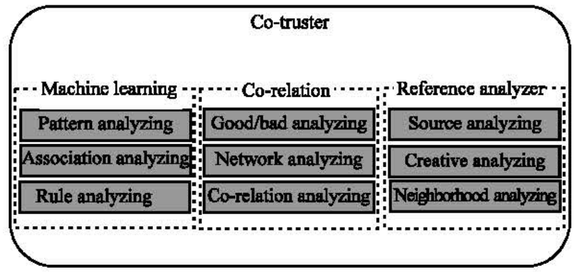

Fig. 8: Co-truster

Data analyzer: Data analyzer consists of data miner module, relationship analysis module and prediction analysis module (Fig. 7). The data miner module performs pattern analysis, connection analysis and rule analysis. The dimension analysis module defines the attributes to be expressed in the dimension, performs the expression and analysis of the vector and the prediction analysis module performs the future prediction through molding analysis, regression analysis and linear algebra analysis.

Co-truster: The co-truster consists of a machine learning module, a mutual analysis module and a reference analysis module (Fig. 8). The machine learning module analyzes the patterns and rules of an entity by performing two functions of generalization and expression. The interrelationship module analyzes the relationship between affirmative and negative analysis of objects and the relationship between objects and the reference analysis module analyzes the relationship between the creator of the object and the neighbors of the researcher.

Visualizer and searcher: The visualizer supports visualization using a graph model and the searcher supports efficient searching through the text index module and the sentence classification module (Fig. 9).

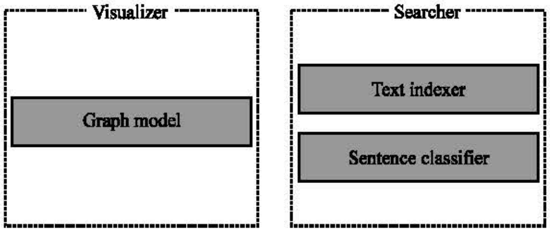

Fig. 9: Visualizer and searcher

CONCLUSION

With the advent of the fourth industrial revolution in recent years, interest and demand for intelligent systems such as industry 4.0 and smart factory are increasing. Therefore, autonomic control systems for productivity improvement and cost reduction are becoming important issues. At present, autonomic control systems have the advantage of abstraction of massive system objectives, easy modeling and strategies for errors.

In this study, we propose a big data platform for high reliability smart factory. Through the proposal we have been able to design a big data platform that provides the reliable data required by the autonomic control engine. However, since the platform is not implemented, it is necessary to verify the performance and reproducibility in the future. Therefore, future research will cover a system that can implement the platform through investigation and verification of open source and can perform machine learning and real-time control of autonomic control engine based on reliable data.

ACKNOWLEDGEMENTS

This research was supported by the ICT and R&D program of MSIP/IITP. [R-20150505-000691, IoT-based CPS platform technology for the integration of virtual-real manufacturing facility.

REFERENCES

AI., 2007. OPC unified architecture. Advosol Inc, Austin, Texas. http://www.opcua.us/

Kim, J. and H. Bahn, 2016. An efficient log data processing architecture for internet cloud environments. J. Inst. Internet Broadcast. Commun., 8: 79-87.

Ko, D., T. Kim, J. Park, S. Kang and I. Chun, 2016. An approach to applying goal model and fault tree for autonomic control. Contemp. Eng. Sci., 9: 853-862.

Landset, S. and T. Khoshgoftaar, `2015. A survey of open source tools for machine learning with big data in the hadoop ecosystem. J. Big Data, 2: 1-36.

OMG., 1997. Traditional messaging solutions lose to data centricity, discovery, QOS and more. Object Management Group Inc, Needham, Massachusetts. http://portals.omg.org/dds/features-benefits/

OPCF., 2017. Introperability Iot and industrie 4.0: From sensor to cloud. OPC Foundation, Scottsdale, Arizona. https://opcfoundation.org /event-details/?ee=76

Shin, S.J., 2012. SNS using big data utilization research. J. Inst. Internet Broadcast. Commun., 12: 267-272.