Cloud-Based Master Data Platform for Smart Manufacturing Process

Lei Ren[1,2(✉)], Ziqiao Zhang[1], Chun Zhao[3], and Guojun Zhang[1]

[1] School of Automation Science and Electrical Engineering, Beihang University, Beijing, China
renlei@buaa.edu.cn
[2] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, China
[3] School of Computer, Beijing Information Science and Technology University, Beijing, China

Abstract. With the development of technology and application of industrial internet of things, a large amount of data is generated in the research and development (R&D) processes in manufacturing domain, including manufacturing procedures, enterprise management, and product transactions. However, these data usually maintained in different departments, which result in information isolation and data with relations cannot be synchronized. This issue leads to the waste of storage space for redundant data and human resources for coordinating essential information. Aiming these problems, we proposed a cloud-based data management platform architecture to collect and maintain the data from isolated domains and distributed departments. A graph database is employed to store the data emphasizing the relations between entities and Master Data Management is deployed to link the entities cross standalone databases. The efficiency of inspecting, managing and updating information across databases shall be improved by the features of the proposed platform.

Keywords: Cloud manufacturing · Master data management · Graph database

1 Introduction

The appliance of smart manufacturing is rapidly increasing with the development of cyber technology. Concepts such as Industrial Internet of Things, Industry 4.0, Cloud Manufacturing were proposed in the progress of manufacturing industry cyberization. The Cloud manufacturing stressed that resources and capabilities in manufacturing should be transformed into services, which means sharing and communicating between enterprises [1]. Domain ontologies in distributed department should be accumulated to knowledge which is a key role in supporting the cloud manufacturing system. Scheduling and optimizing the production procedure requires the directions of the knowledge which indicates the detail of services in the system.

While cloud platforms were establishing widely in the industry to manage the process of productions and the supply chains, they collect and generate an enormous amount of data then analyze them to optimize the work process separately. Isolations between

X. Zhang et al. (Eds.): CloudComp 2019/SmartGift 2019, LNICST 322, pp. 163–170, 2020.
https://doi.org/10.1007/978-3-030-48513-9_13

systems were widely found within and across enterprises. Therefore, the data sharing and synchronization is a complicated and repeated task. The Master Data Management was proposed to effectively manage the data in the distributed systems. The MDM (Master Data Management) extracts information from existing systems and timely processes them automatically [2]. Data service is provided that integrating information from the standalone databases across the whole enterprise and knowledge can be mined utilizing this service. Furthermore, master data can be expressed by ontologies, and the linkage between master data from different enterprises is exploited by analyzing the extracted ontologies. The manufacturing process can be optimized by the guidance from every segment data in the supply chain.

The graph database is developing and gaining attentions rapidly with the growing scale and complexity of big data. A graph database stores data in the form of nodes, edges, and properties using the graph structure which represents the relations between the data [3]. Therefore, a graph database is suitable for maintaining strongly related data and have major advantages in searching information according to relationships between nodes. In the platform we proposed, a graph database is used in storing descriptions and references of data, for instance, the header of tables, the name of tables and the known relations between tables. The establishment procedure and the evaluation results of the master data model often require efforts of experts. With the data descriptions visualized by the function of the graph database, the master data shall be intuitively operated and evaluated. The items in graph database with relations and references can created a dataset for algorithms to establishing and evaluating the master data automatically.

The platform we proposed integrates a graph database and master data management features. The demanding functions, for example, demonstrating the relations of items, coordinating the changes across systems and removing redundant information is also provided.

2 Related Work

The solutions we presented is designed to resolve the data management issues in manufacturing process. The key components and technologies involved in our solutions were continuously studied and developed in recent years.

2.1 Cloud Manufacturing

Cloud manufacturing paradigm utilized technologies to provide manufacturing services with the support of knowledge containing the information to integrate the whole life cycle of products [1]. The knowledge base and data management function are essential parts of cloud manufacturing systems accordingly. Domain experts were often acquired for knowledge editing and knowledge could be generated by data mining as well [4]. A knowledge cloud architecture was presented [5] using big data collected in the manufacturing process to provide support for enterprise manufacturing cloud.

2.2 Master Data Management

The researches on MDM is growing for the actual need created by the developing of big data in varies of fields. A knowledge-based machine scheduling is proposed considering the uncertainties in master data [6]. This approach used a knowledge-based system to provide previously actual data and scheduling parameters is provisioned accordingly. A master data exchange service architecture is presented [7]. An API to manage the messages of master data and algorithms to measure the data quality is also presented in this paper. A data cleaning system which can identifies incorrect data and generates possible repairs were proposed [8]. This system utilizes the knowledge-base and crowdsourcing to identify the data.

2.3 Graph Database

Graph database is tented to be more efficient when storing and processing data with complex relationships. Considering ontology as a large graph, a graph database-oriented approach is proposed [9]. The storing and querying operations on ontologies is claimed to be more efficient and scalable. Data generated by wireless multimedia sensor network can also be stored on the graph database [10]. The experiment showed graph database performs better than relational database in querying the big data related to internet of things. A common information model-oriented graph database was also proposed [11]. Graph database was outperformed relational database again in the multiple test case.

2.4 Summary

These studies indicated that graph database is an effective approach for managing big data with complex relationships, and MDM provides valid method to manage related data in different systems. The approach we proposed utilized these two technologies and aim to provide support in data and knowledge management process in the cloud manufacturing system.

3 Architecture

The functions of cloud-based master data platform are collecting and managing data from distributed database. The data in varies of forms collected were preprocessed and stored in graph database establishing the relations according to the origin data relations. Master data of distinguished departments is generated by the descriptions in the graph database. The master data of different source can be described in ontology form, which enables further analyses and integration. The integrated master data represent the knowledge in this field which included multiple distributed enterprises. Relations between entities in this procedure can be created in the graph database accordingly. The data management can be performed across the isolated databases and coordination of data modification will be an automated and immediate operation.

To realize the functions in the platform, the architecture is depicted in Fig. 1. The cloud-based master data platform can be divided into 4 layers as the data is processed

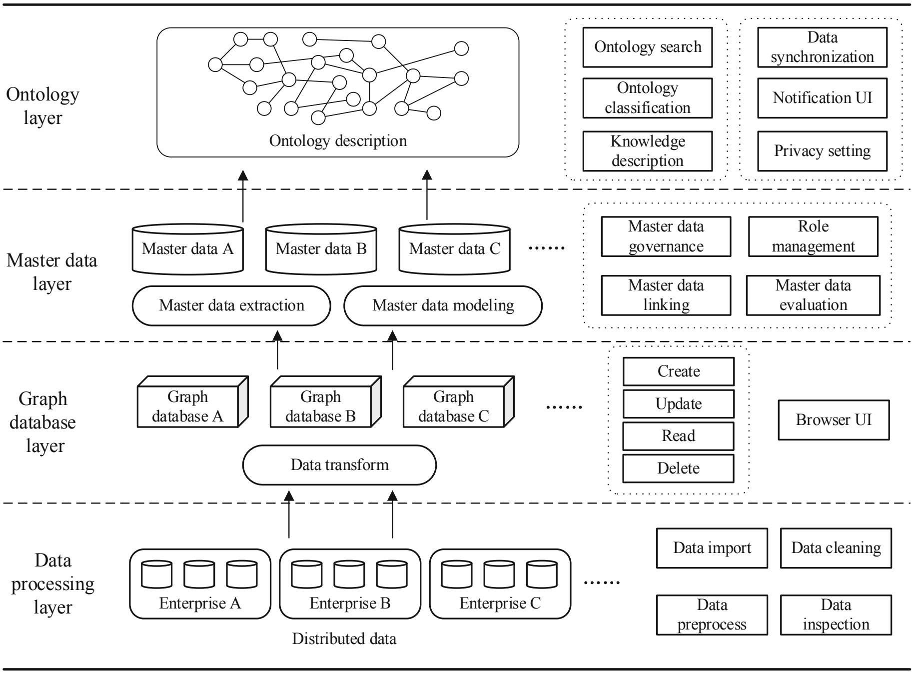

Fig. 1. Architecture of master data platform

and transformed, which are data processing, graph database, master data and ontology description.

The data processing layer collects and preprocesses the data in the disturbed source of original data. The information is described by different form and standard in the systems. A uniformed description of information is generated in this layer from relational databases, tables or raw string files.

A graph database platform is included in the platform. The uniformed data provided by the data processing layer is imported and maintained in the graph databases where data in different enterprises are still stored separately. Typical CURD functions of database are provided as well as the APIs tent to export data to the master data layer and receive update request when data need to be changed in multiple related data source.

The master data layer establishes multiple master database for each graph database which corresponding to the data source. MDM functions are provided to maintain the collected manufacturing data, and the key information are provided to the ontology layer.

The relations between data from different enterprises were established in the ontology layer. Ontologies are constructed by the key information by domain experts or algorithms. Connections revealed in the ontologies are provided for the graph database to update the data relations between isolated databases.

4 Key Techniques

The function of master data platform is to manage the data distributed in different systems. The data is collected, processed and their relations are analyzed in the platform, the procedure can be divided into 4 parts according to the 4 layers described in Sect. 3. Each layer has configuration panels for operators and users to manage the data processing procedures and interfaces are also provided for the data managing.

4.1 Data Processing Layer

In data processing layer, data from different enterprises were accumulated. These data have different sources thus they have diverse formats as well.

Data import functions are provided for users to configure which data were to submitted and what type of data they are submitting. Relational databases from manufacturing systems, information tables from internal management software as well as the raw data files generated by the production equipment is supported to be contained in the platform. The key data describe the structure of data is only needed to be stored in the platform for analyzing and managing. Data processed by the platform also maintained in the platform to insure the accessibility.

The data in different forms are collected, thus the data preprocess techniques is required to uniform the data in more flexible method rather than simple storing. The raw data generated in the equipment is varied and complicated, thus this function is required in this layer. For example, features should be extracted from time sequence data and items with wrong or missing features should be excluded from normal data. Standards like metrics and date forms should be unified in tables or databases.

Data inspection and rectification function are essential in this layer. Users can view the data that imported in the platform and check if there are mistakes in the import and preprocess procedures. The change of items in graph database is simple when the rectification is only performed to data items, while the modification for the structure of databases needs every layer to reconstruct the data involved.

4.2 Graph Database Layer

A graph database platform is maintained in the platform, and the key data processed and provided in the data processing layer were reorganized by their relations to support further analyses.

The key function of the graph database layer is importing and transforming the data from the data processing layer. The data stored in relational database can be described as table which is the most common form. These structural data can be described by the relations between each item using the header of table. The key or the major item of a row in the table can be described in nodes and the other items can be the attribute or another node with relations depend on its value.

Basic CURD function is provided by the graph database. The graph database performs better in search items according to the relations between nodes. Thus, the proper relations in the graph database is vital for the platform to operate correctly. In conversion, simple rules are applied for the automation process. The corrections of errors can be fixed by the manual updates. Update function is also required when the original data were modified.

A browser UI demonstrating the data in graph database is intuitive and simple method to inspect and rectify the items in the database. The CURD functions are integrated in the web UI and the items in database is presented by figures. Relations between nodes can be easily read by users through lines between entities and filtering irrelevant nodes.

4.3 Master Data Layer

In the master data layer, master data in each graph database is extracted and identified, and the master data management technique is available for users. The master data are marked and stored in graph database rather than a standalone database.

Master database establish function is essential for data management using master data. Key entities like customers, equipment ID or products is commonly selected to be the master data [2]. The data in the graph database is collected from multiple systems in one enterprise, so the same entity as master data may be showed up multiple times. Thus, though rule-based filter can be deployed in the platform, domain experts are often needed to extract master data and build master data models. Master database is established for each enterprise and maintained separately.

Master data management functions is supported in the master database layer. Operators in enterprise can utilize the benefit provided by the master data. Data reference is built between master data and items in the graph database which is another reference to the original data in disturbed systems. The graph database in the middle provide a unified and efficient tool to conduct the manage process to the data. MDM can match the information in multiple databases and merge them to provide detailed information. The modification of master data can be applied to all sources and applications across the databases. Historical records of audition trail data are also supported in the platform.

4.4 Ontology Layer

After master data is extracted and identified in the master data layer, each enterprise can manage their data more efficiently. While the master data across the enterprises still need to be identified and integrated, ontology is constructed to analyze the relations of master data. Interfaces are provided to manage the knowledge from the ontology and coordinate the change of data.

Ontology construction from the master data is a necessitate function for analyzing the relations between data from different enterprises. Domain experts are usually required to participate in the establishment procedures, because identifying and integrating master data in distributed enterprises needs practical experiences. Semi-automatic acquisition methods can be adopted to assist the construction procedure with template-based, rule-based ontology generation methods or machine learning algorithms if adequate amount of data is collected and validated.

The constructed ontology describes knowledge and provides support for building semantics, which means the data in distributed enterprises is linked and described in form of knowledge. The domain knowledges described by ontologies make data management, knowledge search, entity match, ontology fusion and other services is available for users. Integrating data across various sources and inspecting information related to the key data from other sources can significantly improve the quality of data and usability of data management.

With the linkages between nodes in all graph database are established by the knowledges described in ontology, the coordination of data changes between different enterprises is an addition function when users manage their own data. The request of changing data in other database of different enterprises can be pushed, and operators for corresponding database will get the notification and inspect the relevant data. This function is suitable for manufacturing enterprises exchanging their information.

5 Conclusion

By the establishment of the cloud-based master data platform, the manufacturing enterprises can manage their data efficiently for the benefits of exploiting the data relations between distributed databases. With the applications of graph database, master data management, and ontology analyses, the knowledge of the manufacturing domain is extracted and expressed. On this basis, several services are provided to support information integrating, exchanging and updating process in the cloud manufacturing systems.

In the future, researches will be focused on the algorithms to process data automatically. The procedure of extracting and evaluating master data, constructing ontology relies on domain experts to operate. With enough concrete data collected, a machine learning model can be trained and applied on the procedures to reduce consumption of manpower. Besides, the method to effectively utilize knowledge of linkage between data in the smart manufacturing system is another topic need further studies.

Acknowledgment. The research is supported by The National Key Research and Development Program of China No. 2018YFB1004001, and the NSFC (National Science Foundation of China) project No. 61572057 and 61836001.

References

1. Zhang, L.: A new manufacturing paradigm. Enterp. Inf. Syst. **8**(2), 167–187 (2014)
2. Loshin, D.: Master Data Management. Morgan Kaufmann, Burlington (2010)
3. Ron, H.: Combining computational models, semantic annotations and simulation experiments in a graph database. Database **2015** (2015). Article ID BAU130
4. Ren, L.: Cloud manufacturing: from concept to practice. Enterp. Inf. Syst. **9**(2), 186–209 (2015)
5. Chun, Z., Ren, L.: Study on a knowledge-based master data management method for manufacturing big data. In: CIE48, vol. 353, pp. 1–6 (2018)
6. Geiger, F.: Knowledge-based machine scheduling under consideration of uncertainties in master data. Prod. Eng. Res. Devel. **10**(2), 197–207 (2016)
7. Rivas, B.: Towards a service architecture for master data exchange based on ISO 8000 with support to process large datasets. Comput. Stand. Interfaces **54**, 94–104 (2017)

8. Chu, X., Morcos, J.: Katara: a data cleaning system powered by knowledge bases and crowdsourcing. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. ACM, pp. 1247–1261(2015)
9. Elbattah, M., Roushdy, M.: Large-scale ontology storage and query using graph database-oriented approach: the case of freebase. In: 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), pp. 39–43. IEEE (2015)
10. Küçükkeçeci, C.: Big data model simulation on a graph database for surveillance in wireless multimedia sensor networks. Big Data Res. **11**, 33–43 (2018)
11. Ravikumar, G., Khaparde, S.A.: CIM oriented graph database for network topology processing and applications integration. In: 2015 50th International Universities Power Engineering Conference (UPEC), pp. 1–7. IEEE (2015)