Cloud-based Data-intensive Framework towards Fault Diagnosis in Large-scale Petrochemical Plants

Zhiqiang Huo[1,2], Mithun Mukherjee[1], Lei Shu[1*], Yuanfang Chen[1,3], Zhangbing Zhou[2,3]

[1]Guangdong Provincial Key Laboratory on Petrochemical Equipment Fault Diagnosis,
Guangdong University of Petrochemical Technology, Maoming, China
[2]School of Information Engineering, China University of Geosciences (Beijing), China
[3]School of Information Engineering, Télécom SudParis, France
Email: zqhuo@cugb.edu.cn, m.mukherjee@ieee.org, lei.shu@ieee.org, yuanfang.chen.2009@ieee.org, zbzhou@cugb.edu.cn

Abstract—**Industrial Wireless Sensor Networks (IWSNs) are expected to offer promising monitoring solutions to meet the demands of monitoring applications for fault diagnosis in large-scale petrochemical plants, however, involves heterogeneity and Big Data problems due to large amounts of sensor data with high volume and velocity. Cloud Computing is an outstanding approach which provides a flexible platform to support the addressing of such heterogeneous and data-intensive problems with massive computing, storage, and data-based services. In this paper, we propose a Cloud-based Data-intensive Framework (CDF) for on-line equipment fault diagnosis system that facilitates the integration and processing of mass sensor data generated from Industrial Sensing Ecosystem (ISE). ISE enables data collection of interest with topic-specific industrial monitoring systems. Moreover, this approach contributes the establishment of on-line fault diagnosis monitoring system with sensor streaming computing and storage paradigms based on Hadoop as a key to the complex problems. Finally, we present a practical illustration referred to this framework serving equipment fault diagnosis systems with the ISE.**

Index Terms—**IWSNs, Fault Diagnosis, Big Data, Cloud Computing, Hadoop, Industrial Sensing Ecosystem.**

I. INTRODUCTION

Along with the rapid development of Industrial Wireless Sensor Networks (IWSNs) [1], [2], a large number of sensors enables to detect the physical environment of industrial production dedicated to monitoring of production, transportation, and surveillance. IWSNs plays a major role of data collection and management with advanced devices in industries. Implementing and deploying IWSN-based industrial monitoring and control systems have shown great potential on promoting efficiency and safety in large-scale petrochemical plants [3]. On this basis, industrial sensing ecosystem (ISE) aims to provide comprehensive and the collaborative sensing environment with the help of distributed topic-specific monitoring systems in the factories. Recently, large-scale petrochemical plants incorporate dense wireless devices for production monitoring, pollution analyzing, leakage detection [4], asset tracking, and safety monitoring (e.g., toxic gas monitoring [5]). For large-scale state monitoring and fault diagnosis to deduce latent damage of equipment and exposure to noxious leakage, it is necessary that the processing platform is highly integrated
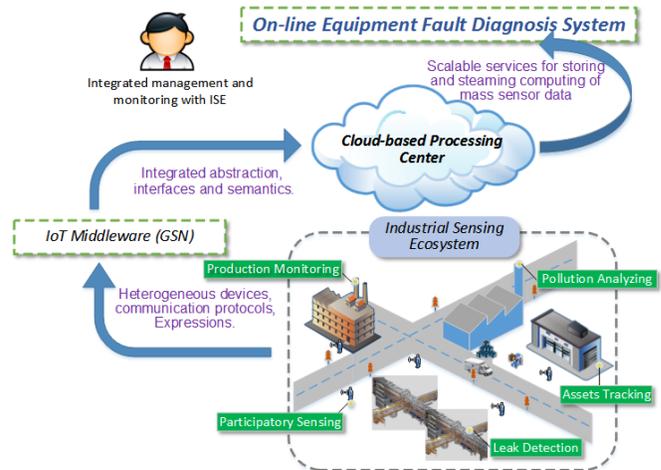
Fig. 1: ISE consists of topic-specific monitoring systems in devices and environmental monitoring. Moreover, Internet of Things (IoT) middleware facilitates heterogeneity issues, afterwards, cloud-based processing center supports the addressing of data-intensive problems, e.g., massive sensor data storage, computing, knowledge discovery.

and capable of addressing issues with massive storage and computing generated by a vast sea of sensor data. However, figuring out heterogeneity and Big Data problems on one general platform for fault diagnosis is a challenging issue demanding innovative solutions. Fig. 1 illustrates key elements and processes for on-line equipment fault diagnosis system with the ISE. Thus, the two major research challenges are presented as follows:

- Heterogeneity and isolation commonly exist in domain-specific and customized sensing systems due to diverse wireless devices with matched softwares to execute on-site monitoring and decision-making tasks. Inconvenient integrated interoperability and transverse data sharing have restrictions on building up all-round and cooperative industrial monitoring applications on general platforms.
- Large amount of sensors, which have been connected to the sensor networks (SNs), have started another informa-

TABLE I: Cloud-integrated WSNs for Equipment Fault Diagnosis with IoT

Scheme/Architecture	Focus	Feature/Contribution
Huang et al. (2015) [9]	WSNs as new signal collection and transmission techniques for state monitoring and fault dignosis of mechanical equipment.	The development process and classifications of fault diagnosis of mechanical equipment based on WSNs are presented.
Tiwari et al. (2007) [10]	To monitor machinery condition-based maintenance in small machinery spaces based on WSNs.	Wireless status monitoring sensor network with a hardware platform, networking architecture, and medium access communication protocol.
Yang et al. (2015) [11]	Fault diagnosis for error detection and location in big sensor data sets generated by large-scale sensor networks.	A data-error detection approach exploits the computation potential of cloud platform and the network feature of WSNs.
Alam et al. (2010) [12]	SOA presents a set of architectural principles that encapsulate the functions into generic services which are transmitted over a communication network.	SOA bridges the gap between sensor nodes and enterprise applications.
Distefano et al. (2012) [13]	A cloud is viewed as computing infrastructure to facilitate the management of IoT resources.	The IoT management functions are regarded as peer services provided by a typical cloud service model.
Chenaru et al. (2015) [14]	To integrate IWSNs with cloud infrastructures, and to enable remote access over secure and scalable real-time communications.	An architecture that implemented RESTful services at a coordinator node level of WSNs.
Ahmed et al. (2011) [15]	An integrated framework to fully utilize the valuable data. A gateway to collect sensory data from WSN, and to provide publish-subscribe services.	Considers the related approaches to user management, access control, storage and retrieval of distributed data.

tion explosion. Traditional computing approaches are not able to handle the increasing demands on processing big sensor data with specific response times for data-intensive applications. Moreover, distributed sensing systems suffer from problems caused by semi-structured or even unstructured data format and heterogeneous semantics.

Due to such problems, there is a strong need to provide higher abstraction, integrated management, and scalable data processing services to facilitate the development and deployment of fault diagnosis monitoring applications based on IWSNs in large-scale petrochemical plants. Therefore, how to solve the problems caused by heterogeneity of these distributed and data-intensive monitoring systems (e.g., on-line equipment and environmental monitoring) are worth to be studied.

Cloud Computing has emerged as a solution to extend the capabilities of a system in massive storage, computing, and software services in a scalable manner at low cost. This advantage of Cloud Computing helps to overcome data-intensive problems. Thus, it is necessary to build up a cloud-based framework that consists of high-level abstraction of diverse sensor streaming and flexible data processing services. For the important components, data processing services with some promising programming models, e.g., Hadoop MapReduce [6], [7] and NoSQL databases [8] have been discussed to handle the data-intensive problems. This will help the service engineers and fault diagnosis managers to focus on the implementation of core transactions without considering the heterogeneity and data-intensive problems.

This paper presents a Cloud-based Data-intensive Framework, named as CDF, that supports large-scale equipment fault diagnosis systems in a cloud environment. The objective of this framework is to enable the integration of distributed monitoring systems, and the development and implement equipment fault diagnosis and analysis system through serving massive storage and computing paradigms in large-scale petrochemical plants.

The remainder of this paper is organized as follows. Section II briefly summarizes related work on fault diagnosis based on IWSNs, and overviews the related work to IoT, service-oriented architecture (SOA), and cloud architecture. We discuss major components and implementation of the proposed framework in Section III. An architecture of CDF for equipment fault diagnosis systems in large-scale petrochemical plants is illustrated as an example in Section IV. Finally, conclusions are drawn in Section V.

II. RELATED WORK

Wireless sensor networks (WSNs) are prevalent in state monitoring and fault diagnosis of mechanical equipment as a new signal collection and transmission technology. Huang et al. [9] presented recent techniques, and the development process and classifications of fault diagnosis for mechanical equipment based on WSNs. An IWSN to monitor machinery condition-based maintenance in small machinery spaces has been developed and reported in [10]. Moreover, in purpose of conducting error detection and location in big sensor data sets, Yang et al. [11] proposed a data error detection approach that exploits the computation of cloud platform and the network feature of WSNs.

In recent years, convergence of the cloud and SNs has been increasingly studied to exploit their intrinsic interoperability. Table I summarizes the related work on equipment fault diagnosis based on cloud-based WSNs with IoT. For instance, the sensor virtualization through semantic abstraction for sensor capabilities is enhanced in [12]. In addition, SOA that bridges the gap between sensor nodes and enterprise applications including factory monitoring, control and tracking systems is a promising technique. SOA presents a set of architectural principles that encapsulate the functions into generic services

which are transmitted over a communication network. A cloud is viewed as the computing infrastructure to facilitate the management of IoT resources. Recently, Distefano et al. [13] proposed an approach to accommodate the growing scale and diversity of IoT integrated with Cloud Computing. The IoT management functions are regarded as peer services provided by infrastructure management with typical cloud service model. Moreover, in purpose of enabling remote access over secure and scalable real-time communication channels, Chenaru et al. [14] proposed an architecture that implemented representational state transfer (REST) services at a coordinator node level of WSNs in the cloud. Meanwhile, in order to fully utilize the valuable data, Ahmed et al. [15] discussed an integrated framework. This framework applied a gateway to collect sensory data from WSN, and to provide publish-subscribe services. Actually, it considered the related approaches to user management, access control, storage, and retrieval of distributed data.

Although, the aforementioned works provide a very useful overview of cloud-based WSNs for equipment fault diagnosis, most of the work focused on the resource management, and the design and implementation of domain-specific monitoring systems based on WSNs rather than the data-intensive oriented service delivery for state monitoring and fault diagnosis. Therefore, there is still lacking desired service deliver paradigms towards the integration and implementation of data-intensive monitoring systems for state monitoring and fault diagnosis, providing scalable storage and computation paradigms for amounts of sensor data with ISE. In fact, the generic nature of the Cloud Computing and techniques of Big Data makes it attractive and suitable to realize integrated general platform. Therefore, it employs the powers of processing large amounts of sensor data, and enables flexible interactions based on cloud.

III. Proposed CDF Framework: Design And Implementation

In this section, we propose the CDF framework in the cloud. Fig. 2 outlines high-level major components with typical cloud service model in this framework. Afterwards, we introduce the particular role and functionality of each component.

Basic service models in Cloud Computing: Cloud Computing has started to emerge on exploiting the service delivery models of cloud to accommodate the growing scale and diversity of data services. It has extensively used three basic types of service models to categorize the cloud services: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). SaaS is software delivery model that provides user services on a hosting environment and allows customers access the plentiful resources through Internet. PaaS is a development platform that supports users to develop and deploy services and applications directly on the PaaS cloud without previous investment. IaaS provides pay-per-use pattern that delivers access to cloud storage, processing, networks and other fundamental computing resources.
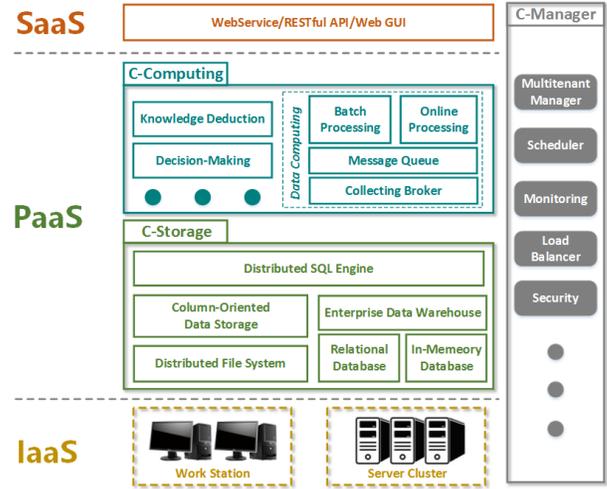

Fig. 2: The proposed framework of CDF and major components are presented with typical cloud service model: IaaS, PaaS, and SaaS levels.

In this proposed framework, at SaaS level, both REST and Simple Object Access Protocol (SOAP) services can be implemented for non-critical data acquisition applications providing increased and scalable services, e.g., acquisition of sensor data streaming, interactions and management of user-cloud. REST is a modern software architecture, designed for scalable web services implementation, and SOAP with standard XML protocol makes it flexible and cross-platform for applications. Meanwhile, the efficient visualization methods of configuration and management not only facilitate the applications development, but also enhance the security in the cloud. At IaaS level, the capabilities of infrastructures can be extended with increased storage, processing power with a scale-out configuration using general-purpose servers. Moreover, through applying visualization, resources (e.g., hard disk, CPU, and memory) of server clusters can be achieved like an integral part without considering the workload and optimal characteristics of a server.

At PaaS level, *C-Storage, C-Computing*, and *C-Manager* take responsibility for supporting major storage and computing paradigms, and service management in the cloud. Roles, functionalities, and implements of these three components are discussed respectively as follows:

1) C-Storage:

Distributed File System (DFS) performs as a distributed file system that provides fault-tolerant and high-throughput access to application data in CDF, e.g., semi-structured and unstructured data and files. In addition, in purpose of serving multi-tenants, it implements a version manager and a multi-tenant manager provided by *C-Manager* module. This module acts as the infrastructure to the implementation of distributed storage.

NoSQL and Relational DataBases (NoReDB) serves as a hybrid data management module comprising unstructured databases and RDBMS. Unstructured data from industrial

sensing systems (e.g., production and transmission equipment monitoring, toxic gas monitoring, and video surveillance), business transactions, and external open data sources could be stored in column-oriented database such as Apache HBase [16]. RDBMS provides solutions of storage, and Atomicity, Consistency, Isolation, Durability (ACID) operations to process structured data from users, production feedbacks and other meta information. Moreover, in-memory database stores all data in main memory, thereby eliminating the need for disk access, resulting in high-speed data acquisition between pub-sub data services and applications. This module provides flexible APIs and object-entity mapping on the top of multiple databases, interfaces data access approaches, and simplifies migration of application databases.

Enterprise Data Warehouse (EDW) provides data strategy collections for creating analytical analysis, reports, and decision-making to promote the quality of transactions, monitoring, and control. This module stores subject-oriented and disparate data sources, e.g., condition monitoring of equipment, RFID systems for assets tracking, and toxic gas leakage monitoring. EDW restructures hybrid data so that it delivers excellent query performance with topic-oriented common data models for all data of interest regardless of the data's source.

Distributed SQL Engine (DSE) provides flexible SQL-based query functions towards large volumes of sensor data, and simplifies hybrid data management in NoReDB module. Moreover, DSE facilitates the complex query processing and supports ACID operations of mass sensor data that supplements the columned-based database. For this purpose, Impala [17] and PostgreSQL [18] can be implemented to support SQL-based operations for data-intensive monitoring applications, as well as plentiful APIs.

2) C-Computing:

Data Computing (DC) carries out computing processing tasks submitted by developers upon a cluster of machines, comprising of off-line *Batch Processing* and *On-line Processing*. *Batch Processing* module handles transactional processing of large volumes of sensor data that tolerates specific response times, e.g., Create, Retrieve, Update, and Delete (CRUD) operations among historic mass data. *On-line Processing* module provides streaming programming framework to satisfy real-time calculation and assessment. There is necessary to satisfy the specific response times when designing and implementing topic-specific programming APIs for fault diagnosis managers, as well as considering allocation of resources. Based on Hadoop, the proposed programming framework is available to apply MapReduce, Storm [19], Spark [20], and Spark streaming computing platforms. Considering features of high concurrency while a large number of sensor devices measuring physical environments, *Message Queue* acts as caching pipelines to balance the flow traffic and provide the non-blocking transmission in the context. *Collecting Broker* pulls sensor data streaming from IoT middleware in which *Agent* relays sensor reading in real time and upload historical sensor data from available data sources. Therefore, DC acts as a computing paradigm of handling massive sensor data, which

work as computing infrastructures supporting data mining and on-line applications.

Knowledge Deduction (KD) helps developers build a programming environment for quickly creating scalable performant data mining applications (e.g., machine learning and data analysis), which targets at helping developers create topic-specific data mining approaches and providing some off-the-shelf algorithm implements (e.g., Apache Mahout [21]). Managers enable regularly analyze massive historic sensor data to discover potential relationships among physical productions and quality of products with expert knowledge.

Decision-Making (DM) is a strategy module that realizes the technical business support for the establishment in large-scale equipment fault diagnosis, monitoring applications, and administrator management. DM provides functional strategies extracted from specific businesses such as public security domain, databases, algorithms, physical monitoring, which are based on existing compositions according to logistics functional management, expert support system and emergency reporting systems.

3) C-Manager:

C-Manager works as a resource scheduler and monitoring center that realizes the management of components to the overall federation in clouds.

Task Scheduler (TS) performs for planning the execution of non-transactional tasks. It determines the execution order of tasks according to the importance and priority of application requests, e.g., analysis of safety monitoring is more crucial that should be given as prior priorities to be executed. Another responsibility of this module regards to the approval control and resource allocation, depending on the deadlines of running applications, availability of idle resources and demands for required performance.

Load Balancer (LB) keeps balanced processing services of load and orders for new requests applying appropriate routing and caching approaches. It charges the allocation of new requests for storing and computing tasks, which provides stable and scalable interactions in CDF.

Resource Monitoring (RM) takes responsibility for performing the resource budgets and keeping track of resource utilization periodically. RM monitors requests and the performance of executions in terms of the execution time, occupation and release of resources, and performance evaluation of transactions. This information can be utilized to update the TS for dynamic resource allocation or execution migration to another machine with rich resources. All the above information related to transactional operations and applications will be recorded in logging systems for further analysis and planning.

Multi-tenant Manager (MM) provides a consolidated view of resources that are accessible by each tenant. Different domain-specific applications are assigned specific physical and computing resources according to their priorities and demands for capabilities of data processing. In CDF environment, the resources include not only cloud resources such as computing capability and software instances in PaaS platform, but also the publish-subscribe services of sensor data streaming. For

instance, the data sharing between horizontal sensing systems helps build up all-around intelligent applications, e.g., cooperations between the toxic gas monitoring system and alarm systems may construct real-time early warning strategies in large-scale petrochemical plants.

Security focuses on providing protecting approaches in authentication, authorization and availability. Authentication and authorization are information protections targeting at verifying and validating permissions of accessing cloud resources and sensor data streaming services for each application request, which is always maintained by system administrator and service providers. OAuth is an open protocol to allow secure API authorization in a simple and standard method from web, mobile and desktop applications that can be considered as protectors in CDF. Meanwhile, there is a need to build up IP blacklist, firewall and restricted synchronized connections in case of some malicious attacks.

IV. AN EXAMPLE OF CDF SYSTEMS: EQUIPMENT FAULT DIAGNOSIS IN IWSNS

We provide equipment fault diagnosis system with IWSNs as a typical user-case scenario that validates the implementation of our proposed CDF prototype in the large-scale petrochemical plants. On-line fault diagnosis and analysis of toxic gas monitoring and production equipment that are critical issues in the large-scale oil and gas industries, contributes to network early warning, facility maintenance, and safety evacuation for first-line workers, reducing the latent danger and enhancing the productivity and safety.

The validating scenario contains necessary elements to implement our CDF system. Particular sensor devices for toxic gas monitoring are presented as follows:

- Wireless sensing devices collaboratively measure air information to administrators, e.g., CO, SO_2, CH_4, wind speed, humidity, temperature, e.g., toxic gas sensor node and wearable equipment. Moreover, The smart helmet [22] with one STM32 processing chip and plentiful sensors enables mobile monitoring, and reports the geographical locations and first-line workers' vital signs. Moveover, ARM devices contribute to information visualization and early-warning approaches through using TFT-LCD touch screen, vibration motor and RF communication module. These promising approaches enable form a network in self-organized manner and continuously upload data streaming leading to amounts of sensor data.
- Wireless sensors allows managers to monitor physical conditions of production and transmission equipment with hazards in remote, hard-to-reach, and prohibited regions. This facilitates the data collection for status monitoring and fault diagnosis of mechanical equipment.

This comprehensive sensing environment not only constructs safety monitoring in working regions, but also contributes to all-round equipment fault diagnosis system in large-scale petrochemical plants. Based on large numbers of sensor data from topic-specific monitoring systems, equipment fault diagnosis and analysis need stable and scalable storage and
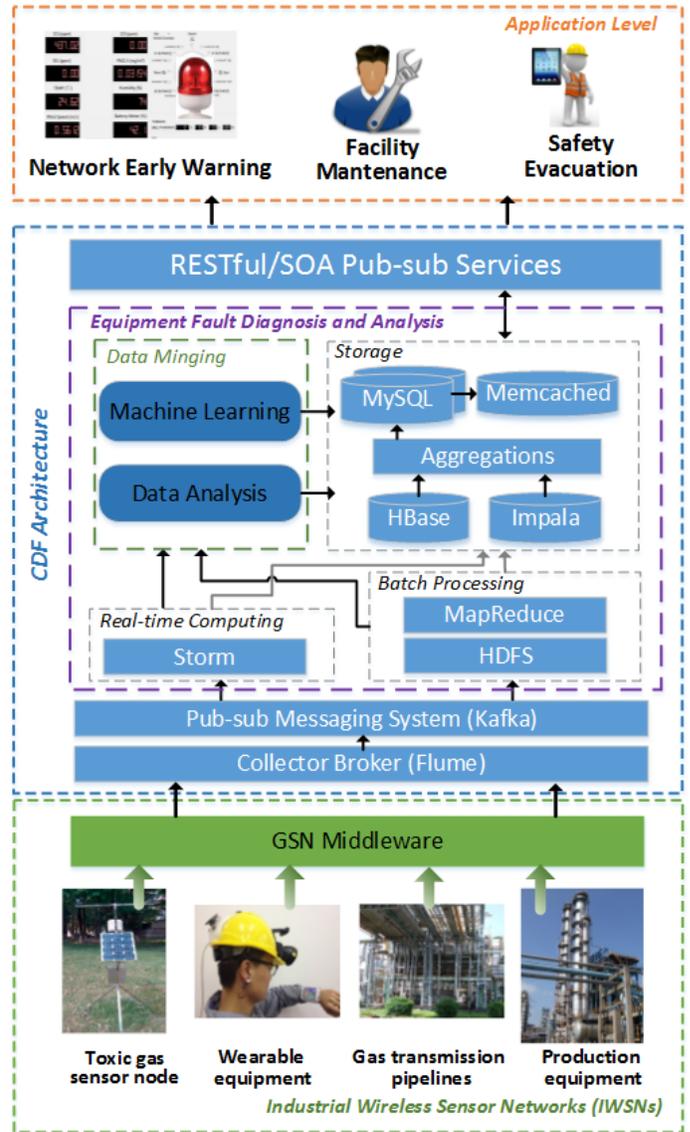

Fig. 3: Equipment fault diagnosis system for toxic gas monitoring in the petrochemical plants as an example of CDF system.

computing platforms to be carried out. To solve this, Fig. 3 illustrates the proposed architecture to facilitate the integration of distributed monitoring systems and provide cloud-based storage, computing, and data mining platforms to support the equipment fault diagnosis system in large-scale petrochemical plants.

IoT middleware facilitates the heterogeneity problem with topic-specific monitoring applications in IWSNs. Therefore, we choose GSN [23] as gateway middleware to interface and enable sensor data streaming uploaded to communication modules in CDF. GSN is an Internet-scale infrastructure for rapid deployment and integration of heterogeneous IWSNs, which targets at flexible configuration, general abstraction and distributed query support, on this basis, CISI middleware [24] is selected to support crowdsensing-based applications based

on GSN. To facilitate the interaction between GSN and CDF, we add *Agent* module for relaying sensor data streaming to *Collector Broker* module. Afterwards, after across pub-sub messaging system, sensor data streaming flows into computing paradigm (i.e., batch processing and streaming computing), which supports real-time industrial monitoring and fault diagnosing and analyzing. Based on computing paradigms for data-intensive applications, fault diagnosis managers enable to conduct data mining of no-obvious correlations in industrial environment with the domain knowledge of workers, i.e., experience, knowledge and wisdom.

For instance, it has great potential to infer toxic gas leakage through deducing the mutual effects with the change tendency of targeted gas concentration and pressure of valve along pipelines for oil and gas in large-scale petrochemical plants. On this basis, a safe and intelligent ecosystem can benefit from real-time equipment fault diagnosis system in large-scale petrochemical plants. Although we discussed the equipment fault diagnosis application for toxic gas monitoring as an example, the core of our proposed CDF architecture is not limited by this application. This architecture can be easily extended to plentiful topic-specific applications by applying domain knowledge according to the monitoring requirements in large-scale petrochemical plants.

V. Conclusion

This paper proposed a cloud-based framework to support the data-intensive application development. This framework is developed to satisfy the increasing requirements on massive sensor data storage and processing for equipment fault diagnosis. Such sensor-based IoT data is generated from large-scale industrial sensing ecosystems. In addition, this framework enables the integration of heterogeneous infrastructures and provides services for storage, computation, and sensing. It makes application developers be able to focus on their special logic. Apart from existing related research work, this paper presents how to design and implement the streaming computing and storage paradigm on sensor data, in industrial sensing ecosystems, based on Hadoop in the cloud. Finally, we provided a typical use case scenario of large-scale industrial ecosystems, equipment fault diagnosis, as a practical example to validate the implementation of the proposed CDF prototype.

References

[1] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches," *IEEE Trans. Ind. Electronics*, vol. 56, no. 10, pp. 4258–4265, Oct. 2009.

[2] 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," in *Proc. Int. Conf. on Industrial Eng. and Eng. Manage. (IEEM)*, Bandar Sunway, Dec. 2014, pp. 697 – 701.

[3] M. R. Akhondi, A. Talevski, S. Carlsen, and S. Petersen, "Applications of wireless sensor networks in the oil, gas and resources industries," in *Proc. 24th Int. Conf. on Advanced Information Networking and Applications (AINI)*, Perth, Australia, 20–23 Apr. 2010, pp. 941–948.

[4] K. Wang, H. Lu, L. Shu, and J. J. P. C. Rodrigues, "A context-aware system architecture for leak point detection in the large-scale petrochemical industry," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 62–69, Jun. 2014.

[5] L. Shu, J. Zeng, K. Li, Z. Huo, X. Wu, X. Wu, and H. Sun, "Wifi-based smart car for toxic gas monitoring in large-scale petrochemical plants," in *Proc. Int. Conf. Consumer Electronics-Taiwan (ICCE-TW)*, Taipei, Taiwan, 6–8 Jun. 2015, pp. 200–201.

[6] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proc. IEEE 26th Int. Symp. on Mass Storage Systems and Technologies (MSST)*, Incline Village, Nevada, 3–7 May, 2010, pp. 1–10.

[7] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters." *Proc. ACM 6th Int. Symp. on Operating Systems Design and Implementation (OSDI)*, vol. 51, no. 1, pp. 107–113, 6–8 Dec. 2004.

[8] R. Cattell, "Scalable SQL and NoSQL data stores," *Acm Sigmod Record*, vol. 39, no. 4, pp. 12–27, May, 2010.

[9] J. Huang, G. Chen, L. Shu, Q. Zhang, and X. Wu, "WSNs-based mechanical equipment state monitoring and fault diagnosis in China," *International Journal of Distributed Sensor Networks*, vol. 2015, 2015.

[10] A. Tiwari, P. Ballal, and F. L. Lewis, "Energy-efficient wireless sensor network design and implementation for condition-based maintenance," *ACM Trans. on Sensor Networks*, vol. 3, no. 1, pp. 1–23, 2007.

[11] C. Yang, C. Liu, X. Zhang, S. Nepal, and J. Chen, "A time efficient approach for detecting errors in big sensor data on cloud," *IEEE Trans. Parallel and Distributed Systems*, vol. 26, no. 2, pp. 329–339, 2015.

[12] S. Alam, M. M. R. Chowdhury, and J. Noll, "SenaaS: An event-driven sensor virtualization approach for Internet of Things cloud," in *Proc. IEEE Int. Conf. on Networked Embedded Systems for Enterprise Applications (NESEA)*, Suzhou, China, 25–26 Nov. 2010, pp. 1–6.

[13] S. Distefano, G. Merlino, and A. Puliafito, "Enabling the Cloud of Things," in *Proc. IEEE 6th Int. Conf. on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, 4–6 July, 2012, pp. 858–863.

[14] O. Chenaru, G. Stamatescu, I. Stamatescu, and D. Popescu, "Towards cloud integration for industrial wireless sensor network systems," in *Proc. ACM 9th Int. Symp. on Advanced Topics in Electrical Engineering (ATEE)*, 7–9 May, 2015, pp. 917–922.

[15] K. Ahmed and M. Gregory, "Integrating wireless sensor networks with cloud computing," in *Proc. IEEE 7th Int. Conf. on Mobile Ad-hoc and Sensor Networks (MSN)*, Beijing, China, 16–18 Dec. 2011, pp. 364–366.

[16] "HBase." [Online]: http://hbase.apache.org/

[17] "Impala." [Online]: http://www.cloudera.com/products/apache-hadoop/impala.html

[18] "PostgreSQL." [Online]: http://www.postgresql.org/

[19] "Storm." [Online]: http://storm.apache.org/

[20] "Spark." [Online]: http://spark.apache.org/

[21] "Mahout." [Online]: http://mahout.apache.org/

[22] L. Shu, K. Li, J. Zen, X. Li, H. Sun, Z. Huo, and G. Han, "A smart helmet for network level early warning in large scale petrochemical plants," in *Proc. 14th Int. Conf. on Information Processing in Sensor Networks (IPSN)*, Seattle, USA, 13–16, Apr. 2015, pp. 390–391.

[23] K. Aberer, M. Hauswirth, and A. Salehi, "Infrastructure for data processing in large-scale interconnected sensor networks," in *Proc. IEEE Int. Conf. on Mobile Data Management (MDM)*, Mannheim, Germany, 7–11 May, 2007, pp. 198–205.

[24] Z. Huo, L. Shu, Z. Zhou, Y. Chen, K. Li, and J. Zeng, "Data collection middleware for crowdsourcing-based industrial sensing intelligence," in *Proc. ACM Int. Conf. on Mobility and MiddleWare Management in HetNets (MobiMWareHN)*, Hangzhou, China, 22–25 Jun. 2015, pp. 3–8.