A Smart Service Warehousing Platform Supporting Big Data Deep Learning Modeling Analysis

Chih-Hung Chang[1], Tse-Chuan Hsu[2], William Cheng-Chung Chu[3], Che-Lun Hung[4], Ping-Fang Chiu[5]

[1] College of Computing and Informatics, Providence University, Taiwan
[2] Department of Healthcare Information & Management, Ming Chuan University, Taiwan
[3] Department of Computer Science, Tunghai University, Taiwan
[4] Institute of Biomedical Informatics, National Yang-Ming University, Taiwan
[5] Nephrology Division, Department of Internal Medicine, Changhua Christian Hospital, Taiwan
ch.chang@gm.pu.edu.tw, tsechuan.hsu@gmail.com, cchu@thu.edu.tw, clhung@ym.edu.tw, 68505@cch.org.tw

Abstract

Chronic disease management is the most expensive, fastest growing and most difficult problem for medical care workers in various countries. Current Health care information systems do not have interoperability characteristics and lack of data model standards, which makes it very difficult to extract meaningful information for further analysis. Deep learning can help medical care giver analyze various features of collecting data of patients and possibly more accurately diagnose and improve medical treatment through early detection and prevention. Our approach uses P4 medical model, which is predictive, preventative, personalized and participatory, which identifies diseases at early stage of diseases development, therefore it helps patients improve their daily behavior and health status.

In this paper, an effective and reliable intelligent service warehousing platform, which is a service framework and a middle layer, is designed to maintain the quality of service of the intelligent health care system and to analyze and design to predict the risk factors that contribute to diabetes and kidney disease. The mathematical prediction model is provided to doctors to support their patient's treatment. At the end we verified the availability and effectiveness of this service platform from the data of hospital.

Keywords: Big data, Dependable system, Deep learning, Modeling, Microservice

1 Introduction

The Ministry of Health and Welfare, Taiwan (R.O.C) investigated in the 2015, and the top ten causes of death were mainly chronic diseases. Care for chronically ill patients requires longer care and high medical costs. In the aging society, the increase of chronic diseases, the geographical dispersion of residents and the shortage of medical staff. The medical system must respond to changes, and the disease care strategy must also be innovative. It is a solution of using technology to reduce the manpower of home care. The trends of medical are preventive medicine, personalized medicine and integrated health management. At present, hospitals have a complete electronic medical record (Hospital Information System, HIS), which is used to assist patients to manage the electronic medical record. However, there is no shortage of data, to integrate with intelligent data decision analysis, big data analysis and deep learning.

In this paper, we propose a smart health care system to help improve care for chronic patients. We adopted the P4 medicine model proposed by the US NIH: *Prediction*, *Prevention*, *Personalized*, and *Participatory* [1]. Taking the medical habits of chronic diseases as an application example, through the collection and analysis of health care information, the purpose of predicting and preventing diseases is achieved. It also provides personalized services and increases the benefits of user engagement. Prevention and early detection help reduce the burden of health care and caregivers, and contribute to chronic diseases and diseases that require long-term care. To implement the P4 medicine model, you must have a reliable and efficient computing platform to provide a data-efficient, high-efficiency cloud service platform for data collection, integration, analysis, and verification. This study combines modeling analysis and deep learning to establish a highly efficient and intelligent chronic disease care system. Through the data collection and integration analysis, learning and calculation of the service platform, users and professional medical personnel can obtain more accurate and efficient analysis results. Professional medical staff assists patients in health care and treatment.

The structure of this paper is as follows. In the second chapter, we will briefly describe the research

*Corresponding Author: William Cheng-Chung Chu, E-mail: cchu@thu.edu.tw
DOI: 10.3966/160792642021032202022

related to this topic. The third section introduces the development process and system implementation of the reliability intelligent service warehousing platform we designed. The fourth section is the conclusion and future work.

2 Problem Formulation

In this section, we will briefly introduce the relevant research cited in this paper.

2.1 Cloud Computing

According to the National Institute of Standards and Technology (NIST), the definition of cloud computing identifies "five essential characteristics" [2]:

- On-demand self-service.
- Broad network access.
- Resource pooling.
- Rapid elasticity.
- Measured service.

In essence, cloud computing is a new application of distributed computing. The basic concept of cloud computing is to automatically split a large number of computational processes into a lot of smaller sub-processes through the Internet, and then hand it over to multiple servers. The huge system is composed of the search and operation analysis, and then the processing result is transmitted back to the user. Through this technology, the Internet Service Provider (ISP) can process millions or even billions of information in a matter of seconds to achieve the same performance as the "supercomputer."

2.2 Big Data

The big data [3-4] issues are how to find useful information from huge data to analyze and discover valuable information. It includes many challenges, such as how to acquire, store, search, share, analyze, and visualize. The challenge of big data has brought about changes in IT technology and commercial software. Of course, the database and data warehousing operators are the first to bear the brunt. Due to the big data and the emergence of demand, the role of data warehousing will be very obvious in 2011. The change is in response to the challenges of massive, complex, and diverse forms of information. One of the factors that has changed the role of data warehousing is that the data are not only increasing in number but also increasingly complex. For data warehousing systems designed 5 to 10 years ago, information must also be processed. Diversity, complexity, huge capacity, and immediate response to system speed.

Basically, big data represent large, complex, and unstructured data. Historically, data warehousing systems have been used in the field of OLAP (On-Line Analytical Processing), focusing on data mining and in-depth analysis. In the past, there was no immediate need for deep data mining and modeling, so the time and immediate requirements were not as high as in the transaction-based OLTP (On-Line Transaction Processing) field. However, with the advent of the big data, the amount of data has not only grown substantially, but the types and patterns of data have become more and more complex. Users have to respond more quickly to the rapid analysis of a large number of messages.

In response to the demand for big data, how to efficiently process and analyze huge amounts of data has become an important issue. Many traditional technologies have been applied to massive analysis, such as association rule learning [5], classification, cluster analysis, genetic algorithm (GA), machine learning, support vector machine (SVM) [6-7], neural networks, natural language processing, time series analysis and data visualization.

2.3 Machine Learning and Deep Learning

Machine learning [8-9] is a multi-disciplinary subject that has emerged in the past 20 years and involves many disciplines such as probability theory, statistics, approximation theory, convex analysis, and algorithm complexity theory. The theory of machine learning is mainly to design and analyze algorithms that allow computers to automatically "learn". Machine learning algorithms are a class of algorithms that automatically analyzes and obtain rules from data and use rules to predict unknown data. Machine learning has been widely used, such as: data mining, computer vision, natural language processing, biometrics, search engines, medical diagnostics, detecting credit card fraud, securities market analysis, DNA sequencing, speech and handwriting recognition, strategy Games and robots.

Deep learning [10-12] is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised.

Deep learning has been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, Bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts. [13-15]

2.4 Middleware

In cloud computing technology, Middleware [16-17] is located between the service and server clusters to provide management and services. The mediation software focuses on system integration in the cloud computing environment and provides services for mediating pre-operations. It provides a unified, standardized programming interface and protocol for applications, hides the underlying hardware, integrates the heterogeneity of the operating system and the network, and manage network resources. For establishing an efficient software middleware layer, how to make a trade-off between technology and efficiency among various integrated components in the environment is a very important issue. Appropriate penetration performance reduces the complexity of the application software and optimizes the use of system resources when conditions are limited.

Because of this unique feature, the software middleware layer can integrate systems that implement a secure pervasive computing environment to achieve the development of applications that integrate and manage intermediaries. Such as: consistent application development support, user management, resource management, and mapping management.

2.5 Microservice

Microservices [18] are a software development technique—a variant of the service-oriented architecture (SOA) architectural style that structures an application as a collection of loosely coupled services. In a microservices architecture, services are fine-grained and the protocols are lightweight. The benefit of decomposing an application into different smaller services is that it improves modularity. This makes the application easier to understand, develop, test, and become more resilient to architecture erosion.

3 System Architecture

The platform is the overall system coordination role. The task of the platform has included the integration of service process integration and service data between internal and external systems as a whole structure, including: (1) user query and data collection (2) external database Extract-Transform-Load (3) Data Training and (4) Prediction, the architecture is shown in Figure 1:

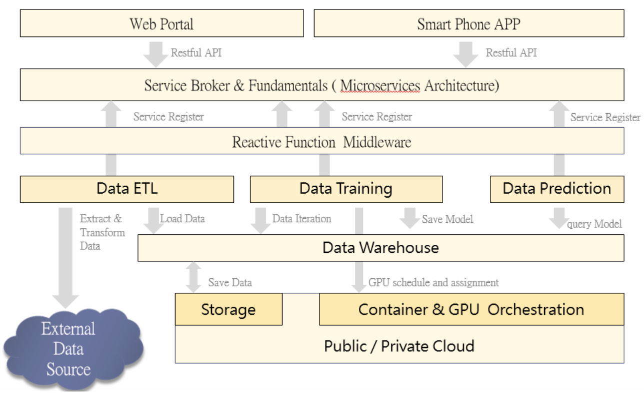


Figure 1. Our proposed system architecture

3.1 Analysis and Confirmation of the Requirements of the Smart Service Warehousing Platform

system design is collecting and analyze the existing health service systems, related applications and research, and integrate the necessary features and needs of health services. The intelligent chronic care system is extremely user-friendly, and the access rights are different depending on the identity (such as doctors, nurses, patients), and the system contains multiple settings such as privacy settings, medical information and functions, and the system will be extremely complicated. If there is no good-planned user interface and related auxiliary functions, it will cause others to operate smoothly and even endanger health. Therefore, it is extremely important to improve the usability. We will analyze the above functions in detail to find out the complete requirements and the system validation functions to be interfaced with the future.

3.2 Container Based Microservices Integration Architecture

Microservices have become popular in recent years. The essence and spirit of microservice are consistent with SOA. They are also suitable for service integration of large-scale systems, but microservices is simpler and more efficient. With the Docker, the container technology is mature and the Docker runs under Container. The microservice architecture is more effective in achieving high integration and high performance. The microservice integration architecture is an integration framework for inter-subsystem modules. The intelligent data warehousing, big data analysis, and AI deep learning related services, all are integrated in the microservice integration architecture as shown in Figure 2.

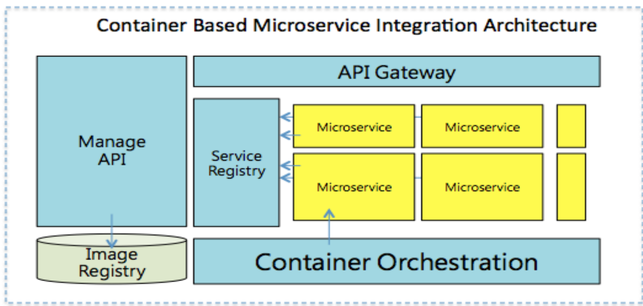


Figure 2. Service integration architecture

The microservice integration architecture consists of four main modules:

- Container orchestrator

Responsible for scheduling the computing resources provided by the cloud platform to provide SLAs that meet the service operation.

- Service Registry

Provide a unified registration mechanism for all microservices so that services can find and call each other.

- API Gateway

  Provides all microservice API access interfaces.

- Manage API

  Provides a resource management API that can be used to dynamically expand or shrink computing resources.

Figure 3 is showing the process of container manage and dynamic deployment monitor.

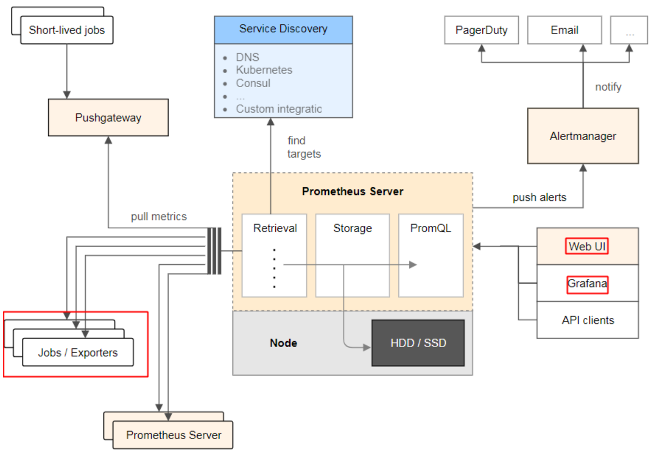


Figure 3. Container Management and Dynamic Deployment Monitor

3.3 Function Reactive Middleware

Middleware plays the role of centralized service process definition and control when the all system modules are integrated with microservices. Functional Reactive Programming (FRP) is a program development technology in recent years, it is suitable for decentralized and asynchronous architectures. This system uses the architectural concept of FRP to be imported into the architecture of Middleware. The purpose is to control the execution efficiency of the microservices to achieve more flexibility and high-speed execution efficiency. Therefore, it is called Functional Reactive Middleware, and its architecture diagram is shown in Figure 4.

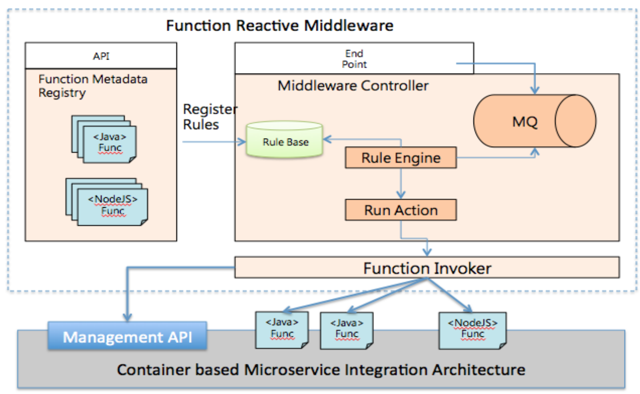


Figure 4. Function Reactive Middleware architecture

Functional Reactive Middleware is stacked on top of the microservice integration architecture and it is responsible for the establishment and execution of Function Chaining between microservices. This architecture consists of three subcomponents:

- Function Metadata Manager

  It is responsible for managing function code and its metadata description, and dynamically parsing its operational process rules and execution environment requirements according to the metadata description.

- Middleware Controller

  It contains a high-capacity message queue management, can accommodate a large number of demand messages, and processes messages through the clustered rules engine to drive the execution of functions after the decision is made.

- Function Invoker

  It is responsible for function execution calls, dynamically scheduling the underlying container, runtime environment for demand, ensuring SLAs (Service-Level Agreements) for service execution.

3.4 Metadata Driven Data Warehouse

The diversity and complexity of data is a common problem in information systems, and various types of databases, including RDB, NoSQL DB, Time series DB, Graphic DB or In memory DB. Different storage technologies have their application materials, so it is important to properly direct the different types of storage technologies to different storage technologies.

The Metadata Driven Data Warehouse using metadata to define the data structure, process method. It is also performed to drive the data processing and normalization, and then guides the storage to correspond databases. The architecture of Metadata Driven Data Warehouse is as Figure 5 as follows:

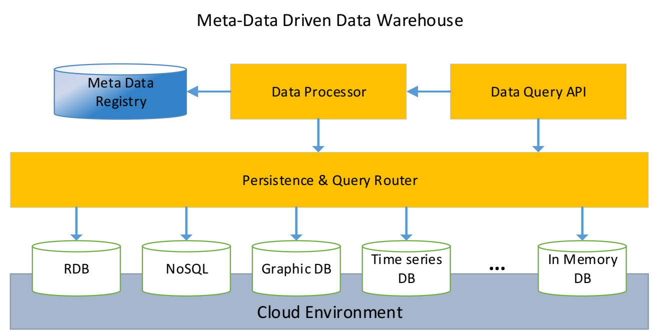


Figure 5. Meta-Data Driven Data Warehouse

The Metadata Driven Data Warehouse architecture consists of four sub-modules

- Metadata Registry

  It is responsible for the definition of Metadata model for data.

- Data Processor

  According to the Metdata model definition, the processing of resources and data normalization.

- Data Query API

  It provides a consistent and efficient data query API.
- Data Persistence Router

  Automatic routing to the appropriate storage database based on the characteristics of the data.

3.5 Deep Learning based Data Analysis

In the deep learning part of data analysis, the computing platform is a software and hardware integrated computing architecture. The hardware architecture includes CPU and GPU computing power. The software architecture includes decentralized computing, such as Hadoop and Spark. The parallel calculation section contains CUDA. The data storage includes MongoDB, HDFS, and HBASE, and the upper analysis tool includes deep learning computing frameworks. Before any data analysis is performed, it must be able to access the data effectively. We use Docker technology to install all analysis modules and the required software in a Container. In another, it's allowing application analysis tools to be automatically deployed when needed.

This Docker integrates Hadoop & Spark ecosystem and R, as well as a deep learning computing framework (including Tensorflow, Caffe, DL4J and other common frameworks). The entire software stack is shown in Figure 6 below. Figure 7 shows the data modeling process using deep learning.

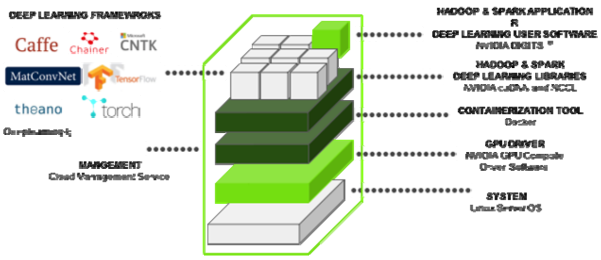


Figure 6. Software Stacks of Deep Learning

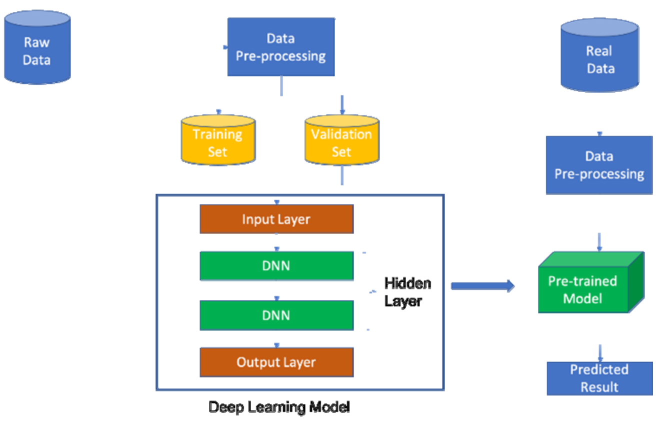


Figure 7. AI deep learning data modeling

Figure 8 to Figure 11 are the system screens of automatic data analysis using deep learning models. Figure 8 is the screen to select analysis model process. If there are no suitable analysis model process, user can modified form existing model or create a new model. Figure 9 and Figure 10 is the screens of set data training parameters. Figure 11 and Figure 12 are the screen of training result.

Figure 8. Select Analysis Model Process of Data Process Management

Figure 9. Setet Parameters of Analysis Process

Figure 10. Setup Deep Learning Parameters

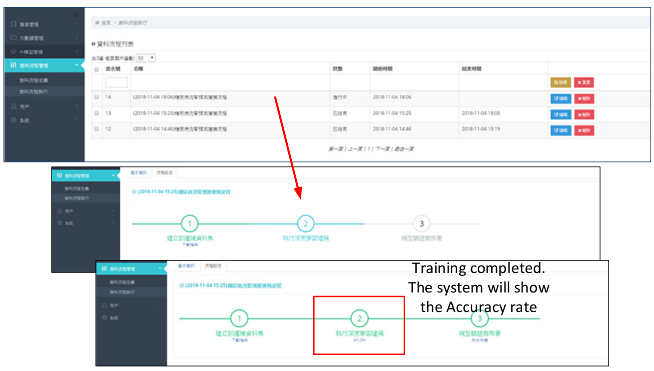


Figure 11. Data Training Results

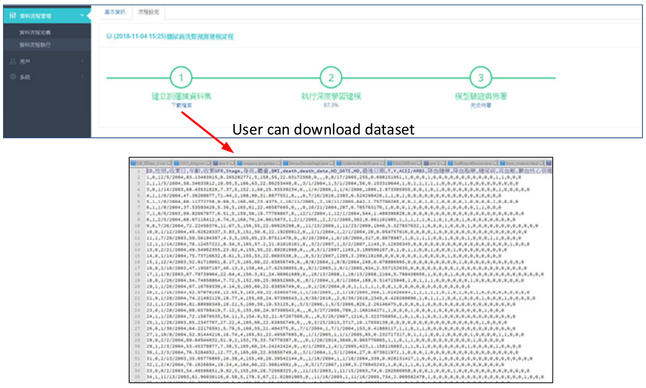


Figure 12. Download Dataset

4 Conclusions and Future Works

As the use of information technology and medical care becomes more and more widely used, and the cost of data storage and retrieval increases, more and more hospitals are expected to adopt cloud computing through medical systems that integrate IT technology. However, users or IT departments in hospitals still lack sufficient capabilities for the development and management of cloud-related big data platforms. Without an easy-to-use system architecture, faced with emerging information technologies such as cloud and big data, related users and managers are still unsuccessful.

In this paper, we propose a smart service warehousing platform which support big data deep learning modeling analysis. According by big data analysis and modeling using deep learning a based on the P4 Medicine model to predict and prevent users from chronic diseases. The possibility of early prevention and treatment to assist health care providers to reduce the burden of caring for chronic patients. This analysis results of the prototype has adapted to support some CKD research of Changhua Christian Hospital. This system has ongoing development with Changhua Christian Hospital for chronic kidney disease.

In the future work, we will continue to systematically develop and verify the integration and feasibility of the system. We also will cooperate with hospital to integrate the analysis and modeling of the existing database and medical data on chronic diseases and family medicine and apply it to chronic diseases to disease prevention analysis of medical health care systems.

Acknowledgments

This research was supported in part by Ministry of Science and Technology, Taiwan ROC, under grants number MOST 106-3114-E-029 -001, MOST 105-2221- E-164-006, MOST105-2622-E-126-002-CC, MOST 106-2221-E-126-012-MY2, and MOST108-2221-E-126- 002.

Dr. Ping-Fang Chiu has equal contribution as the corresponding author. Dr. Ping-Fang Chiu's email is 68505@cch.org.tw.

References

[1] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, D. Walker, P4: Programming Protocol-Independent Packet Processors, *ACM SIGCOMM Computer Communication Review*, Vol. 44, No. 3, pp. 87-95, July, 2014.

[2] P. Mell, T. Grance, *The NIST Definition of Cloud Computing*, NIST Special Publication 800-145, September, 2011.

[3] T. White, *Hadoop: The Definitive Guide*, Electronic book text, 3rd ed., O'Reilly Media, 2012.

[4] MIKE2.0, Big Data Definition, Retrieved from http://mike2.openmethodology.org/wiki/Big_Data_Definition

[5] N. V. Lakshmi, K. S. Rani, Privacy Preserving Association Rule Mining in Vertically Partitioned Databases, *International Journal of Computer Applications*, Vol. 39, No. 13, pp. 29-35, February, 2012.

[6] S. Peng, Q. Xu, X. B. Ling, X. Peng, W. Du, L. Chen, Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines, *FEBS letters*, Vol. 555, No. 2, pp. 358-362, December, 2003.

[7] I. W. Tsang, J. T. Kwok, P. M. Cheung, Core Vector Machines: Fast SVM Training on Very Large Data Sets, *Journal of Machine Learning Research*, Vol. 6, No. 13, pp. 363-392, 2005.

[8] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[9] R. O. Duda, P. E. Hart, D. G. Stock, *Pattern Classification*, 2nd Edition, Wiley, 2000.

[10] Y. Bengio, A. Courville, P. Vincent, Representation Learning: A Review and New Perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1798-1828, August, 2013.

[11] J. Schmidhuber, Deep Learning in Neural Networks: An Overview, *Neural Networks*, Vol. 61, pp. 85-117, January, 2015.

[12] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *Nature*, Vol. 521, No. 7553, pp. 436-444, May, 2015.

[13] D. Ciregan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 3642-3649.

[14] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *NIPS 2012: Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, 2012, pp. 1106-1114.

[15] J. Russell, Google's AlphaGo AI wins three-match series against the world's best Go player, *TechCrunch*, 25 May 2017.

[16] R. H. Campbell, M. Montanari, R. Farivar, A middleware for

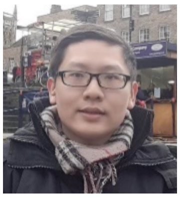
assured clouds, *Journal of Internet Services and Applications*, Vol. 3, No. 1, pp. 87-94, May, 2012.

[17] H. P. In, C. H. Kim, U. Yun, S. S. Yau, Q-MAR: A QoS Resource Conflict Identification Model for Situation-Aware Middleware, *Proceedings of the Ninth IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS'03)*, San Juan, Puerto Rico, USA, 2003, pp. 1-7.

[18] L. Chen, Microservices: Architecting for Continuous Delivery and DevOps, *The IEEE International Conference on Software Architecture (ICSA 2018)*, Seattle, WA, USA, 2018, pp. 39-46.

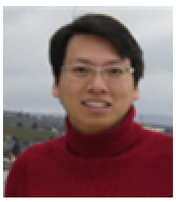
Biographies

Chih-Hung Chang received the M.S. and Ph.D. degrees in computer science from Feng Chia University, Taichung, Taiwan. He is an Associate Professor with the College of Computing and Informatics, Providence University, Taiwan. His research interests include software engineering, cloud service, and deep learning.



Tse-Chuan Hsu received the M.S. degrees in computer engineering from Tunghai University, Taiwan. received the Ph.D. degrees in BathSpa University Centre for Creative Computing for Ph.D degrees and an Assistant Professor with Department of Healthcare Information & Management, Ming Chuan University, Taiwan. His research interests include software engineering and cloud computing data analysis systems.

William Cheng-Chung Chu is a Distinguished Professor in the Department of Computer Science and the director of the Software Engineering at Tunghai University, Taiwan. Chu received a PhD in computer science from Northwestern University. He is a member of the IEEE Computer Society.



Che-Lun Hung received the M.S. degrees in computer science from Feng Chia University, Taichung, Taiwan. He received the Ph.D. degrees Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan. He is a Professor in Institute of Biomedical Informatics, National Yang-Ming University. His research interests include Genomics, Proteomics, Bioinformatics, Bio-medicine, GPU Programming and Applications, Parallel and Distributed Processing.

Ping-Fang Chiu is the Director of Department of Nephrology, Changhua Christian Hospital, Changhua, Taiwan. He is major in Acute and chronic renal failure, Diagnosis and treatment of hematuria and proteinuria, and Various kidney diseases.