Data System Configuration and Organizing and Management Technologies of Information System

Hua Wang, Haidong Meng, Lianwei Zhang, Zhengmao Chen and Shan Wang
Tianchen Information Technology and System Engineering Institute, Beijing, China
Email: 919wh_cn@sina.com

Abstract. Based upon the synopsis of data system and information system, this paper puts forward the data system configuration of information system and execution process of data organizing and management, and sets forth a series of related technologies, such as data warehouse construction technology, data governance technology, data training and learning technology, data syncretization and analysis technology, and data subscription and distribution technology, and moreover, the effectiveness analysis is carried out with a performance comparison between the prototype and the existed system of data organizing and management. This will benefit the further evolution of information system and the continual improvement of data engineering.

1. Introduction
Information system is a distributed man-machine system that effectively coordinates related resources, implements information processing and carries out business activities. With the rapid development of things internet, cloud computing and artificial intelligence technologies, data resources and data services are more and more important during the development of information system, and information organization models and service modes need to adapt to business information and decision activities. In order to further satisfy dynamic demands of system flexible availability and business decision, data collection, data management, data analysis, and data application technologies are widely investigated and applied; and a great deal of structured, semi-structured and unstructured data is organized by new methods into data system that has independent function and flexible availability, and provides precise trend prognostication and decision report for users to finely get hold of business rules and dominate activity courses [1][2].

2. Data System Configuration
Data system configuration of information system is mainly composed of primary and collection type, basic and support type, application and product type.

- Primary and collection type is the source data that aims at business requirements from various related units, and comprises business data, specialty data and system data.
- Basic and support type is the fundamental data that aims at business assignments, and comprises attribution data, situation data and synthesis data.
- Application and product type is the practical data that aims at system operation, and comprises system function data and departmental business data.

Considering the complexity, speciality and expansibility of data system, data collection, data management, data analysis and data application technologies are investigated to support the development and operation of information system, and data organizing and management plays an important role in the smooth running of data system and information system [3][4].
3. Data Organizing and Management Process

Data organizing and management is to carry out unitive data standardization storage, and construct data system by means of model and related facilities. The execution process of data organizing and management is as follows.

- Data pretreatment. The isochronous data is structured and pretreated;
- Data import. The pretreated data is tagged with time label, and automatically introduced into base database;
- Data handling. Object oriented wide tables are constructed in light of application scene, and extract data from primary data layer and fill in corresponding field;
- Data analysis. Index computation model is established on the basis of algorithm development and model training platform, and data excavation can be carried out so as to store up the outcome to respective database;
- Data utilization. The data is released in the form of services and utilized by applications.

4. Data Organizing and Management Technologies

Data organizing and management is mostly composed of data warehouse construction, data governance, data training and learning, data syncretization and analysis and data subscription and distribution [3][5][6]. Data organizing and management technologies of information system are depicted in figure 1.

![Figure 1. Data organizing and management technologies of information system.](image-url)
- Data warehouse construction is to implement data layered storage, and comprises normalized operation data, common dimensionality model, and application data service technologies.
- Data governance is to implement the establishment and management of data governance process and data warehouse model, and comprises data quality management, data standard management, and data authorization technologies.
- Data training and learning is to implement data computation and data excavation by means of data model and algorithm library in diverse layers, and comprises machine learning algorithm frame, basic algorithm development, and natural language processing technologies.
- Data syncretization and analysis is to implement open data analysis so as to provide data inquiry and data visualization for various business departments, and comprises multi-dimension data analysis, human-machine interaction exploration, and programmed data visualization technologies.
- Data subscription and distribution is to implement data service management, and provide authorization management of subscription and distribution, and comprises tiny service framework, service quality monitoring, and service authorization supervision technologies.

5. Data Warehouse Construction Technology

5.1. Normalized Operation Data
Normalized operation data is to save increment or full data image of data source, and add a field of synchronization batch at a time; it carries out unitive data extraction, and ETL interface to prepare data, so as to ensure coupling untying of data warehouse and business system, and maintain record of basic data to help trace data and inspect location.

5.2. Common Dimensionality Model
Common dimensionality model includes detail data of data warehouse that is formed by data cleaning from normalized operation data, and then summary data of data warehouse is formed by light and heavy summary. Related and similar data is combined by means of fine table and correlative design; unitive processing is implemented to establish statistic indexes of standard naming and unitive algorithm for upper data application and service; and logic and collective tables are constituted that has consistent dimensionality.

5.3. Application Data Service
Application data service aims at special business, sets up different data fair, and actualize practical data business over the common data model layer. It consists of individuation index processing, applicative data assembling, flexible tactic engine and module mechanism. Data storage, access and processing model is set up in light of principles, such as high cohesion and low coupling, separation of kernel model and extended model, and data roll back and standardization.

6. Data Governance Technology

6.1. Data Standard Management
Based upon dimensionality modelling theory, data standard management is to carry out naming rules of table and field, such as field type selection, table subarea mode and data storage position and mode, by means of bus matrix construction, partition and definition of data region, business process, dimensionality, time cycle, and derived index, and etc.

6.2. Data Quality Management
Data quality management is to adopt/ introduce related inspection measures on data cushion layer, data operation layer, common basis layer, and data application layer, so as to guarantee data correctness, integrality and pertinency during the nodes of data import, conversion and eduction, and the related feedback can be obtained about data quality problem via data quality inspection.
6.3. **Data Authorization**
Data authorization provides available and safe data access service, and consists of data safe setting and data authority supervision. Data safe setting afford table and field level setting, user classification and access authority, so as to reduce the risk of data gathering; data authority supervision is to ensure relevant user the proper data via data safety requirement definition, processing flow and safe procedure institution.

7. **Data Training and Learning Technology**

7.1. **Machine Learning Algorithm Frame**
Algorithm optimization and issue location can be carried through by modelling tensor, grads, weight and renewal value; and model effect is fully evaluated via distribution estimation and index analysis. Deep reinforced learning is also implemented in respect to the existed models.

7.2. **Basic Algorithm Development**
Basic algorithm development frame provides abundant MPI, PS, BSP programming support and data access interface, and incorporates in proper deep learning algorithms, so as to provide IDE+ visualized experiment console and algorithm customized development for users.

7.3. **Natural Language Processing**
Natural language processing service perform effective analysis and excavation of non-structured text, and resolve non-structured text into structured data, so that business knowledge atlas can be set up, and human-machine interaction and machine reading and comprehension be carried out. Natural language processing can afford new capabilities, such as participle service, speech label, deep text excavation, sensibility analysis, speech identification, key word survey and speech synthesis service.

8. **Data Syncretization and Analysis Technology**

8.1. **Multi-dimension Data Analysis**
Multi-dimension data analysis is to achieve the unitive data view and data exploration of many data sources of complex system via automation identification and relation analysis. It can provide metadata management and real time analysis, accomplish multi-dimension analysis and business exploration of big data, and meet the capability needs for multivariate data analysis and correlation data analysis.

8.2. **Human-machine Interaction Exploration**
Human-machine interaction exploration is to combine machine computation capability and man cognition ability via concise and natural mode, so as to obtain discernment for big data, achieve valuable information and knowledge, and accord with user requirements and system characteristics.

8.3. **Programmed Data Visualization**
Programmed data visualization is to sustain the protraction of slot, route, section, and multilayer cascade drawing, and actualize the presentation of dynamic information, so that users can carry out flexible layout on the basis of inner templates, and quickly implement data excavation and visualization display in accord with business requirements.

9. **Data Subscription and Distribution Technology**

9.1. **Tiny Service Framework**
A large and complex application is divided into many tiny services that are independently configured, maintained and extended to provide distributed applications. Tiny service framework helps in service administration, operation control, route tracing and groupware stability, and the fast establishment of applications system on the basis of distributed computation frame.
9.2. Service Quality Monitoring
Service quality monitoring is to incorporate and encapsulate groupware of data collection, message route way, real time computation, column storage, online report and etc. and carry out integrative end to end service real time monitoring; it mainly consist of log and data incepting, real time computation assignment, assignment arrangement and warning, and report forms customization.

9.3. Service Authorization Supervision
Service authorization supervision is to perform detailed purview control on service API with the assurance of safe accessing, and includes purview management, role management, user group management, authorization application and authorization examination; it stipulates transfer source, business range, transfer availability and transfer flow, so as to insure safe service sharing.

10. Effectiveness Analysis
Based on the above data architecture and technologies of information system, the prototype system of data organizing and management is developed, and related technology validation testing is implemented. The prototype data ordering time of the mixture of structured, semi-structured and unstructured data is averagely 19.1 minutes, and the existed system data ordering time of the structured data of a size is averagely 35.1 minutes; and moreover, the prototype field inquiry time of testing data set is averagely 1.611 seconds, and the existed system field inquiry time of the same data set is averagely 3.090 seconds. Thus the effectiveness is remarkably promoted, and the performance comparison between the prototype and the existed system is shown in table 1.

|     | The prototype data ordering time | The existed data ordering time | The prototype field inquiry time | The existed field inquiry time |
|-----|---------------------------------|--------------------------------|---------------------------------|-------------------------------|
| 1   | 22.8                            | 36.1                           | 1.766                           | 3.329                         |
| 2   | 20.6                            | 37.8                           | 1.503                           | 3.201                         |
| 3   | 18.7                            | 35.5                           | 1.609                           | 3.173                         |
| 4   | 16.1                            | 33.9                           | 1.362                           | 2.986                         |
| 5   | 17.3                            | 32.2                           | 1.815                           | 2.761                         |
| average | 19.1                          | 35.1                           | 1.611                           | 3.090                         |

11. Conclusions
During the development of information system, data configuration of information system should be effectively constituted, and the related technologies such as data warehouse construction, data governance, data training and learning, data syncretization and analysis, and data subscription and distribution technology. Moreover, we should further advance the construction of data organizing and management system, increase data service and decision capabilities, and strengthen system synthesis integration and joint operation tests, so that information system can continually fulfil dynamic and knowledgable requirements and arrive at satisfactory objective.

12. Acknowledgments
We are respectful to the reviewer and secretary for their guidance, understanding and encouragement.

13. References
[1] Li Jianhui, Shen Zhihong, Meng Xiaofeng. 2017 *Journal of Computer Research and Development* 54 235
[2] Liu Wei, Xia Cuijuan, Zhang Chunjing. 2013 *New Technology of Library and Information Service* 4 2
[3] Yi X. Liu F. Liu T. et al. 2014 *IEEE Network* 28 5
[4] Fan Dandan, Yu Huiyon, Liu Chen. 2018 *Electronic Measurement Technology* 10 50
[5] ZHENG X P, ZHANG L Y, PENG J, et al. 2014 Computer Measurement & Control 22 41
[6] Robak S, Franczyk B, Robak M. 2013 Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE p1215