

# A Data Sharing and Integration Technology for Heterogeneous Databases

Yongjie Zhu and Youcheng Li  
Information Management Center  
Xuchang University  
Xuchang, 461000, China

Received: July 3, 2021. Revised: December 19, 2021. Accepted: January 8, 2022. Published: January 10, 2022.

**Abstract**—For a long time, there are a large number of heterogeneous databases on the network, and their heterogeneity is manifested in many aspects. With the development of enterprise informatization and e-government, the system database of each department constitutes a real heterogeneous database framework with its independence and autonomy in the network system of many different functional departments. This paper will design information sharing between heterogeneous databases of network database system of many similar functional departments by using XML data model. The solution of data sharing between heterogeneous databases can accelerate the integration of information systems with departments and businesses as the core among enterprises, form a broader and more efficient organic whole, improve the speed of business processing, broaden business coverage, and strengthen cooperation and exchange among enterprises. In addition, heterogeneous database sharing can avoid the waste of data resources caused by the heterogeneity of database, and promote the availability rate of data resources. Due to the advantages of XML data model, the system has good scalability.

**Keywords**—Data Sharing, Data Integration, Heterogeneous databases, XML data model

## I. INTRODUCTION

Database integration is the basic condition to realize the sharing of information resources. Whether the information between different databases can be integrated without gap will directly affect the comprehensive utilization of data information resources[1-3]. The data integration provides a unified storage and management model and shields the independent operation for all kinds of heterogeneous data[4-5].

Therefore, the integrated heterogeneous data achieves the data sharing between heterogeneous data sources, and effectively utilizes the effective information between different data sets. However, there are various kinds of semantic heterogeneity and data heterogeneity between different databases[6-7], which leads to the unsuccessful integration of heterogeneous database in the matching process.

The academic research on heterogeneous data integration started relatively early in foreign countries with more in-depth researches[8]. However, the corresponding prototype system developed abroad is not suitable for large-scale in commercial field. Although the importance of information integration has been realized in recent years, the research depth is insufficient due to the late start. Most domestic researchers only improved foreign research results on the basis of the theory. The data integration system which could meet the requirements of domestic large-scale utilization hasn't been developed [9]. Many domestic system integration companies only use foreign data system software for simple information integration and conversion. The integrated data is relatively rough, which can't solve the problem of data conflict fundamentally, and the real data sharing is difficult to be achieved[10].

However, it is the key factor to solve the problem of database integration to describe the data semantic features and to find out the semantic attribute matching and entity matching between heterogeneous databases. Data integration is to realize data sharing among subsystems, effectively use resources and promote the overall performance. The ideal goal of data integration is to provide a single system image for end users in a distributed environment [11]. Single system image means that the interaction between subsystems must be transparent. Transparency is to hide the details of interaction from users, so that users can see each subsystem as a seamless integrated system. Therefore, in view of above problems, this paper will design information sharing between heterogeneous databases of network database system of many similar functional departments by using XML data model, which can avoid the waste of data resources caused by the heterogeneity of database, and improve the utilization rate of data resources.

## II. DESIGN OF HETEROGENEOUS DATABASE INTEGRATION SYSTEM

In the network system of many different functional departments, the system database of each department constitutes a real heterogeneous database framework with its independence and autonomy[12-13]. This paper will discuss the design and implementation of information sharing between heterogeneous databases of network database system of many

similar functional departments by using XML data model.

The premise of heterogeneous database information sharing is to assume that there are many heterogeneous, distributed and independent database systems [14]. Taking the government network database as an example, there may be many specific government systems in normal operation in actual work. Now what we need to do is to design a simple and easy scheme to realize information sharing.

In order to share information and distribute databases, a frame should be established on the basis of these heterogeneous database systems [15]. Users can access multiple heterogeneous database systems transparently only through one communication platform. At the same time, the database of each department's information system is an independent entity with relative independence and autonomy. They do not want external users to operate or control their own database at will for the security of their own department's database. That is to say, in the process of sharing, users only have query permission to the shared database, and have no modification function. The best way to realize the query function is to provide services for users through B / S mode. Therefore, we adopt the system structure of B / S three-tier model based on XML, and the structure diagram is shown in Figure 1.

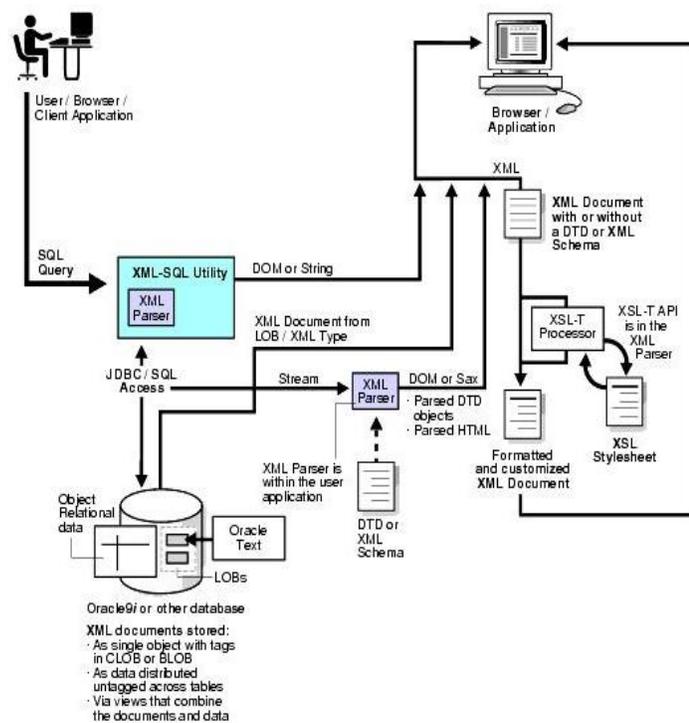


Fig. 1. Three layer structure of B / S based on XML

A Web service middle layer is established, which serves as a communication platform for information sharing. The middle layer uses ADO and other related data access technologies to connect with heterogeneous data sources[16]. In the client, users only need to have a Web browser to submit the corresponding query requirements through the Web, but the requirements are limited in scope, and the scope of query information is defined by the database administrator of each

department. According to the query requirements submitted by users, the Web service middle layer obtains data from heterogeneous databases, dynamically generates standard XML format documents, and then sends them to the client browser through the Web server, so as to meet the query requirements of users.

In the realization of data sharing of multiple heterogeneous databases, the management of shared data is very important, especially in the government network system, using the Web service middle layer to realize the connection of multiple heterogeneous data sources. However, due to the business particularity of each government network system, it is not suitable to disclose part of the Department's data in practical work. For example, data information of privacy nature is not allowed to spread freely through the network, but through data access technologies such as ADO, the Web server connects to the whole database system, not limited to a table or even some fields in a table in a database. If the client browser can query and master all the information of a certain data source through the web server, it may bring some losses and errors to the work of the data providing department, which the data providing department does not want to see[17]. The data providing department is willing to provide some tables or even some fields of a table in its system database for users of other departments to share. In view of the difficulty in solving this problem between Web Services middle tier and heterogeneous data sources, this paper proposes a solution to solve this problem by providing limited range query in Web Services middle tier.

In this scheme, because the management of the service layer is relatively simple, the amount of data is small, and the data relationship is not very complex, the web service middle layer establishes its own small database[18], and the administrators of various heterogeneous data sources submit the tables and tables that can be shared in the network system database of the Department through the management module provided by the web service layer. The specific fields in the table and the relationship between tables are saved in the database of service layer. When a user queries a database on the web through a browser on the client side, the web server queries the access database through APS, so as to determine the tables, fields and the relationships between tables that can be shared by the database, and generate the corresponding SQL statements according to these information. Then the generated SQL statements are submitted to the database through the database access interface protocol of the web service layer, so the data returned by the database is the local data defined by the administrator and can be shared. After returning the data, ASP and TDD are used to dynamically generate XML documents and related internal TDD, which are returned to the client browser in XML format to complete a complete data query.

### III. IMPLEMENTATION OF LOCAL DATA INTEGRATION IN HETEROGENEOUS DATABASE

In order to query the limited local data of heterogeneous data sources in the middle layer of Web services, we build a small database in the middle layer to store the shared tables, fields

and the relationship between tables of each data source. The information is submitted by the database administrator of heterogeneous data sources through web, The administrator users of heterogeneous data sources are pre-defined in the middle tier database of Web services, and the corresponding management users are assigned for each heterogeneous data source.

Establish a database administrator information table in the database of the service middle layer to record the user information that can use the management module. The information should include user name, user password, Department identification and other basic information, and other user information, such as user email, password retrieval verification information, etc. The Department ID corresponds to the actual department information. The structure design of administrator user information table is shown in Table 1.

Table 1. Structure design of administrator user information

Name	Code	Type	P	M
User name	UserID	varchar(15)	Yes	No
User password	Password	varchar(30)	No	No
Department ID	Department_ID	varchar(30)	No	No

The structure design of department information table is shown in Table 2.

Table 2. Structure design of department information

Name	Code	Type	P	M
Department ID	DepartMent_ID	varchar(15)	Yes	No
Department Name	Department_Name	varchar(30)	No	No

An administrator user corresponds to a department. The user can define the connection parameters of the Department's database and submit them to the database of the web service middle layer. Therefore, the connection parameter table of each heterogeneous data source should be set up in the web service middle layer. The structure design of the table is shown in Table 3.

Table 3. Structure design of connection parameter

Name	Code	Type	P	M
Department ID	DepartMent_ID	varchar(15)	Yes	No
Server name	Server_Name	varchar(20)	No	No
Database name	Database_Name	varchar(50)	No	No
Database user	Database_User	varchar(15)	No	No
Database password	DataBase_Pwd	varchar(30)	No	No
Service provider	Provider	varchar(40)	No	No

The connection parameter table provides the connection parameters of various heterogeneous data sources. The data

access protocol interface uses these connection parameters to connect to one or more heterogeneous data sources. Once the connection is established, the data can be read freely. In the process of database access, the access scope is limited. The limit of the scope is determined by the heterogeneous data source administrators according to the data security requirements of their own departments, It is realized by defining the data table structure that can be shared by the Department in the web service middle tier database. When defining these structures, it is possible that the whole table or some fields of a table can be shared externally. At the same time, it is also necessary to provide the relationship of external keyword links between tables. We build a shared information table to save the information of tables and fields that can be shared by heterogeneous databases. The table structure design of the shared information table is shown in Table 4.

Table 4. Structure design of shared information

Name	Code	Type	P	M
Department ID	DepartMent_ID	varchar(15)	Yes	Yes
Database name	Database_Name	varchar(50)	Yes	Yes
Table name	Tab_Name	varchar(20)	Yes	Yes
Table name (China)	Tab_Name_CN	varchar(20)	No	No
Column name	Column_Name	varchar(20)	Yes	Yes
Column name (China)	Column_Name_CN	varchar(20)	No	No

At the same time, an association information table is established to record the relationship between tables in heterogeneous databases. The structure design of association information table is shown in Table 5.

Table 5. Structure design of association information

Name	Code	Type	P	M
Department ID	DepartMent_ID	varchar(15)	Yes	Yes
Database name	Database_Name	varchar(50)	Yes	Yes
Table name	Tab_Name	varchar(20)	Yes	Yes
Key field (Foreign)	Foreign_Key	varchar(20)	Yes	Yes
Link table	Link_Tab	varchar(20)	No	No
Link column	Link_Column	varchar(20)	No	No

The basic design of database structure in Web Service middle layer can realize the limited access to heterogeneous database. When users query a data source, the related information recorded in the database is controlled by the user APS.NET It is used to dynamically generate relevant SQL data query statements, so as to limit the query scope to the database structure system defined by the database administrator.

#### IV. DATA STRUCTURE OF STORAGE MODEL OF HETEROGENEOUS DATABASE

Traditional methods improve the control performance and throughput performance of data systems by analyzing heterogeneous data structures and optimizing data storage algorithms. However, there is a problem of uneven spatial distribution of algorithm grid in existing databases. Some

related references put forward new improvement methods, but these results show that the data control performance and the robustness of the algorithm are not very good, and the storage overhead is too large

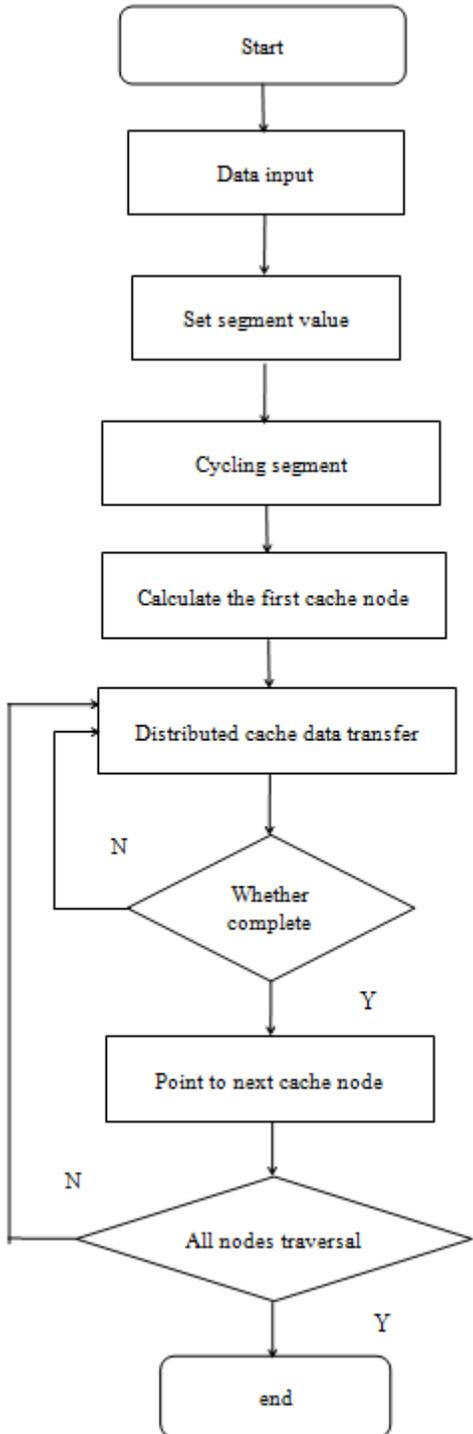


Fig.2 Cloud storage mode in information data structure

This paper proposed a novel optimization storage algorithm as cloud storage model in information data structure, which has the best adaptive matching of basic functions. Through the multi-mode data source storage mode information data structure, the user nodes in heterogeneous database adopt the

three-tier distribution mode. Heterogeneous databases distribute massive data for multiple service nodes, control the distribution of multi-source information, and store and manage information data in the database. The general establishment and implementation of cloud storage mode in information data structure are shown in Figure 2.

Sub-optimum regulatory factors in memory is in the formula (1).

$$\begin{aligned} \lambda(k) &= c(k), c(k) > 1 \\ \lambda(k) &= 1, c(k) \leq 1 \end{aligned} \quad (1)$$

In the space of characteristic data, control data is

$$c(k) = t[N(k)/t \cdot c(k)] \quad (2)$$

and it should be in the formula(3),

$$\begin{aligned} N(k) &= V_0(k) - \beta R(k) - \dot{H}(k) \dot{Q}(k-1) \dot{H}^T(k) \\ C(k) &= \dot{O}(k-1) P(k-1) \dot{O}^T(k-1) \dot{H}^T(k) \dot{H}(k) \end{aligned} \quad (3)$$

Among them,  $R(k)$  is a real positive definite symmetric matrix.  $R(k) = \text{diag}\{r_1(k), r_2(k), \dots, r_{nq}(k)\}$  is positive diagonal matrix.  $V_i$  stands for vector quantification characteristic of the  $i$  times data in storage structure.

Choose  $S = \sqrt{\Delta t / \Delta f}$ , and isolation between two data is  $\Delta x = \sqrt{\Delta t \Delta f}$ , which shows that separable state function in informative data is the formula (4),

$$X_p(u) = g'(u) \exp\left[-j\pi t^2 \tan\left(\frac{\alpha}{2}\right)\right] \quad (4)$$

Further designing separable dynamic scheduler of data storage based on above data structure analysis, if  $x_t^r = k_t^r, x_t := x_t + x_t^r, X := X \cup \{x_t\}, \gamma_{it} = b_t \frac{c_{it} \delta_{ir}}{d_i}$ .

Column vector composed by discrete sample in informative data structure of Heterogeneous database is the formula (5),

$$\begin{aligned} \bar{x} &= D \left[ x\left(\frac{1}{\Delta x}\right), x\left(\frac{2}{\Delta x}\right), \dots, x\left(\frac{N}{\Delta x}\right) \right]^T \\ \bar{X}_p &= J \left[ X_p\left(\frac{1}{\Delta x}\right), X_p\left(\frac{2}{\Delta x}\right), \dots, X_p\left(\frac{N}{\Delta x}\right) \right]^T \end{aligned} \quad (5)$$

Among them,  $D$  and  $J$  stand for interpolation and extracting matrix in informative data structure of Heterogeneous database. Through above method, it lays a foundation for optimum storage with the help of analyzing time scale of informative data stream in informative and integrating data structure of Heterogeneous database.

The index set of basic function in informative data stream is the formula (6),

$$\Lambda_0 = \beta \in \Gamma : |\langle f, d_{\gamma_0} \rangle| \geq \alpha \cdot \sup_{\gamma \in \Gamma} |\langle f, d_{\gamma} \rangle| \quad (6)$$

Vector sets AHP - PSO scheduling equals to  $\alpha$  in formula (6), and  $\beta$  is exactly the storage state index  $\gamma_0$  of data by using projection matching to have a dimension reduction for the load of storage cost, and choosing certain basic function to have a system matching. Storage distributed space determined by estimated value of Doppler frequency is the formula(7),

$$\lambda^n(d_{\gamma_0}) = \int_{-\infty}^{+\infty} f(t) \otimes d_{\gamma_0}^*(t) dt \quad (7)$$

It can realize optimum storage of data structure by searching optimum storage space projection based on optimum basic function adaptive matching before projecting each time.

## V. DISCUSSION

### A. Generating XML pages with ASP.NET

The transformation from relational data source to XML data source is generally generated through asp.net, rather than statically creating XML files. By using asp.net, the above relational data table can be converted into XML data source (XML document or dynamic XML page). At this time, the data source is ASP file. Just like the general ASP file output HTML page, the ASP file output at this time is the XML data source.

The generated XML data source is also updated automatically. In this way, only through the data maintenance and management of different database management systems, the user query results can be consistent with the data in the current database.

The key to what type of dynamic page is to set the content type property of the response object in asp.net. When the value of content type attribute is set to "text / HTML", the generated dynamic page is HTML page. When it is set to "text / XML", the generated dynamic XML page is dynamic XML page. By default, the value of the content type attribute is "text / HTML".

When using asp.net to dynamically generate an XML page, all text of the XML page should be generated dynamically. We can use the response object of ASP.NET to generate each statement of the page, such as

```

"<author><lastname>MINWEI</lastname></author>"
response.write "<title>Fundamentals of Computer</title>"
response.write "<publlsher><name>Machinery Industry
Press
</name></Publisher>"
response.write "</book>"
response.write "<book year= "1998" >"
response.write
"<author><lastname>LEICHENG</lastname></author>"
response.write
"<author><lastname>XUANSHI</lastnmae></author>"
response.write "<title>Computer Principle</title>"
response.write "<publisher><name> Electronic Industry
Press
    
```

```

</name></publisher>"
response.write "<book>"
resPonse.write "</lib>"
%>
    
```

In the process of data query on heterogeneous databases, the system mainly dynamically generates SQL statements for query according to the shared data structure of the system database of the Department defined by the administrator of each heterogeneous data source, so it is easy to know the specific data in the query results.

### B. Embodiment of XML Document in Client Query Document

The structure and content of XML documents are separated from display control. Therefore, the application of XML also includes the control methods of XML document display. There are two main methods to control XML document display: cascading style sheet (CSS) and extensible language (XSL).

Cascading style sheet (CSS) provides a technical means to make the display of web pages more attractive in the browser, without adding standard annotations to control the display as frequently as HTML documents. XML allows any custom tag to be used in the document to organize the information controlling the display of XML document elements into the form of style sheet.

In the heterogeneous database information sharing system, after the heterogeneous database data is transformed into XML documents, the client display also needs corresponding cascading style sheets to restrict the output format. In the process of dynamically generating XML documents, the generation of tags is uncertain. Therefore, defining the display format of each tag in CSS has become a difficult problem. However, during the generation of XML document, the root element tag is fixed (set as result in the front), and the CSS style has the feature that the child elements automatically inherit the attributes of .

## VI. RESULT ANALYSIS

The statistical efficiency of using this model to optimize the quality of data sharing and integration is higher than that of the traditional analysis model. Using this model and the traditional analysis model, the statistical experiments of data sharing and integration were carried out. Compare the statistical stability (%) of data sharing and integration with the two models. The comparison results are shown in Figure 3.

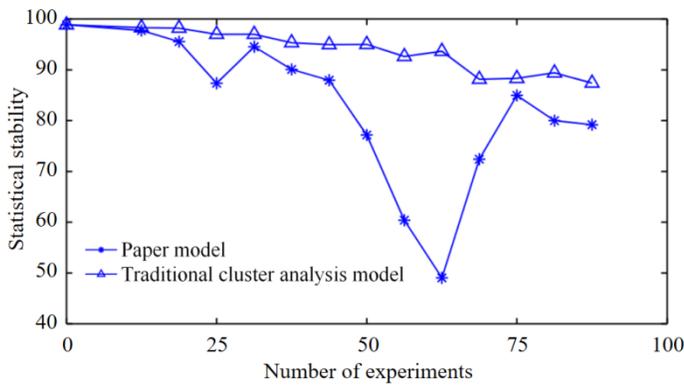


Fig. 3. Comparison of statistical stability of different models

It can be seen from the figure that the statistical stability of data sharing and integration by using this model is better than that of the traditional model. Using this model and the traditional analysis model, the statistical experiments of optimizing the quality of data sharing and integration are carried out.

The influence of different integration scales on the accuracy of data sharing and integration is shown in Figure 4.

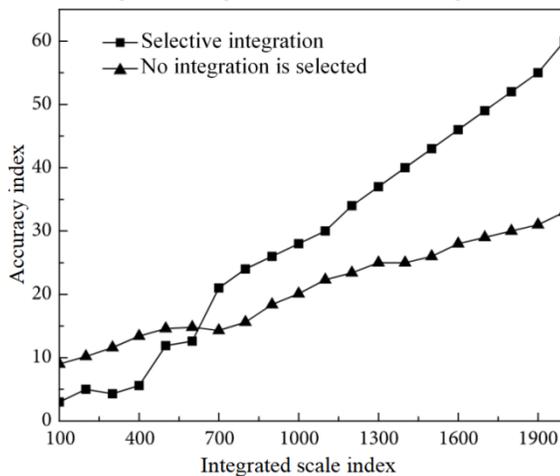


Fig. 4. The impact of different integration scales on the accuracy

Through the simulation test of several groups of data, it can be seen that the error between the predicted results of each group and the true values is very small, and the fitting degree of the input data basically meets the required requirements.

## VII. CONCLUSION

In view of the problems of data exchange and departmental coordination among multiple government network systems in most cities, this paper puts forward a data co implementation scheme which makes use of existing network resources, low investment construction and easy popularization, and introduces the relevant technologies involved in the implementation of the scheme. The scheme adopts B / S three-tier architecture mode and XML based system design. Users can access and query heterogeneous databases only by web application (from web browser) on the client. In the design and implementation of the scheme, four key problems are mainly solved:

(1) A resource information directory composed of shared information of heterogeneous data sources is dynamically generated on the web middle tier server, which makes the heterogeneous database a completely transparent global database relative to users.

(2) The administrator of each heterogeneous data source defines the shared data range of the Department's database to dynamically generate relevant SQL query statements, so as to realize the local data access to the heterogeneous database, so as to ensure the security of the business data of each heterogeneous data source.

(3) Through web, the administrator of each heterogeneous data source can easily define and change the range setting of database sharing information and the setting of database connection parameters.

(4) According to the characteristics of database structure and the user-defined characteristics of XML tag elements, the field names of database table structure are dynamically mapped into tags in XML documents to realize the transformation from different types of relational database data to unified format XML model data.

The limitations of this paper and the following research will be done in the future: in addition to CSS constraints, the display of XML documents on the client can also use the extended style language XSL, which makes the display definition of XN more complex and more powerful. This article only briefly discusses the use of CSS and more complex style definitions. Using XSL to define the display format can achieve better display effect. The design of displaying query results in XML format itself includes the idea of further Web Query in the results. XML document is the best method to support web page query at present; After the XML document is generated and successfully displayed on the client, this paper does not do further web page query on the client. The web query of results should be further considered in the heterogeneous database data sharing system discussed in this paper.

## REFERENCES

- [1] Asghari A , Sohrabi M K , Yaghmaee F . Online scheduling of dependent tasks of cloud's workflows to enhance resource utilization and reduce the makespan using multiple reinforcement learning-based agents. *Soft Computing*, 2020, 24, pp. 1-23.
- [2] Prasad V K . MD Bhavsar. Monitoring IaaS Cloud for Healthcare Systems: Healthcare Information Management and Cloud Resources Utilization. *International Journal of E-Health and Medical Communications*, 2020, 11, pp. 122-131.
- [3] Skrami E , Carle F , Villani S , et al. Availability of Real-World Data in Italy: A Tool to Navigate Regional Healthcare Utilization Databases. *International Journal of Environmental Research and Public Health*, 2020, 17, pp. 65-72.
- [4] Xu X . Material database management system based on heterogeneous multi-processor and computer embedded system. *Microprocessors and Microsystems*, 2021, 82, pp.103926.
- [5] Asfand-E-Yar M, Ali R. Semantic Integration of Heterogeneous Databases of Same Domain Using Ontology. *IEEE Access*, 2020, 8, pp. 77903-77919.
- [6] Mollero R , X Pennec, Delingette H , et al. Population-based priors in cardiac model personalisation for consistent parameter estimation in heterogeneous databases. *Communications in Numerical Methods in Engineering*, 2019, 35(2), pp. e3158.1-e3158.25.

- [7] Yu H K , Min B L , Nam S H , et al. Enhancing the Accuracies of Age Estimation with Heterogeneous Databases Using Modified CycleGAN. *IEEE Access*, 2019, 99, pp.1-12.
- [8] Yanni Zhao, Hualei Guo. Research on Heterogeneous Database Migration Technology Based on XML. *Computer and Digital Engineering*, 2018, 46, pp.129-133.
- [9] Yan Z, Chen X, Tang X. A Novel Linear Model Based on Code Approximation for GNSS/INS Ultra-Tight Integration System. *Sensors*, 2020, 20, pp. 3192.
- [10] Aa A , Fm A , Kg A , et al. XRepo - Towards an information system for prognostics and health management analysis. *Procedia Manufacturing*, 2020, 42, pp. 146-153.
- [11] Sousa M , Pires R , Del-Moral-Hernandez E . SOMprocessor: A high throughput FPGA-based architecture for implementing Self-Organizing Maps and its application to video processing. *Neural Networks*, 2020, 125, pp. 349-362.
- [12] Junior P O, Conte S, D 'Addona D M, et al. An improved impedance-based damage classification using Self-Organizing Maps. *Procedia CIRP*, 2020, 88, pp. 330-334.
- [13] J Feng, Wang F, Wang Q , et al. Intraseasonal variability of the equatorial Pacific Ocean and its relationship with ENSO based on Self-Organizing Maps analysis. *Journal of Oceanology and Limnology*, 2020, 38, pp. 1108-1122.
- [14] Oleg Milder, Dmitry Tarasov, Andrey Tyagunov, The Artificial Neural Network Structure Selection Algorithm in the Direct Task of Spectral Reflection Prediction, *WSEAS Transactions on Systems and Control*, Volume 14, 2019, Art. #9, pp. 65-70.
- [15] Ashkan Tashk, Jrgen Herp, Esmacel Nadimi, Automatic Segmentation of Colorectal Polyps based on a Novel and Innovative Convolutional Neural Network Approach, *WSEAS Transactions on Systems and Control*, Volume 14, 2019, Art. #47, pp. 384-391.
- [16] Fan Z, Alley A, Ghaffari K, et al. MetFID: artificial neural network-based compound fingerprint prediction for metabolite annotation. *Metabolomics*, 2020, 16, pp. 104.
- [17] Wang J , Zhao Z , Liu Y , et al. Research on the Role of Influencing Factors on Hotel Customer Satisfaction Based on BP Neural Network and Text Mining. *Information (Switzerland)*, 2021, 12, pp. 99.
- [18] Wu L, J Zhou, Li Z. Applying of GA-BP Neural Network in the Land Ecological Security Evaluation. *IAENG Internaitonal Journal of Computer Science*, 2020, 47, 11-18.
- [19] Panda S , Panda G . Performance Evaluation of a New BP Algorithm for a Modified Artificial Neural Network. *Neural Processing Letters*, 2020, 51, pp. 330-337.
- [20] Suboh Alkhushayni, Taeyoung Choi, DuAlzaleq, Data Analysis using Representation Theory and Clustering Algorithms, *WSEAS Transactions on Computers*, Volume 19, 2020, Art. #38, pp. 310-320.
- [21] Stella Vetova, Workflow Model for Big Data Analysis and Visualization, *International Journal of Applied Mathematics, Computational Science and Systems Engineering*, Volume 3, 2021, pp. 99-106.

**Creative Commons Attribution License 4.0  
(Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)