Research on Risk Evaluation of Human Resource Management Outsourcing Based on Data Driven

Tao Liu

7891950@qq.com

School of Business Administration, Wuhan Business University, Wuhan, China

Abstract—This paper proposes a set of data-driven human resources information system architecture, and specifically clarifies the technical implementation method of each layer in the architecture, the program flow of the system and the management of employee basic data, employee information management, reward and punishment management, salary management, service The setting, function and key technologies in development of five subsystems including management have been deeply analyzed and studied. The weight coefficient of each index is determined by the Analytic Hierarchy Process (AHP), the cloud model is introduced, and the quantitative and qualitative concepts are converted to each other using the forward and reverse cloud generator.

Keywords- Human resource management; information systems; digital transformation; outsourcing risk; analytic hierarchy process; cloud model

1 INTRODUCTION

With the rapid development of cloud computing technology, big data technology, artificial intelligence technology and Internet of Things technology, in order to adapt to the development of the times and the wave of information development, enterprises must carry out digital transformation. As an important part of enterprise management, enterprise human resource management has the core demands of digital transformation: strengthening business empowerment, quick and agile response, accelerating service transformation, vigorously cultivating talents, and transforming its own functions [1]. Under this demand, improving the intelligence level of the enterprise's human resources system, integrating cutting-edge information technology into the human resources management information system, and rapidly building a human resources system that adapts to the current enterprise development have become the top priority of the enterprise, and it is also the enterprise's top priority. One of the key points of digital transformation [2].

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322783

2 HUMAN RESOURCE INFORMATION SYSTEM

2.1 Decentralized construction architecture

The essence of “decentralization” is an open and shared idea. The current human resource system is designed and implemented from the perspective of the management. With the development of microservice technology, the application of human resource information system can be realized. Loosely coupled, each application realizes complete service and componentization [3]. Based on this, enterprises can also realize the digital transformation of decentralized organizational structure. Based on microservices and low-code technology, enterprises can provide employees with personalized services. In this way, the intentions and feelings of employees are analyzed, combined with HR’s operational data, to achieve a better decentralized human resource management transformation [4].

2.2 Data Model Analysis

The current enterprise human resources system mainly has two construction and deployment modes. One is the traditional e-HR software, which is commonly found in large enterprises. It has carried out informatization earlier and has a set of human resources systems with relatively complete functions [5]. The current problems are: The software functions are outdated and it is difficult to upgrade and iterate. The other is to deploy lightweight HR SaaS products, which are common in small and medium-sized enterprises. However, due to the many human resource management modules and complex systems, vertical SaaS deployment is difficult to get through the underlying data. A “chimney-style” independent information system, and at the same time, the greater the difference in the service methods of different SaaS service providers, the lower the standardization level and the greater the uncertainty. In the future, the new human resource system will adopt a data integration model, based on data warehouse technology, integrate multi-source heterogeneous data of various vertical systems, perform hierarchical data processing, and perform labelled and structured hierarchical storage of data[6].The human resources system built in this way deploys new applications in the form of micro-services during upgrade iterations, without the need to build data interfaces and transmission channels with other applications and systems, which greatly reduces development costs, and the construction of data warehouses is also realized. The data of different modules is fully connected to prevent the formation of a “chimney-like” business system.

3 DESIGN AND RESEARCH OF DATA-DRIVEN HUMAN RESOURCE INFORMATION SYSTEM

According to the above-mentioned development status and future development direction of the human resources system, this paper proposes a new type of human resources system architecture as shown in Figure 1. Based on emerging technologies and centered on the internal needs of enterprises’ digital transformation, it forms A data-driven human resource system has been developed.

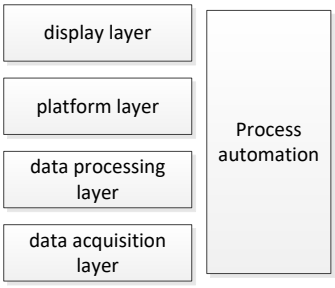

Figure 1 Data-driven human resource information system architecture diagram

3.1 Data acquisition layer

Data is one of the most important assets of an enterprise. The purpose of the data acquisition layer is to obtain all data related to human resources of the enterprise, including local data, company intranet data, and third-party platforms, covering all applications of the human resources system. , and other system data related to the Human Resource Information System. Local data is mainly data assets on the company's local deployment server, such as personnel information, salary information, organizational personnel and other locally stored structured data. Intranet data includes relevant data uploaded by employees who connect to the company's intranet through VPN, as well as data of each branch, for example, in the retail industry, the acquisition of data such as personnel information and attendance information of each store. Third-party data, mainly data stored by enterprises on third-party platforms, such as data stored on public cloud platforms after purchasing SaaS services, and external data obtained from third-party platforms, such as those obtained from third-party recruitment platforms CV, etc. In terms of data acquisition, some domestic scholars have learned from foreign data standards in related fields, and tried to explore research such as data quality standards and metadata standards. Combined with the existing research results, the following dimensions should be considered when acquiring human resources-related data. Comprehensiveness of data acquisition: First, the caliber of the data should be consistent.

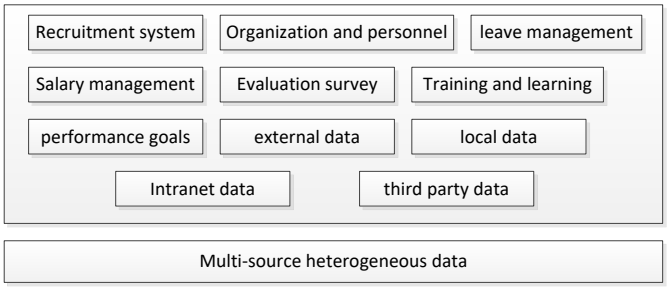

Figure 2 Data acquisition layer

Every time the data from the same source should be obtained, the statistical method and scope should be consistent. Second, the data integrity and validity should be maintained. Third, the obtained data should be timely. , formulate different acquisition strategies for data from

different sources and applications, and acquire data with high urgency in time. Data acquisition layer as shown in Figure 2.

3.2 Data processing layer

The bottom layer of the data processing layer is composed of a data warehouse server cluster, which collects, cleans and transforms data from multiple heterogeneous data sources through the ETL (extraction, transformation and loading) process. Research shows that in order to better complete the data cleaning process, certain To combine the knowledge of specific application fields, the data acquisition layer of this architecture directly acquires data from various business systems for data cleaning, and uses the data logic of existing business systems for ETL.

The business of human resource information system is complex, data sources are diverse, and data modalities vary widely, such as relational data, XML data, graph data, stream data, scalar data, vector data, data warehouses gather data from different sources into a single A centralized, centralized, and consistent data store to support data analytics, data mining, artificial intelligence, and machine learning. Data warehousing systems enable organizations to perform powerful analysis on petabytes of historical data that standard databases cannot. A data warehouse centralizes data from various data sources to cleanse, eliminate duplication, and normalize data to create a single source of data. This will give decision makers faster insight into the nature of the business. Because data from disparate sources limits decision makers' ability to confidently formulate business strategies, data warehouses support data integration, allowing users to leverage all of a company's data in every business decision.

A data mart is a subset of a data warehouse that focuses on a specific line of business, department, or subject area. Data marts make specific data available to a defined set of users, which allows those users to quickly access key insights without wasting time searching the entire data warehouse. The construction of the data mart takes a bottom-up analysis model with the aim of identifying all potential star schemas, starting with each Entity-Relationship (ER) model and performing an exhaustive analysis to continuously discover candidate entities, using these entities as the center nodes, which generate a large number of directed graphs. Each graph corresponds to an underlying star schema, which consists of a fact entity and all possible dimensions around it. The subsystems of the human resources system will have their corresponding data marts, such as compensation systems, recruitment systems, and so on. Data marts are designed to meet specific needs and their data topics are relatively narrow. While the data mart can still contain millions of records, its goal is to provide business users with the most relevant data in the shortest amount of time. Data marts simplify data access and increase efficiency, allowing users to leverage targeted data insights to achieve their specific goals. Businesses can benefit from accelerated business processes and higher productivity as users identify and extract valuable data in less time. At the same time, the maintenance of the data mart is also easier. Compared with the huge data warehouse, the deployment and operation and maintenance of the data mart are also more efficient. When building a data mart, first of all, it is necessary to clarify the needs of use, that is, the corresponding business line, determine the data source, then construct data subsets and subdivisions, and finally determine the relationship between the data mart and a larger-scale data set. Association to determine the deployment logic. Data processing layer as shown in Figure 3.

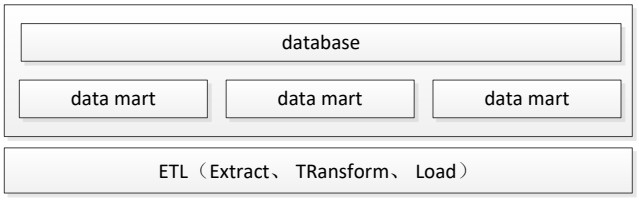

Figure 3 Data processing layer

3.3 Platform layer

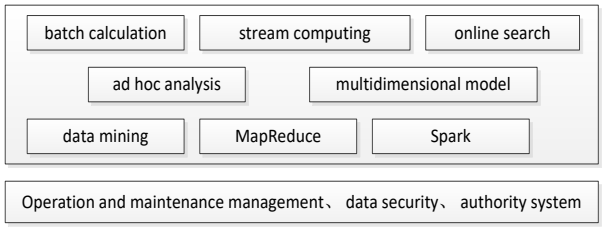

Figure 4 Platform layer

More in-depth data analysis requires the use of data mining technology to achieve some high-level data analysis requirements. The platform layer provides reliable analysis components for the analysis and utilization of human resource data, mainly including data that integrates multiple data mining algorithms. The platform layer as shown in Figure 4.Model creation and visual display functions mainly include classification algorithms and clustering algorithms. Classification algorithms include hierarchical clustering algorithms, segmentation clustering algorithms, constraint-based clustering algorithms, clustering algorithms in machine learning, and clustering algorithms for high Clustering algorithm for dimensional data…, the classification algorithm mainly includes decision tree classification algorithm, Bayes classification algorithm, association rule-based classification algorithm, database-based classification algorithm, and provides two main big data processing engines, Spark and MapReduce. The use of data processing engines can extract important value from huge amounts of data with low value density. The two data processing engines face different data processing needs. Spark is performed in memory, which is faster, while MapReduce is performed in disk. Processing of large amounts of data.

It will take longer. Batch computing is to decompose a large-scale data set into multiple small-scale data sets, and then distribute the same computing tasks to each small cluster, realizing mobile computing power instead of mobile data, greatly reducing the workload of each part, and parallel computing, which improves the computational efficiency. Stream computing mainly performs real-time computing and processing on streaming data, and can process and display incoming real-time data in real time. For example, during the online written test, through the real-time image analysis and processing of the video stream, abnormal behaviors can be found to ensure the fairness and impartiality of the test. Online query not only needs to deal with large-scale data result sets, but also provides some fast computing capabilities, such as online retrieval, conditional filtering and filtering, which can quickly filter and retrieve

information from a large number of data sets, and has high concurrency and low latency. Features. These capabilities are not available in batch computing and stream computing.

OLAP analysis refers to the multi-dimensional cross-analysis and processing of data in the face of large-scale data sets. The multi-dimensional data model used by OLAP includes two parts: dimension and fact. Its core technology is to find the mapping relationship between dimensions and facts, and to perform data association. , most of the aggregated associations, such as clustering, counting, summation, etc., are different from online queries, which are usually known and set in system design and implementation, while ad hoc analysis is usually in relational Real-time multi-dimensional modeling is performed in the data warehouse to realize the response to the temporary data demand generated by the user.

3.4 Display layer

The presentation layer provides a variety of data visualization tools, and Harvard Business Review divides data visualization into four main purposes: idea generation, idea presentation, visual discovery, and daily data analysis. Information visualization is the application of visualization technology in the field of non-spatial data, which can enhance the data presentation effect and allow users to observe and browse data in an intuitive and interactive way, so as to discover hidden features, relationships and patterns in data. According to the data type and different business requirements, the data is presented in the best way. For example, the employee roster, salary schedule, etc. are presented in the form of statistical reports. The analysis indicators are displayed in the form of disk, and the analysis of decision-making levels such as organizational effectiveness, organizational develop-ment, and human kinetic energy is displayed in the form of free reports that can be designed by themselves. Comprehensive evaluation model based on cloud model

3.5 Model description

Cloud Definition. Suppose a set of quantitative universe U, the qualitative concept on U is represented by C, if the quantitative value X is a random realization of C, and X $\in$ U, the degree of certainty $\mu(x)\epsilon[0,1]$ of X to C is a random number with a tendency to be stable, see formula (1):

$$\mu: U \to [0,1] \;\; x\epsilon\mu \qquad x \to \mu(x) \tag{1}$$

Then the cloud can be defined as the distribution of x on the universe U, and each cloud droplet is represented by $(x, \mu(x))$.

Digital features of clouds. Generally, the cloud model concept is represented by three digital features of expectation E_x, entropy E_n and super entropy H_e. These numerical characteristics of the cloud reflect the overall characteristics of the concept and belong to the overall qualitative characteristics of the qualitative concept.

Expectation E_x: It represents the expectation of the distribution of cloud droplets on u, which is the point that can best represent the qualitative concept.

Entropy E_n: used to measure the uncertainty of qualitative concepts.

Hyperentropy H_e: It is used to measure the uncertainty of entropy.

3.6 Cloud generator

The cloud model generation algorithm that is modularized by software or solidified by hardware is called cloud generator. The connection between qualitative and quantitative is mainly established through cloud generator. Qualitative and quantitative are interdependent. Cloud generators mainly include: forward cloud generator, reverse cloud generator, X condition cloud generator and Y condition cloud generator, etc. The forward and reverse cloud generators are mainly introduced here.

According to the three digital characteristics of the cloud: expected Ex, entropy En and super entropy He (Ex, En, He), cloud droplets $\mathrm{drop}(X_i, \mu_i)$ are generated, which are called forward cloud generators. Given a set of cloud droplets $\mathrm{drop}(X_i, \mu_i)$ conforming to a normal distribution as a sample, three numerical features (Ex, En, He) of the qualitative concept corresponding to the cloud model are generated, which is called a reverse cloud generator. The principle is shown in Figure 5.

Figure 5 Cloud Generator

The actual data is processed by the reverse cloud generator, and the single-factor cloud model and its digital characteristics are obtained. The calculation process is as follows:

Input index: sample point Xi (i=1, 2, ···, n).

Output metrics: Three numerical features (E_x, E_n, H_e) of the cloud model.

The calculation formula is:

According to each sample data x; calculate its average $\mathrm{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$,First-order sample absolute center distance $\frac{1}{n}\sum_{i=1}^{n} |x_i - X|$, and sample variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - X)^2$.

Using the comprehensive algorithm in the cloud model, the evaluation cloud of each index is integrated, and the comprehensive evaluation cloud of the human resource management outsourcing risk of logistics enterprises is obtained. The calculation process is as follows, ,see formula (2-4):

$$E_x = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{E_{x_1}\omega_1 + E_{x_2}\omega_2 + \cdots + E_{x_n}\omega_n}{\omega_1 + \omega_2 + \cdots + \omega_n} \tag{2}$$

$$E_n = \sqrt{\frac{\pi}{2}} \times \frac{1}{n} \sum_{i=1}^{n} |x_i - E_x|$$
$$= \frac{\omega_1^2}{\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2} E_{n_1} + \frac{\omega_2^2}{\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2} E_{n_2} + \cdots + \frac{\omega_n^2}{\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2} E_{n_n} \quad (3)$$

$$H_e = \sqrt{S - E_n^2} = \frac{\omega_1^2}{\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2} H_{e_1} + \frac{\omega_2^2}{\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2} H_{e_2} + \cdots + \frac{\omega_n^2}{\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2} H_{e_n} \quad (4)$$

4 CONCLUSION

In view of the shortcomings of the current human resources information system, this paper proposes a new data-driven human resources information system construction framework, and provides a full-link data processing solution from the acquisition of raw data to the display of statistical analysis reports. It is a human resources information system construction solution that meets the needs of digital transformation of enterprises. It can realize the integration and digitization of human resources business in all scenarios, and provide an excellent system foundation and digital base for human resources analysis.

Acknowledgment: "Human Resource Management" of the Core Curriculum Construction Project for Undergraduate Majors of Wuhan Business University in 2021

REFERENCES

[1] Takao Kato,Hideo Owan.Market Characteristics,IntraFirm Coordination, and The Choice of Human Resource Management Systems:Theory and Evidence.Journal of Economic Behavior Organization, 2011, 80(03): 375-396.

[2] S.C. Kundu,Rajesh Kadian.Applications of HRIS in Human Resource Management in India.European Journal of Business and Management, 2012, 4(21):58-60.

[3] Achola A.Management Information System Design on Human Resource Management of Kampala International.Information and Knowledge Management, 2013, 3(06):22-27.

[4] Ali Hadian.Comparing Electronic Human Resource Management Systems Efficiency In Production Organization Amp Service Organizations.International Journal of Scientific:Technology Research, 2014, 4(08):53-56.

[5] Gurhan Uysal. Impact of Human Resource Management on Economy.Chinese BusinessReview, 2015, 14(12):86-92.

[6] Lourdes Souto-Anido, Irene García-Rondón.Human Resource Management Model Based on Fuzzy Subsets Theory.Gecontec:Revista Internacional de Gestión delConocimiento y la Tecnología 2016, 14(02):14-34.