

# Data Analysis of EMS Sorting System on the Basics of Graph Database Neo4j

Rui Wang

Shandong University of Finance and Economics  
Jinnan, China

**Abstract**—With the rapid development of the express delivery industry, competition among industries is becoming more and more fierce. As an important part of the logistics process, sorting is facing severe pressure. Therefore, using data analysis to improve the efficiency of distribution centers has become a hot topic in current research. This paper firstly introduces the graphic database neo4j from the aspects of data storage, and then briefly describes its function for the distribution center. Finally, the orders of Jinan EMS distribution center are taken as the research object, and the graphic database neo4j is fully utilized to display the EMS shipment. The order status of the distribution center is conducive to the order sorting of the distribution center and it improves the sorting efficiency of the distribution center.

**Keywords**—neo4j; sorting center; sorting system; order analysis

## I. INTRODUCTION

At the same time of rapid economic development, the logistics industry is showing more and more vitality. Rapidly changing consumer demand continues to drive the development of the express delivery industry, and the impact of the express delivery industry on people's daily lives and the national economy cannot be ignored. Among them, express sorting is an important part of express circulation, which has reached 20% of logistics cost. If we can start from the quick sorting process and effectively improve the sorting efficiency of express mail, it will bring considerable economic benefits to the enterprise.

The main focus of this paper is to improve the efficiency of the sorting center by integrating orders from the sorting centering order to improve the sorting efficiency, all the orders of the Jinan Distribution Center and the relevant information of the order are imported into the graphic database neo4j[1-4]. There is a visualization window in the neo4j that can display the map [5-6]. In the visualization window, we can see the integrated order map so that We better conduct order analysis, and then make reasonable planning for the sorting equipment and layout of the distribution center, so as to improve the sorting efficiency of the distribution center.

Through the analysis of the order data of Jinan EMS distribution center, the orders sent to Jinan and the order details are displayed on the order map, which is convenient for visual analysis of the distribution of EMS orders, data analysis of order sorting, and discovery of current distribution. The

shortcomings of the center, in order to carry out the work of later improvement.

## II. BIG DATA AND GRAPHIC DATABASE NEO4J

Data are created constantly. New digital technologies particularly Mobile, Social, Cloud, Automation device and Big Data have transformed the computers industry [7-8]. Big Data describes the continuous increase in data, and the technologies which are needed to collect, store, manage, and analyze them. These have impact on people, processes and technology. Big Data, from technology point of view, means hardware and software that integrate, organize, manage, analyze and present data which are characterized mainly by three Vs: Volume-large amounts of data; Variety-various of data sources and formats, and Velocity-speed of generation data [9-11].

A knowledge map is a graph-based data structure consisting of nodes (entities) and labeled edges (the relationships between entities), which is essentially a semantic network. Each node in Neo4j expresses the "entity" that exists in the real world, and each edge is the "relationship" between the entity and the entity. In other words, Neo4j is a relational network that connects all kinds of different kinds of information together, thus providing the ability to analyze problems from the perspective of "relationship" and return more comprehensive and rich information.

The main advantages of Neo4j are as follows:

- It is very easy and faster to retrieve/traversal/navigate of more connected data
- It is very easy to represent connected data.
- It represents semi-structured data very easily
- Neo4j CQL query language commands are in humane readable format and very easy to learn
- It uses simple and powerful data model
- It does NOT require complex Joins to retrieve connected/related data as it is very easy to retrieve its adjacent node or relationship details without Joins or Indexes

The creation of the graph database is very easy to operate. Each node corresponds to a record in the relational database. The attributes of the nodes and edges are equivalent to the fields in the record. The content and number of the attributes

can be dynamically changed, and the edges between the nodes are also It can be freely pruned and does not affect the logic of existing data structures [11-12].

The way the map is constructed is generally divided into three types: the top-down approach, the bottom-up approach, and the combination of the two methods. The top-down approach first builds the top-level relationship ontology and then updates the extracted entity matches to the top-level ontology being built. The bottom-up approach directly incorporates the categories, entities, attributes, and relationships found in the extracted data into the knowledge map. Regardless of which method is used to build the knowledge map, the build process is divided into four modules: knowledge acquisition, knowledge representation, knowledge storage, and knowledge visualization [13-14].

The node records contain only a pointer to their first property and their first relationship (in what is often termed the *\_relationship chain*). From here, we can follow the (doubly) linked-list of relationships until we find the one we're interested in, the *LIKES* relationship from Node 1 to Node 2 in this case. Once we've found the relationship record of interest, we can simply read its properties if there are any via the same singly-linked list structure as node properties, or we can examine the node records that it relates via its start node and end node IDs. These IDs, multiplied by the node record size, of course give the immediate offset of both nodes in the node store file.

### III. DISTRIBUTION CENTER EXPRESS SORTING

#### A. Delivery Enterprise Sorting Center

The distribution center is the node of the logistics network system in logistics systemization, and it is the place where the basic functions of logistics are fully displayed. The distribution center is an important node in the express delivery network and plays an important role in the management of the business organization in the region and the communication with other regional businesses. A number of express operations can be realized in the distribution center, including distribution, sorting, inspection, and assembly[15-16].

#### B. Distribution Center

The distribution center is an indispensable place for the logistics process because it has an important simplification effect on the flow of express mail. Its specific functions are as follows:

##### 1) Collection function

The distribution center can effectively combine the goods required by the users and form an economical and reasonable scale transportation effect. The distribution center shows such functions as distribution functions in circulation; using the distribution center to collect goods can increase the full load rate of logistics vehicles and reduce logistics costs.

##### 2) Sorting function

The sorting center can efficiently deliver a variety of goods to customers everywhere, so the distribution center must properly sort the received goods, and on this basis, build and deliver the goods according to the logistics operation plan.

##### 3) Articulation function

Through the cargo transportation and sorting activities, the distribution center can transport all kinds of goods to the customers, which can objectively play the role of production and consumption. This is an important performance of the function of the distribution center.

##### 4) Information function

The distribution center is the centralized place for information transmission, collection, processing and transmission in the logistics system. This information plays a very important role in the modern logistics system and is also an important guarantee for the complex logistics units to be integrated into an organic whole.

##### 5) Management Function

The distribution center plays an important role in optimizing the entire logistics network. From the perspective of development, it not only performs general logistics functions, but also more and more functions such as command and dispatch, information and other nerve centers. It is the soul of the entire express service network. Where.

#### C. Jinan EMS Distribution Center

##### 1) Introduction to the distribution center

The Jinan Mail Distribution Center is located on the west side of the runway of Jinan Yaoqiang International Airport. It is located at No. 7 Airport North Road in Zhangqiu Airport Industrial Park. The west side of the site is adjacent to Airport North Road and the east side is the airport runway. The total investment of the project is 120 million yuan, with a total area of 107 mu, a total construction area of 27,000 square meters, a production area of 23,000 square meters, and a north-south annex building area of 3,000 square meters.

##### 2) Production and operation

The Jinan Mail Distribution Center was put into operation in September 2013 and is a sub-center of China Post Express Express. Responsible for Jinan, Zibo, Liaocheng, Binzhou, Heze, Jining, Laiwu, Tai'an, Zaozhuang, Dezhou, Linyi, Weifang, Dongying, Qingdao, Yantai and other 17 cities express transit package transfer and the province's logistics mail distribution tasks, responsible for each The handling, processing and transshipment of airmail and car mail general contract mail. On the south side of the production site is a cross-border e-commerce processing area and customs inspection and customs declaration area in cooperation with Jinan Customs. The Jinan Mail Distribution Center received one or two orders per day, 226 times for the car-to-car route, and 148 flights for the day. Every day, the daily import and export mail is 150,000 pieces, and the total package is 100,000 bags.

### IV. THE KNOWLEDGE MAP OF EXPRESS SORTING

The construction of the order knowledge map requires the integration of order information. In this knowledge map, the Jinan Distribution Center receives the express mail as the receiving and dispatching bureau, and sends the express mail to the delivery bureau. Among them, 35 express shipments are issued to the Binzhou branch, and 152 express shipments are sent to the Dezhou branch to Heze City. The branch issued 364

express shipments, issued 312 express shipments to Jinan Branch, 91 express shipments to Jining Branch, 19 express shipments to Laiwu Branch, and 103 express shipments to Liaocheng Branch to Linyi City. The branch issued 424 express shipments, issued 53 express shipments to Qingdao Branch, 162 express shipments to Weihai Branch, 27 express shipments to Dongying Branch, and 66 express shipments to Weifang Branch, to Yantai The city company issued 46 express shipments, issued 13 express shipments to the Zaozhuang branch company, and issued 54 express shipments to the Zibo branch company.

Each order data includes city name, district and county name, delivery bureau number, mail number, receiving and dispatching bureau name, recipient's telephone number, delivery date, delivery staff number, and receipt method. Save the data details of each order to Excel, convert the file into a csv file, and use the code to import the order information into neo4j, which is a batch import method. Of course, you can also use the database statement to import data, this situation is suitable for the case of fewer orders, because the import of each order requires an input statement. Each order data exists in the form of entities in the software. The knowledge map will be displayed in the visualization window of the software. Click on a separate entity, and the detailed data of the order can be displayed at the bottom of the window. After the data is imported into the database, we can also use the corresponding query statement to query the order data you want from the order. Then import the entities in the receiving and dispatching bureaus in Shandong.

The distribution and distribution bureaus are in all parts of Shandong. Clicking on a separate receiving and receiving office will display the specific information of the receiving and receiving office in the window. For example, clicking on the Weifang branch will show that the total number of orders sent to Weifang City is 66. When all entities have been imported, the relationship between the entities is established. Organize orders in different regions by establishing relationships. The relationship between the receiving and receiving bureaus is shown by the knowledge knowledge map as shown in Figure 3-1:

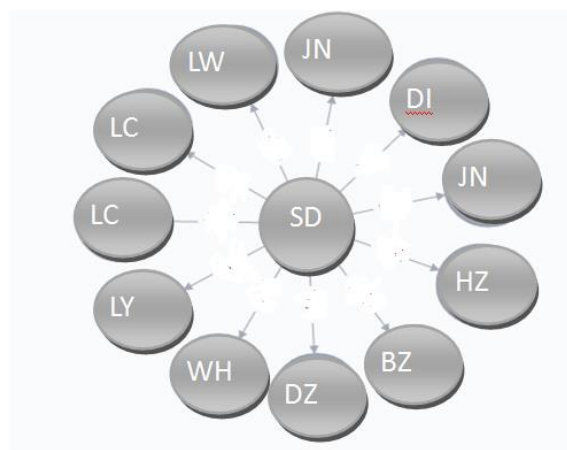


Fig. 1 The knowledge map

Through the display of the knowledge map in Figure 1, we can intuitively see that there are 11 directly sent to Shandong receiving and dispatching bureaus centered on the Jinan Distribution Center. Consistent with the establishment of the above relationship, we can establish a relationship between the receiving and dispatching bureaus and the numerous orders sent to the receiving and dispatching offices.

## V. CONCLUSION

The data is stored and analyzed by the graphics database neo4j and the order data platform of the sorting center. Neo4j can handle complex and dynamic data. Through the establishment of the order map, we can visually see the distribution of orders, facilitate the analysis of order data in the EMS sorting center, and rationally arrange the equipment and facilities of the sorting center. Place the location and make improvements to the sorting center process.

Neo4j is an open source database based on Java development, compatible with a variety of operating platforms, combined with other data powerful data analysis and analysis tools to further analyze large-scale data. All graphic databases can also sort out more orders and perform powerful data analysis.

## REFERENCES

- [1] Design and Implementation of Curriculum System Knowledge Mapping System Based on Neo4j Graph Database[j]. Xiao Qingdu, Qu Liangliang, Hou Xia. Computer Knowledge and Technology. 2017(36).
- [2] Optimization and Implementation of Document Retrieval Method Based on Graph Database[j]. Lin Qisheng, Wang Lei, Zhou Xi, Zhao Fan, Ma Bo. Microelectronics and Computer. 2017(10).
- [3] A Survey of Knowledge Mapping Construction Techniques[j]. Liu Wei, Li Yang, Duan Hong, Liu Yao, Qin Zhiguang. Computer Research and Development. 2016(03).
- [4] Application Research of Visualization Library D3.js[j]. Zhao Cong. Information Technology and Informatization. 2015(02).
- [5] Using D3.js knowledge organization system Web dynamic interactive visualization function[j]. Zhang Yunliang, Zhang Zhaofeng, Zhang Xiaodan, Xu Deshan. Modern Library and Information Technology. 2013(z1).
- [6] Research on Embedded Application of Graphic Database Neo4j[j]. Wang Yulan. Modern Electronic Technology. 2012(22).
- [7] <http://www.gartner.com/it-glossary/big-data/>. accessed: 2017.31.03.
- [8] <https://ec.europa.eu/digital-single-market/en/big-data/>. accessed: 2017.08.04.
- [9] <https://neo4j.com/developer/guide-data-modeling/>. accessed: 2017.16.04.
- [10] Wessler M. Big Data Analytics Dummies; 2013
- [11] B.M. Weedy, B.J. Cory, N. Jenkins, J.B. Ekanayake, G. Strbac. Electric Power Systems; 2012
- [12] <https://www.big-data-europe.eu/energy/>. accessed: 2017.01.04
- [13] <https://github.com/neo4j/>. accessed: 2017.06.03
- [14] S. Cetiner, C. Sepil, H. Sural, Hubbing and routing in postal delivery systems. ANNALS OF OPERATIONS RESEARCH, Vol. 181. No 1. P. 109-124.
- [15] T. Corejova, E. Imriskova, Convergence at the postal market. EKSPLOATACJA I NIEZAWODNOSC-MAINTENANCE AND RELIABILITY, Vol. 3. P. 74-76.
- [16] M. Daskin, Network and Discrete Location. Models, Algorithms and Applications. WILEY, New York. 2013.