ABSTRACT

ABSTRACT: Data warehouse is a data set which is subject-oriented, integrated, steady and time-variant, it is useful for the managers to make decisions .The technology of data warehouse has been widely used in the foreign countries, but it is still in the starting phase in China. The automatic fare collection clean (ACC) system of rail transit has developed very quickly with the development of the city's rail transit in China, and the on line transaction processing system has been built completely. In the past years, the operation companies of rail transit have stored a large amount of data during using the ACC system, and there is a large number of valuable information in the data. How to use the data and how to extract valid and actionable information from the data to support the business operation is at the top of the agenda for the operation companies of rail transit.

The author has looked up technical literature to understand the application status of the data warehouse technology and data mining analysis of rail transit industry in China at present, and the author has learned that the infrastructure of the information system has been doing well, and ACC system is also widely used, and the database of the ACC system has stored a large number of data. But for a variety of reasons, the status is still in a wait-and-see in the aspects of technology applications of data warehouse and data analysis, and without really carrying out.

This paper analyses the status and makes a summary introduction for the knowledge about the data warehouse, and make a detail introduction about my works, such as the definition of the logical model and the physical model, the development of ETL (extract, transform, loading) scripts, and introduce the ETL tools used in the project and application process of the tools. In terms of the basis of the ACC systems in Beijing rail transit, the author has researched some processes and relative techniques to build a data warehouse on the business systems.

Finally, the system has been tested. The system runs smoothly and has a good performance. It proves that the ETL scripts can complete ETL requirements of the data warehouse very well, and have the features of easy-to-use and good flexibility.

KEYWORDS： Data warehouse ;ETL—extract / transform / load ;OLAP—on line analysis processing; AFC--automatic fare collection

CLASSNO：[请输入分类号，以分号分隔。]