China Railway Transportation Intelligent Decision Support System

Sun Yuanyun
Information Technology Center of MOR
Agenda

- Today’s China MOR
- Background of MOR DSS System
- MOR DSS Solution
- Summary
China MOR Overview

- **Operating line at mainland** = 68,650 km
- **Organization Structure**
  - 5,785 stations
- **Equipment**
  - 14,472 locomotives
  - 439,943 cars
  - 565,642 containers
  - 40,033 bridges
- **Revenue**
  - accounted for 48% of the total of the country
  - 165.5 million tons /per year
Rank the first in Asian
Agenda

- Today’s China MOR
- Background of MOR DSS System
- MOR DSS Solution
- Summary
Market Challenge

- Deregulation
- More competition
  - Airline
  - shipping
  - Highway
  - ...
- Business Strategy
  - Objective
  - Why
  - How
  - Assess
- Technological innovation
MOR DSS System Target

- Information Access
- Transportation Performance Analysis
- Market Demand Analysis
- Influence Analysis
- Violation Practice Recovery
- Customer Relationship Management
How to choose the right Software and Vendor

- End to End solution
- Support Standards All Over the World
- Railway Industry experience
- Performance and Functional Scalability
- Methodology, Services and Support
- Usability
- Rapid Return on Investment
Agenda

✔ Today’s China MOR
✔ Background of MOR DSS System
✔ MOR DSS Solution
✔ Summary
System Hardware Environment

ES/9000

Intranet

ES/9000

ES/9000

PC Server

PC Server

PC Server

x.25/FM/DDN

RS/6000

RS/6000

Bureau Intranet

Sub-bureau intranet

Station intranet
System Logic Model

Enterprise Information Portal

Business Intelligence Application

Decision Support Tool (Query & Report & Data Miner)

Access Support Tools

EDW \[\rightarrow\] Data Mart \[\rightarrow\] Other Information

Operational data and External data
Agenda

- Today’s China MOR
- Background of MOR DSS System
- MOR DSS Solution
- Summary
Summary

- Thank you for your attention
- Many thanks to SAS China
- Question & Answers
- Contact
  ✓ http://www.chinamor.cn.net/
A Study on Architecture of Railway Transportation Intelligent Decision Support System Based on Data Warehouse Technology

Sun Yuanyun
(Information Technology Center of MOR of P.R.C)

Abstract: The theory suggests that Railway Transportation Intelligent Decision Support System is built based on the existing railway information systems for better decision making by railway managers. In the paper, the system functions, running environment, data warehouse architecture and software structure are given.

Keyword: Railway Transportation Intelligent Decision Support System, data warehouse, OLAP, metadata

1. Preface

Chinese railway is an important part of China national infrastructure. The operating line at mainland is amounting to 68650 kilometers, rank the first in Asian and the fourth in the world. At present, the railway has a total of 14,472 locomotives, 439,943 cars, 565,642 containers, and 40,033 bridges with its overall length of 2,317,644 meters, 4,893 tunnels with the total length of 2,698,085 and 5785 stations. As the backbone of the country's transportation network, the railway of China has been undertaking a major position in freight transportation. Freight dispatched by the railways per year is up to 165.5 million tons, accounted for 12.4% of that of the total freight traffic. The railways total freight turnover is 13,336 tones, accounted for 48% of the total turnover of the whole country.

However, along with China's reform and opening up to the outside world and rapid development of socialist market economy, transportation industry in China is now undergoing a profound change. Its remarkable characteristic is that scrabbling for customers in the same market is extremely fierce among various transportation modes. Such as the air, the water, the highway, etc. Therefore, the dominated situation of Chinese railway in transportation market is progressively stepping down and its competitive advantage are facing a serious threat, which is probably even worse following China's accession to WTO. In order to maintain and enhance the competitive power and the advantage of Chinese railway in transportation market, it must adjust its business strategy in accordance with the demand of the market and always make out wise and correct decisions on the basis of objective reality.

In the beginning of 1980's, MOR has built many computer operational systems automate
the tasks necessary to perform business process. At the end of the 1990’s MOR was developing the Transportation Management Information System (TMIS) of Chinese railway. These systems have been production and implicated in various kinds of business data. So fully utilizing the modern information technologies and the existing systems of MOR and turning all the related business data into practical decision-making knowledge of Chinese railway management personnel at various levels is the key for them to make correct decisions and hold the best business opportunities at all times. In the paper we will introduce the Railway Transportation Intelligent Decision Support System’s function first, then we will discuss the data warehouse’s architectural and software components.

2. Business Function

The system is geared to establishing a decision-making support environment for the transportation management and business analysis personnel of Chinese railway. The system make full use of modern information technology (such as data warehouse, on-line analysis process, data mining, etc.) to realize that all the business processes and procedures such as information accessing, report generating, transaction analyzing and decision-making, etc. are of high automatic and intelligence. The typical business functions include seven models to the following aspects:

Data Management Model

To build freight data warehouse and multi-objective-oriented data marts through extracting, transmitting, integrating and loading data from the existing systems of MOR, such as the Waybill Management Information system, Train Consists Management Information System, Automatic Train Identification System, and the marketing and Technical plan Management Information system.

Information Access

Provide the following information access functions:

- Waybill information search.
- Goods status and location trace (sent out, on the way, arrival, delivery)
- Carrying equipment (cars, containers, etc.) status and location tracing (empty.loaded, at station/on the way, commodity name, etc.)
- Special goods information search

Transportation Performance Analysis

Major transportation indicator statistics, comparing and analysis, including traffic volume, income, cost, benefits, etc. and freight transport indicator forecast
Market Demand Analysis
Commodity distribution and flow analyze based region, arrival and delivery data. Freight transportation requirement forecast using history and demography data.

Influence Analysis
Forecast freight transportation income if adjust the route or tariff. Influence of liquidating freight line grades dividing on liquidation enterprise's income. The model enables the business users to harness the power of the world's strongest analytic capability to design marketing campaigns that will result in higher return on investment and more profitable customer relationships.

Violation Practice Recovery
Present fraudulent bills, Payment made for less-than-carload freight at the price of carload, Decreasing freight charges and other violation practices against rules and regulation.

Customer Relationship Management
Classify customers according to different kinds of professions, districts and their contributions to railway transportation. Analyze customers' trade behavior, e.g. their partiality for railway transport products, to provide specific character services. Forecast customers' transportation demands.

3. Architecture of TIDS

Hardware Environment
The IT environment of TIDS consists of three IBM mainframe systems (ES/9000), The existing 3 mainframe systems have made up of two (2) clusters. The three (3) mainframes among the cluster structure share two (2) disk system together and one (1) tape library. The hardware configuration of TIDS is shown in Figure 1. The first cluster is used as OLTP and ODS (part of Enterprise Data Warehouse) server and Enterprise Data Warehouse data management platform, for data extraction and large volume data management. The second one is used as application development.

In order to ensure and maintain the performance of the production system and avoid too much influence to the production, TIDS system has installed two (2) RS/6000 systems to meet the need of a large volume of inquiries and operation analysis. Two (2)-set RS/6000 servers can realize cluster configuration by using HACMP for RS/6000 software. One of
the RS6000 box is used as both common Analysis Application server and Web Server, for subject oriented data and DSS application access. The other RS6000 box is used as computing intensive Analytical Application server.

For system administrator, a dedicate PC is used as the administration platform, for centralized data warehouse administration, and data model design / integration with data warehouse.

Now the system has the capability to process high volume data up to TB level, Support more than 50 concurrent-clients and 1000 potential clients accesses, support 7 x 24 clients' access mode.

![Figure 1 TIDS System Hardware Environment](image)

**TIDS Data Warehouse Logic Model**

Data Warehousing is the process of making your operational data available to your business managers and decision support applications. Data warehousing doesn't just make data available; proper warehousing focuses on efficient information access. Of course, this efficiency doesn't happen magically. You must evaluate your current operational data and determine how to transform that data into what the business user requests. The tools that you choose for your warehousing solution will take data from your operational systems (extract it), convert your operational data into business information using your defined business rules (transform it), and create a data warehouse (load it).

✧ **Operational Data**

3
The data warehousing process begins with your operational data. It is this data that you must analyze and evaluate, you must cleanse and translate it, and ultimately will populate your warehouse with it. The term operational data is most often used to refer to data that is generated by your online transaction processing (OLTP) system. Operational Data in the system included Waybill Management Information system, Train Consists Management Information System, Automatic Train Identification System, and the marketing and Technical plan Management Information system.

✧ **Staging Area**
This contains usually unprocessed data from the operational source systems for later loading into the data warehouse. Not all warehouses require a staging area. It is used most often when the operational database is on line 24 hours a day and cannot be directly extracted by warehouse extract-transform-load processes. It is also used when the load frequency of the data warehouse differs from the required extract frequency.

✧ **Operational Data Store**
An operational data store (ODS) is an integrated, subject-oriented (often customer based) collection of data directly from operational systems. The records are updated as the transactions in the operational systems are updated.

✧ **Data Warehouse**
A data repository in which information is stored in a format that is suitable for business intelligence and decision support systems.

✧ **Enterprise Data Warehouse**
The enterprise data warehouse is a central store of data common to the organization. It is a central repository of information drawn from disparate and physically distributed operational systems of an enterprise, as well as external data.

✧ **Departmental Data Warehouses and Data Marts**
Departmental data warehouses or data marts contain data and information focused on the narrower needs of a specific department or business unit. Data marts can be built based on data in the central enterprise data warehouse (that is, they are dependent on the data in the data warehouse) or they can be created independently from independent data sources. Independent data marts are not recommended because they create islands of information that data warehouses avoid.

✧ **Application Specific Extracts**
Application specific extracts contain subsets of data extracted from the departmental or business unit data mart or a data warehouse to support some other business application or process. Application specific extracts do not require specific exploitation tools as they are used to provide input into some other business process. These can include varied business applications such as: credit scoring, market segmentation, householding, lifetime customer value estimation, supply chain management, sales or inventory management, quality
management, statutory or management reporting, forecasting or feeding back directly into the operational systems.

✧ **Business Intelligence Portal**
Business intelligence portals allow the exploitation of the data warehouse through a single web interface. This interface provides a personalized view of the data warehouse such that individuals within the organization receive the information they want, when they want it. The information available to each person is controlled centrally via security to ensure only appropriate data is viewed. The portal mechanism also allows people to share information so that others interested in the same subject can quickly access business problem or opportunity found by one person.
Application Software Architecture

As shown in the following figure, TIDS system included five logical layers. Which is data access and management layer, data warehouse building layer, enterprise information portal layer, business intelligence layer and decision support layer.

- **Extract, Transform, Load**
The Key Data Warehousing processes are to extract data from the operations source systems, transform it to support business decision process, load it into the enterprise warehouse: this is known as the Extract, Transform, and Load (ETL) process. In the project we used SAS/ACCESS interfaces to create a SAS data set from operational system data, update or write data to ODS from within SAS.

- **Metadata**
Metadata is data about the data in the warehouse, the processes supporting the warehouse, the business, and the organization. This is used to document Data Warehouse contents, manage processes and change, establish security, and control and promote standards. Metadata is an important and vital part of any Data Warehouse initiative regardless of scope and size. Derived from a variety of disparate sources including technical, as well as
business sources, a metadata strategy and understanding the importance of metadata ownership, accountability, and reliability are critical. The owners of metadata may include information owners, information producers, process owners, information consumers, and system, database and data quality administrators. In TIDS system SAS/Warehouse Administrator software provides a central point of control for creation, maintenance, and management through a metadata-driven architecture. It supports the creation, management and use of metadata by warehouse developers, administrators, applications developers and end-users in all phases of warehouse development and usage. As an integral part of the SAS System it enables seamless integration of metadata with SAS exploitation tools.

Stars:

Data Quality
Data quality is a major concern in warehouse and information environments. Often, it is the creation of the first Data Warehouse that surfaces data quality issues, rather than creates the problem. The SAS/Warehouse Administrator enables IT to surface and deliver consistent, accurate, and reliable information representing a "single version of the truth" to the business community. Furthermore, the metadata helps to define and promote standard terminology across the enterprise. SAS Enterprise Miner have been used for rule discovery. SAS/QA and some of the SAS descriptive statistics capabilities are ideal for measuring the problems, often providing analysis key to improvement.

Data Warehouse Storage
We know data in the warehouse is organized by subject, rather than application, so that the warehouse contains only the information necessary for decision support processing ('subject-oriented'). SAS SPD Server combines optimal performance capabilities with a highly scalable solution enabling IT to provide the user community with reliable and secure information. By utilizing the latest parallel process and data server capabilities, the SPD Server delivers a fully integrated and seamless way to access large volumes of data and serves large numbers of concurrent users. An open data server, it enables extremely efficient data access for hundreds of network clients across multiple processors, a requirement for 24/7 operations.

Olap server
SAS OLAP technology support multi-dimension conceptual view, such as slicing and dicing, rotate, etc. Users can access distributed data transparently with no awareness about which MDDB structure and Viewer are adopted. SAS OLAP Server and OLAP Viewers, such as SAS/EIS or SAS Enterprise Guide, can provide the full Client/Server capability. SAS can use its HOLAP and MDDB partition strategy to ensure the consistent reporting performance, when data volume increased.

Closed-loop Integration
SAS Integration Technologies uses industry standards for object transport, message-oriented middleware, LDAP directory services and a publish/subscribe
framework to make SAS software solutions an easily integrated element of enterprise information flow.

❑ **Ad hoc Reporting**

We understand that users will want to produce reports that they use today from the Data Warehouse. SAS OLAP Technology provides the ability to analyze the warehouse data. From high-level overviews, to the minutest detail, the technology provides the ability to drill out to see how it all fits together. With Enterprise Guide Business analysts, end users and SAS programmers can run any SAS processes, and return professional reports and graphics to their PC. Enterprise Reporter software provides an interface designed with a Microsoft Office "look and feel". With a familiar palette for report design, users can simply paint, preview and publish reports, while the system takes care of the data.

❑ **Business Intelligence Portal**

Business intelligence portals allow the exploitation of the Data Warehouse through a single WEB interface. This interface provides a personalized view of the Data Warehouse such that individuals within the organization receive the information they want, when they want it. The information available to each person is controlled centrally via security to ensure only appropriate data is viewed. The portal mechanism also allows people to share information so that others interested in the same subject can quickly access a business problem or opportunity found by one person. This speeds up the collaborative process of decision-making. A component of SAS Integration Technologies is Publish and Subscribe -- our proactive information delivery framework. It allows users to subscribe to information channels that pertain to a particular topic. In effect, they can create their own personalized information portals. Information is then pushed out as a personalized package directly to consumers, who can receive it via e-mail or the Web.

❑ **Data Mining and Analysis**

Analytical Solutions are seen in the market today as the foundation for decision support applications. An integral part of BI applications such as OLAP gives us technical modeling tools to infer patterns and insights in our data that can help support the management decision-making process. Examples of applied analysis applications that feed into decision support would include business applications such as database marketing, market research, forecasting, quality analysis, and operations research. Analytical Solutions is the bridge between the technology of Data Mining and decision support output. The core competency of the SAS System provides a complete environment for advanced statistical analysis. Applications can incorporate any number of techniques including: probability analysis; statistical inference; hypothesis testing; regression analysis; correlation analysis; composition analysis; value-added analysis; time series analysis; indices; what-if analysis; sensitivity analysis; impact analysis; and proportion disparity (Pareto) analysis. Enterprise Miner is the first and only Data Mining solution that addresses the entire Data Mining process -- all through an intuitive point-and-click graphical user interface (GUI). Combined with SAS Data Warehousing and OLAP technologies, it creates a synergistic,
end-to-end solution that addresses the full spectrum of knowledge discovery.

4. Summary

Based on the SAS rapid warehousing Methodology and products, the TIDS system provided well defined, well-organized and well-managed data to the railway staff, to improve the quality of railway production and the decision-making level, achieve scientific and intelligent decision-making mode, hold market demands in time, optimize the distribution of railway resources, speed up the transition from an extensive management mode to the intensive and make a great contribution to the development of national economy.

Reference Document

3. SAS Whitepaper: Finding the Solution to Data Mining
4. SAS Whitepaper: SAS ® Rapid Warehousing Methodology