# THE ROLE OF EXPERT SYSTEMS, AND RDBMS STRATEGIES IN A DSS FOR URBAN BUS TRANSPORT MANAGEMENT

**by**

**Varghese George**
**S.L. Dhingra**
**P.K. Sikdar**
**Department of Civil Engineering,**
**Indian Institute of Technology, Bombay - 400076.**

## ABSTRACT

Efficient and well-managed urban bus transport systems supported by information systems with decision support capabilities developed within the framework of an advanced Relational Data Base Management System (RDBMS) can provide critical information at the right moment and assist transport managers in conducting performance evaluations. This paper includes discussions on the use of the entity-relationship (E-R) model (a semantic data modeling technique) in the development of the structure of the conceptual database for the information system planned to be implemented using RDBMS. The E-R data modeling approach enables database designers in obtaining the third normal forms of related databases for the efficient functioning of the information system. The inportance of decision tables in the development of DSS modules using Expert System shells are also discussed. The DSS modules will assist transport managers in the analysis of operational performance for bus depots or the organization as a whole. The E-R diagrams generated and the decision files developed will serve as important documents that can enhance the adaptability of the DSS to the changing needs of the organizations. The conceptualization of the information system to support decision-making in an RDBMS framework provides the advantage of a very low 'disk seek' time and facilitates frequent generation of reports.

## 1 INTRODUCTION

Large cities are powerful engines of economic growth due to their inherent higher productive efficiencies. Cities will be able to maintain their high productivity and high degree of mobility as long as the mass transport facilities they provide meet the travel needs of the people. Hence, efficient management of public transport services in metropolitan cities is of great importance. But transport supply levels especially in developing countries remain low due to limited financial and energy resources. The increase in population of urban conglomerations has further widened the gap between the travel demand and the transport supply.

In addition to the above problems, the non availability of critical information at the right moment on various aspects of the functioning of bus transport undertakings will have an adverse effect on their performance. The losses suffered by transport undertakings in the operation of buses necessitate constant monitoring of operational activities. Efficient and well-managed bus transport systems supported by information systems with decision support capabilities developed within the framework of an advanced Relational Data Base Management System (RDBMS) can provide solutions to the above problems.

A Decision Support System (DSS) relies to a large extent on a huge amount of processed data supplied by a well-designed information system, and the effectiveness of expert system (ES) modules for its proper functioning. The information system and the ES modules constitute the main components of a comprehensive decision support system. The information generated from the operation of buses can be effectively used in monitoring and analyzing the performance of bus transport systems. Toward this end, the well tried and most popular semantic data modelling technique (Meyer et al. 1996) that uses Entity-Relationship Diagrams (ERD) for the design of a conceptual database system based on the relational model was adopted. This paper focuses on the application of advanced concepts of relational data modeling using the ERDs. It also demonstrates the use of decision tables in the design of selected DSS modules.

A relational database management system that uses the powerful features of relational-calculus and/or relational algebra is an ideal tool that can be used in the design of information systems that support modules for decision-making. An RDBMS that works on the principles of relational calculus permits the use of non-procedural structured query languages (SQL) that can significantly enhance the productivity of database programmers. Decision support sub-modules that utilize techniques in the field of statistics, pattern recognition, and expert systems can be coupled to information systems in order to impart decision-making capabilities in an interactive environment (George et al. 1996).

The development of decision-support modules was performed in a systematic manner by using decision files or tables. A demonstration on the use of decision tables in the design of selected decision-support modules, and a brief discussion on some aspects of interfacing the modules to the information system are also included in this paper. The work discussed here is part of a comprehensive decision support system (DSS) developed and implemented using the ORACLE RDBMS. The data entry, update and display forms for the same were created using the SQLFORMS utility of ORACLE.

## 2 THEORETICAL ASPECTS

Relational data base management systems (RDBMS) are capable of retrieving a number of rows from a data base in a single disk seek operation unlike non-relational systems. They can also access and extract data from a number of databases simultaneously. The semantic data modeling technique that uses entity-relationship (E-R) diagrams constitutes one of the best tools for modeling of relational data bases. The following sub-sections discuss the theoretical aspects of semantic data modeling and the use of standard indices for performance evaluations.

### 2.1 The Use of E-R Diagrams in Semantic Data Modeling

Semantic Data Modeling aims at capturing the meaning of data in a formal way so that database design becomes systematic and the data base itself can behave intelligently. It has its roots in the methodologies developed for artificial intelligence (Prabhu, 1992). Semantic data models include models proposed as extensions to classical data models (such as the network, heirarchical, and relational models) and models that are significantly different from classical models. The entity-relationship model belongs to the former. In semantic data modelling, information is modelled in terms of entities or objects. These do not require strict encapsulation as object-oriented data models do, and hence are considered to be 'structurally object-oriented' (Dittrich, 1986). The Entity-Relationship model proposed by Chen in 1976 was the first formal definition of a semantic model. It offers high data independence and is one of the most popular design tool used in systems analysis, design, and documentation for databases to be implemented by conventional (classical) data models (Meyer et al. 1996; Prabhu, 1992). It was thus originally proposed as an extension of classical data models.

E-R diagrams consist of rectangles to represent entity sets, ellipses to denote attributes, diamonds to depict relationships among entity sets, and lines to link attributes to entity sets and entity sets to relationships. At times, an attribute itself may be considered as an entity based on the objective of the design. E-R diagrams assist in obtaining third normal relational data bases which ensure that transitive dependencies do not exit. This requirement also ensures the elimination of update anomalies. ERDs, therefore, serve as a basis for the design of stand-alone data base management applications.

Basic documents such as existing forms, monthly reports, yearly reports, descriptions on manual procedures, and clerical operations, ticket memo forms used by conductors, and trip cards filled by drivers assist in the design of the databases. These can assist in identifying the important entities in the system and in understanding their interrelationships. The important steps in the construction of E-R diagrams include a study on the existing business environment, and the associations between the entities identified. Data tables for every entity set must be defined to simplify or eliminate the need for normalization of data, and the primary keys for accessing each record in the data base must be established.

### 2.2 The Role of Performance Indicators in the DSS Submodules

Standard performance indicators widely used by transport undertakings are used in the design of expert system modules for performance evaluation and monitoring. The use of these indicators in the DSS modules will assist the bus transport managers in strategic and tactical planning and in decision-making. The insights and suggestions provided by the modules will guide transport managers in performing rationalization of routes, services, and the related operations (George et al 1996).

Studies conducted by Fielding et al (1985), Fielding and Anderson (1983), and Wright and Thiriez (1987) provided the required background for the application of performance indicators related to resource effectiveness, resource efficiency, service efficiency, and cost efficiency for performance evaluation of urban bus transport systems. The standard values for comparison of performances were obtained from a study on quarterly and yearly reports related to performance evaluation of urban bus transport systems published by Central Institute of Road Transport, Pune and the works of Wright and Thiriez (1987).

## 3 DEVELOPMENT OF THE DATA STRUCTURE

The problem of development of the information system with decision support capabilities was partitioned horizontally based on the major functions to be performed by the DSS. This resulted in the generation of the basic modules FHCRM (fleet held and capacity relationship module), TMEM (ticket memo processing module), COST_DSS (cost module), and STAFF_DSS (staff module). The FHCRM and TMEM modules were further partitioned and decoupled vertically into various submodules to expose their associated sub-functions. The submodules of FHCRM, and TMEM were designed so that their data bases could be linked to the REV_EFU (revenue and effective fleet utilization) module at later stages. The submodules of STAFF_DSS and COST_DSS, and the REV_EFU module were developed such that their related data bases could be connected to the HLDSS module. The details of the activities proposed to be performed on the basic modules identified, and the related submodules developed are mentioned below:

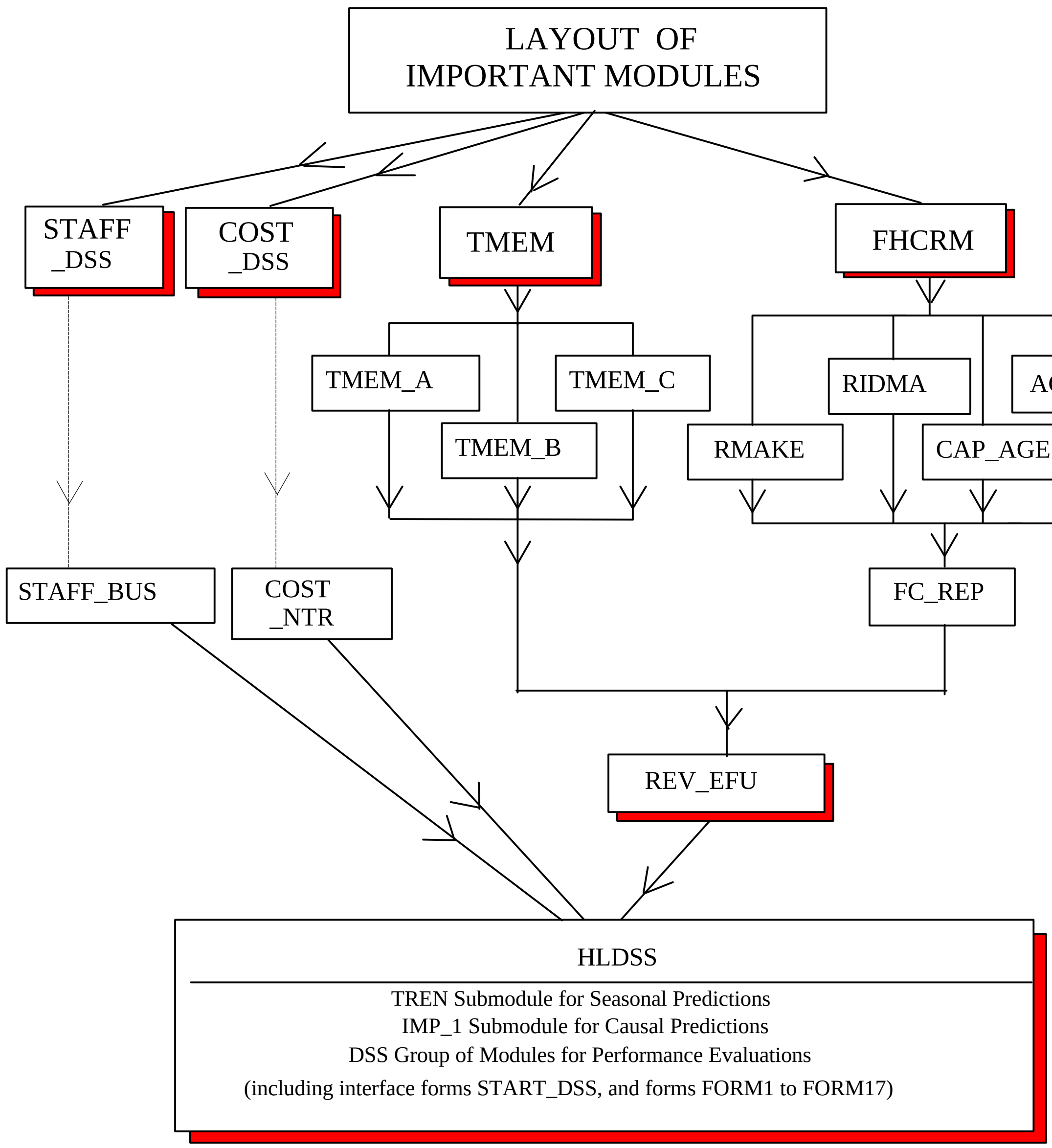
- Modeling of the structure of the databases for the 'fleet held and capacity relationship' module (FHCRM) and development of data manipulation and display submodules RIDMA, RMAKE, CAP_AGE, AGES, and FC_REP.

- Data modeling for the structure of the databases for the 'ticket memo processing' module (TMEM) and the development of related data manipulation submodules TMEM_A, TMEM_B, and TMEM_C. These submodules deal with the data entry and update of three main parts of the ticket memo form filled by bus conductors at the end of each bus trip.

- Formulation of the data structure for the 'revenue and effective fleet utilization' module (REV_EFU) comprising data bases REV_KM_M and REV_KM_D that can supply important information to the DSS for performance evaluation.

- Design of a suitable data structure for the COST_MON data base of the 'cost' module (COST_DSS) to hold information on various cost components relevant to the DSS.

Figure 1: **Overall Layout of DSS with Modules and Interconnections**




- Formulation of the data structure for the STFBUS_M data base of the STAFF_BUS submodule (designed as part of STAFF_DSS) to store information on the staff employed and buses held by the depots.

- Design of databases for 'higher level DSS' modules (HLDSS) that assist managers in performing causal predictions, forecasting of seasonal variations, and performance evaluations, and the development of their interfacing modules.

The design and development activity related to the TMEM module provided the data base structure for the data tables TKMEM_A, TKMEM_BN, and TKMEM_C that could permit route level analyses at later stages if required. The TMEM module supports entry of data contained in three parts of the ticket-memo form filled by the conductor at the end of each trip through the submodules TMEM_A, TMEM_B, and

TMEM_C. These submodules assist in the insertion of data into the third normal databases TKMEM_A, TKMEM_BN, and TKMEM_C respectively. Transport managers with knowledge on the use of SQL*Plus utility of ORACLE can directly access the required information on the performance of a bus for a particular trip on a particular route.

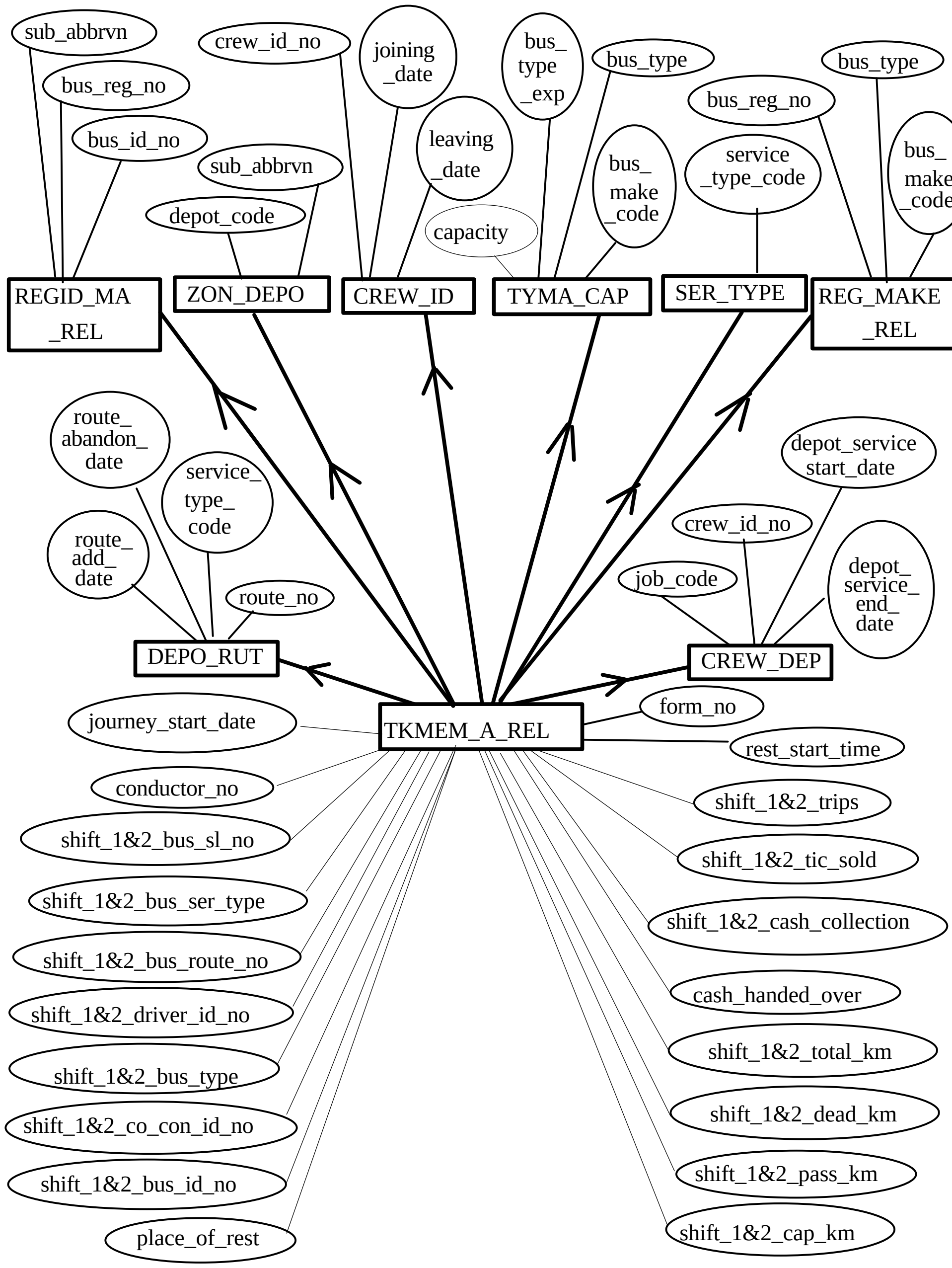
The data bases for the REV_EFU module were designed to facilitate the acceptance of data from TKMEM_A, TKMEM_BN, and TKMEM_C data bases at later stages if required for route level analyses. This has not been attempted at present since the aim was to develop a DSS for depot level analysis.

The partitioning and decoupling activity that aimed at decomposition of the problem into easily understood components provided the basic overall programme structure illustrated in figure 1. This hierarchical representation provided a clear idea on establishing the information content required by each data base and the flow of information. The overall programme structure also assisted in planning the interfacing of various modules at various stages. The following subsections provide relevant details on the TMEM_A submodule. The TMEM_A submodule was designed for the entry of data contained in the first part of the ticket-memo sheets filled by bus-conductors at the end of each bus trip.

### 3.1 The TMEM_A Submodule

The TMEM module consists of three connected data entry submodules, TMEM_A, TMEM_B, and TMEM_C. These modules correspond to the three main parts of the ticket-memo data sheets filled by bus-conductors at the end of each bus trip. The TMEM_A submodule comprises the TKMEM_A_REL relationship table with details on the date of journey, form serial number, details of the crew, details of the place of rest after the first shift duty, and the cash collected by the conductor in the first and second shift duties. The E-R Diagram for the TMEM_A submodule is illustrated in figure 2. It shows the important entities (represented by rectangles), the attributes (denoted by ellipses), the relationship developed for display of the report (depicted by a diamond), and the interrelationships shown by the lines and arrows. The entity ZON_DEPO comprises basic details on the code numbers for the bus depot, its sub-abbreviation, and the zone of the city where the depot is located.

Figure 2: **E-R Diagram for TMEM_A Submodule of TMEM**




The TYMA_CAP entity consists of data on the general types of buses (single-decker, double-decker, and midi buses), the categories of make, and the carrying-capacities for various categories. When a new bus is purchased by a bus transport organization, the details on the type of bus, its category of make, the year of manufacture, the name of the body-building company, the engine and chassis numbers and the registration number (or license plate number) are entered in a separate database (REG_MAKE_REL) accessible at the organization level. When this bus is allotted to a depot, the depot authorities maintain details on the date of entry into the depot, the type and make of the bus, and the unique identification number allotted to the bus in the data base REGID_MA_REL.

The entity CREW_ID consists of general information on the crew such as the name, date of birth, date of joining the organization, date of leaving the organization, crew identification number, address, etc. CREW_DEP comprises details of the crew entered at the depot level such as the job-code at the depot, date of joining the depot, and the date of leaving the depot. The DEPO_RUT entity contains data related to the route numbers served by each depot, the type of service (ordinary, or limited), the date when the route was commissioned, and the date when the route was abandoned. The SER_TYPE entity provides explanations on the code numbers given to the type of services.

Similarly, the TMEM_B submodule comprises the TKMEM_B_REL relationship that serves as a temporary data table. This helps in feeding data regarding ticket-sales in each trip for tickets of each denomination as appearing on the ticket-memo data sheet. Immediately after entry of data into this form, the data is copied into a permanent third normal database TKMEM_BN automatically. The TMEM_C submodule comprises the TKMEM_C_REL relationship that records information on the timings of the starting and ending for each trip at the terminal stations, and the serial numbers for the terminal stations.

The data entry and update forms are designed such that when the data is correctly entered in all the three parts of the forms, the total km., the dead km., the passenger km., and the capacity km. will be automatically calculated and entered into the TKMEM_A data base. Provision has been made for the automatic deletion of data pertaining to incomplete ticket memo forms. Other related data bases and data manipulation forms were also designed in this fashion. The data entry and display forms developed using the SQL*Forms utility of ORACLE were incorporated with numerous safety features to prevent entry of erroneous data.

### 3.2 The Structure of Other Important Data Bases

Managers of urban bus transport organizations use monthly and quarterly data and reports to a large extent in decision-making at the depot level. DSS modules must be able to access and make use of these data bases for effective decision making at the depot level and at the system level. Three important data bases were developed to cater to this requirement. The STFBUS_M database that forms part of the STAFF_DSS submodule provides details for each month for every depot on the number of traffic staff, workshop staff, administrative staff, other staff, and the average buses held for the month.

In this manner, the COST_MON database that is part of the COST_DSS submodule provides information for each month on the personnel costs, fuel costs, cost of spares, cost of tyres and tubes, taxes, interest paid, miscellaneous expenses, depreciation, fixed cost, and non-traffic revenue generated (in the form of rent, etc.) for the month. The REV_KM_M database that forms part of the REV_EFU submodule provides information for every month for each route of the depot on the passengers carried, average buses on road, average bus trips on each route, effective km. operated, total km. operated, traffic revenue generated, capacity km., and the passenger km.

These three major data bases will assist the DSS in determining other related performance indices for decision making. In addition to the three data bases mentioned above, the structure of a daily database for storing information on the revenue generated for each route for each day similar in structure with that of the REV_KM_M database was also developed.

E-R Diagrams were not developed for STAFF_DSS, COST_DSS, and REV_EFU submodules since our ultimate aim was to give a structure to the database for performing decision-making through DSS submodules. Decision files were developed for the formulation of DSS submodules for conducting performance evaluations on the bus system. These decision files provided the fundamental information required for the development of rule-based expert system submodules developed using VPExpert system shells.

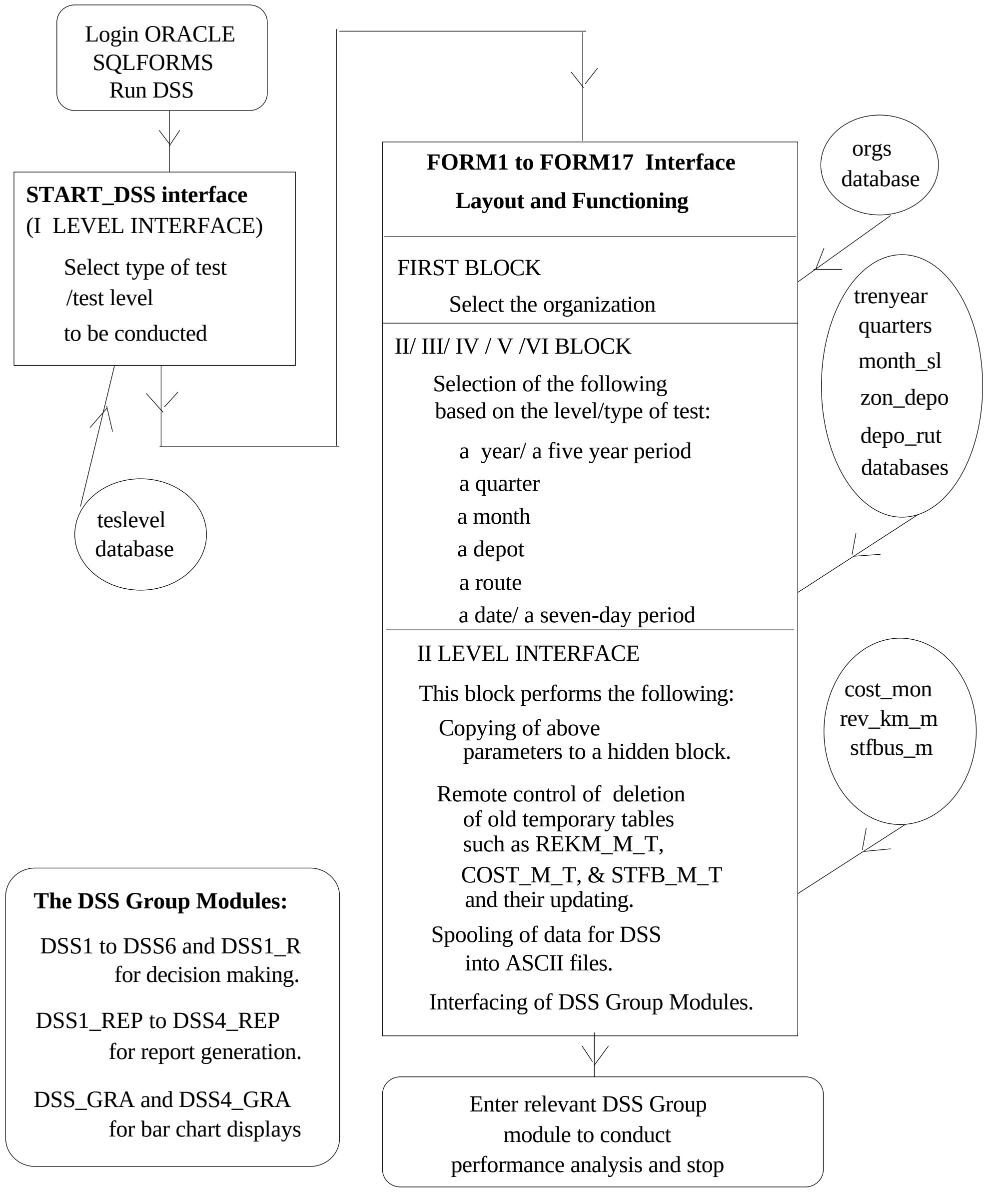
### 3.3 Interfaces for DSS Group Modules

The SQL*Forms utility of ORACLE was used for designing the form 'DSS'. that could be used as a 'starting front-end' in selecting the type of decision-making activity to be performed by the manager. Managers are provided with options to conduct performance analyses at the system level or the depot level (sub-system level). Analysis can be carried out for a particular year, quarter, or a month.

Once the test level has been selected, the form 'DSS' then interfaces to one of the 'intermediate front-end interfacing forms', 'FORM1' to 'FORM17', that performs the required task. Some of these forms also permit tests to be conducted on selected routes. These 'intermediate interfacing modules' obtain information regarding the year, quarter, month, route, and date for which data has to be retrieved and analysed from the manager interactively.

On receiving the relevant data, these forms perform background operations including deletion of old temporary tables, retrieval of relevant data from the data bases and storage into temporary data bases. The forms then create ASCII files containing the required data and run the relevant DSS submodules for performance evaluation. The details of the interfacing mechanism are provided in figure 3. The names of data bases connected to the interfacing forms are also shown in the figure.

Figure 3: **Interfacing Mechanism for DSS Group Submodules**




## 3.4 Sample Decision File for a DSS Module

Table 1 provides the decision file used for the development of a part of a decision support module that comments and advises on the vehicle utilization co-efficient for buses on road. The interface forms retrieve the required data for the analysis and create ASCII data files. These files are read by the expert system module to calculate the vehicle utilization coefficient. The expert system then matches the result obtained to standard *operation condition*. It then searches for the related *action* to be performed in the form of a terminal display that provides comments on the present status of the system and the corrective actions to be taken if any. A number of such decision files are formulated for the development of decision support modules.

Table 1: **Decision File for Vehicle Utilization Coefficient Checks**

| **Operating Conditions & Action Levels** | **1** | **2** R | **3** U | **4** L | **5** E | **6** S | **7** |
|---|---|---|---|---|---|---|---|
| Vehicle Utilization Coefficient>= 400 | ✓ | | | | | | |
| Vehicle Utilization Coefficient >=350<br>Vehicle Utilization Coefficient <=400 | | ✓ | | | | | |
| Vehicle Utilization Coefficient >=300<br>Vehicle Utilization Coefficient < 350 | | | ✓ | | | | |
| Vehicle Utilization Coefficient >=250<br>Vehicle Utilization Coefficient < 300 | | | | ✓ | | | |
| Vehicle Utilization Coefficient >=200<br>Vehicle Utilization Coefficient < 250 | | | | | ✓ | | |
| Vehicle Utilization Coefficient >=150<br>Vehicle Utilization Coefficient < 200 | | | | | | ✓ | |
| Vehicle Utilization Coefficient < 150 | | | | | | | ✓ |
| **Action Level 1***<br>Module displays "HIGHLY EXCELLENT (for rule 1, 'excellent' for rule 2, and 'very good' for rule 3).<br>You are requested to check the occupancy ratio and the passengers carried per bus on road." | ✓ | ✓ | ✓ | | | | |
| **Action Level 2***<br>Module displays "GOOD (for rule 4, 'satisfactory' for rule 5, 'poor' for rule 6, and 'very poor' for rule 7).<br>You may think of improving the vehicle utilization coefficient further by:<br>using well-maintained buses,<br>employing efficient and productive work shop staff,<br>employing highly efficient crew and traffic staff,<br>implementing telescopic fare structure,<br>merging shorter routes into longer ones where possible,<br>plying buses on well maintained and less congested roads with high traffic demand." | | | | ✓ | ✓ | ✓ | ✓ |

## 4 CONCLUSION

Techniques in the field of *software engineering*, were applied to advantage in the design of various components of the DSS. A *top-down* approach was used in identifying the major modules to be developed as part of the exercise of conducting a *software requirement analysis*. The *partitioning* and *decoupling* activity performed at the initial stages assisted in decomposing the complex problem into easily understood components. It provided a clear idea for establishing the information content required by each data base and the flow of information to various interlinked modules.

The design of the basic structure of databases using E-R Diagrams for the efficient functioning of various submodules constituted part of the *software requirement* *specification* for the development of the DSS. The submodules were developed with a view to maintain high *modularity* and *cohesiveness*.

The use of Entity-Relationship modeling technique for a problem conceived in the relational database framework provided the combined advantages of both semantic data models and relational data models. The potentially powerful concept of interactive and flexible DSSs (Sprague and Watson, 1986) coupled with the use of RDBMSs that have capabilities to use non-procedural structured query languages (SQL) ensured increased productivity right from the development phase that involved coding, testing, and implementation.

## 5 REFERENCES


Chen, Peter P. (1976) The Entity-Relationship Model: Towards a Unified View of Data, *ACM Transactions on Database Systems, 1(1)*, 9-36.

Dittrich, K.R. (1986) Object-oriented Database Systems: The Notion and the Issues, in K.R.Dittrich and U. Dayal (Eds.), *Proc. 2nd International Workshop on Object-oriented Database Systems*, Bad Munster, FRG, IEEE Computer Society Press, pp.2-4.

Fielding, G.J. and Anderson, S.C. (1983) Public Transit Performance Evaluation: Application to Section 15 Data, *Record 1947*, Transportation Research Board, National Research Council, Washington, D.C.

Fielding, G.J., Babitsky, T.T. and Branner, M.E. (1985) Performance Evaluation for Bus Transit, *Transportation Research, 19A(1)*, 73-82.

George, Varghese, Dhingra, S.L., and Sikdar, P.K. (1996) Development of a Decision Support System for Urban Transport Planning and Monitoring, *Proc. 7th Int. Conf. on Urban Transport in Developing Countries (CODATU-VII)*, CODATU Association, New Delhi, 12-16 February, pp.IV-55-65.

Meyer, Bernd, Westerman, Gerd D., and Gogolla, Martin (1996) Drafting ER and OO Schemas in Prototype Environments, *Data and Knowledge Engineering, 19(3),* 201-240.

Prabhu, C.S.R. (1992) *Semantic Database Systems: A Functional Introduction*, Universities Press (India) Pvt. Ltd., Hyderabad.

Sprague, Ralph H.,Jr., and Watson, Hugh J. eds. (1986) *Decision Support Systems: Putting Theory into Practice*, Prentice Hall International Inc., Englewood Cliffs, New Jersey.

Wright, A.A., and Thiriez, S. (1987) Bus Services: Reducing costs, Raising Standards, World Bank Technical Paper No. 68 (July), *Urban Transport Series,* The World Bank, Washington, D.C.