

# A RELATIONAL MODELING SYSTEM FOR LINEAR AND INTEGER PROGRAMMING

A. ATAMTÜRK, E. L. JOHNSON, J. T. LINDEROTH, and M. W. P. SAVELSBERGH

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205  
atamturk@ieor.berkeley.edu*

(Received October 1996; revisions received April 1998, April 1999; accepted July 1999)

We discuss an integer linear programming modeling system based on relational algebra. In this system, all modeling related activities, such as model formulation, model instantiation, and model and instance management, are done using simple operations such as selection, projection, and predicated join.

## 1. INTRODUCTION

Anyone who has ever attempted to apply mathematical programming in practice knows that it is usually not a simple and straightforward exercise. The road from a real-life problem situation to a satisfactory solution can be quite long and full of complications. There are many factors that contribute to this, but one of the most important is the large amount of data that needs to be handled.

Because it is standard practice in industry to store information in databases, Mitra et al. (1995) argue that for mathematical programming to gain better acceptance as a modeling tool within corporate decision studies, a unified approach that integrates a modeling language with a relational database is necessary because it will provide a more powerful tool for constructing models that are truly data driven. They propose to achieve this by incorporating relational database structures into the syntax of the algebraic modeling language MPL (Maximal Software 1993).

The goal of our research has been to design an integrated modeling environment in which data management plays an even more central role and in which all modeling activities, such as model formulation, model instantiation, model solution, model validation, and solution analysis, are done in a common paradigm. Our efforts have resulted in a modeling environment based on relational algebra. Model formulation, instance generation, and solution manipulation are all done using relational operators such as selection, projection, and join. Many other desirable features of modeling environments, such as model management, instance management, and report writing are facilitated because they can be done using available relational database tools. Furthermore, model builder as well as end-user can work with the same system and users can easily share models. A prototype has been implemented to provide a "proof of concept." The prototype demonstrates that it is possible to develop a modeling environment for mathematical programming using a single paradigm: relational algebra.

Several other researchers have observed the potential of relational algebra for mathematical programming modeling.

Our research was partly motivated by the ideas presented by Johnson (1989). Choobineh (1991) designed SQLMP, an extension of SQL (Structured Query Language) (ANS, Date 1987) for mathematical programming that uses the algebraic paradigm for model conceptualization. We do not extend SQL and use the block-schematic paradigm for model conceptualization. The block-schematic paradigm is described by Baker (1983) and Welch (1987) and forms the basis of MathPro (1989) and MIMI (Baker 1992). The reason we use the block-schematic paradigm is purely pragmatic: The block-schematic approach is easier to embed in a relational modeling scheme. We are not claiming that the block-schematic paradigm is better than the algebraic approach used in systems such as GAMS (Brooke et al. 1988), AMPL (Fourier et al. 1993), MPL (Maximal Software 1993) and AIMMS (Bisschop and Entriken 1993). Which approach to use is largely a matter of taste, although it is claimed that many industrial users, particularly those in the process industry, prefer the block-schematic paradigm as it is closer to their "activity-based" view of the model. One disadvantage of the block-schematic paradigm is that it is mainly appropriate for linear optimization models. Dolk (1988) shows how structured models (as introduced by Geoffrion, 1987) can be represented and manipulated easily using SQL. Dolk also discusses how SQL might be used to facilitate the solution of mathematical programming models. He expects that this will require nonstandard SQL features. Our research shows that interfacing with an optimizer can be done completely with the tools used to access a relational database system. By using a standard interface language like SQL, the interface can be represented in a database-independent way, with some minor exceptions.

The paper is organized as follows. In §2, we present our view on the characteristics of good modeling environments. In §3, we give a brief introduction to relational operators. In §4, we introduce the basic concepts of relational modeling. In §5, we show how to use these basic concepts to model the fleet assignment problem. In §6, we illustrate how these concepts can be implemented in a standard database

environment. In §7, we present some conclusions. In the appendices, we give relational models for several well-known planning problems and an overview of the functionality of our prototype.

## 2. MODELING ENVIRONMENTS

A model is an abstraction of a real-life decision situation. Therefore, its solution has to be interpreted with care and not as the definitive answer to the real-life problem.

The process of generating a satisfactory solution to a real-life problem involves developing a model (which typically means making simplifying assumptions), generating an instance of the model (which typically involves gathering huge amounts of data), solving the instance (which typically involves transforming the instance data into a machine-readable form), validating the solution and the model (which involves verifying the appropriateness of the simplifying assumptions), and if the need arises, repeating these steps. In addition, models may have to be modified when changes occur in the real-life decision situation or user needs become different. This iterative process represents the *modeling life-cycle* in which a model evolves over time.

A computer based linear programming modeling environment has to nurture the entire modeling life-cycle. Such a system must facilitate the ongoing evolution of models and support the management of resources used in the modeling life-cycle, such as data, models, solvers, solutions.

Nowadays, highly accurate data gathering and processing technologies are widely available in industry. Typically, the availability of more and more accurate data leads to the development of more detailed models, which means that data management facilities in modeling environments are crucially important. Most of the data required for an instance of a model will be stored in corporate databases and must be processed before they can be used to construct an instance of the model at hand.

Typically, it is necessary to solve many instances of one model with varying data. Therefore, it is important that data and model are separated, i.e., the model should be stated independently from any data. Consequently, a modeling environment should support, if not enforce, the separation of model and instance.

These are only a few, though very important, features that an effective modeling environment should have. Other desirable features include support for model documentation and report writing, and the availability of different views, such as lists, schemas, figures, and charts of the model, the instance data, and the solution. The modeling concepts we propose are well suited to form the basis of a modeling system that has these desired features.

## 3. RELATIONAL OPERATORS

The relational modeling system we propose makes frequent use of the relational operators *selection*, *projection*, and *predicated join*. Relational operators are part of relational

algebra and formally operate on relations. However, for presentational convince, we think of a relation as a table (as is done in relational databases), and discuss these relational operators in terms of how they operate on tables. The selection operator constructs a new table by taking a horizontal subset of an existing table. The projection operator constructs a new table by taking a vertical subset of an existing table. The predicated join operator constructs a new table from two existing tables. Each row in the resulting table is formed by concatenating two rows, one from each of the original tables. Rows in the resulting table that do not satisfy the condition (or predicate) are eliminated.

SQL is the most popular data manipulation language that implements these operators. SQL statements have a natural and intuitive interpretation, and we will use SQL to present the concepts of our relational modeling system. We briefly introduce the SQL statements used in the design and implementation of our relational modeling system. For more detailed explanations on SQL we refer the reader to Date (1987) and Elmasri and Navathe (1994).

The CREATE TABLE statement is used to define and create a relational data table.

```
CREATE TABLE table (attribute attribute_type,
..., attribute attribute_type);
```

### EXAMPLES.

```
CREATE TABLE production (plant CHAR(10),
product CHAR(10), capacity NUMBER,
cost NUMBER);
CREATE TABLE demand (center CHAR(10),
product CHAR(10), amount NUMBER);
```

The fundamental operation in SQL is the *mapping*, represented syntactically as SELECT-FROM-WHERE. The mapping operation is effectively a horizontal subsetting (selection operator) followed by a vertical subsetting (projection operator). If the mapping operates on two tables, a predicated join is performed before the horizontal and vertical subsetting. More formally, a query to the database is formulated as follows:

```
SELECT attribute(s)
FROM table(s)
WHERE predicate(s)
```

In SQL each query is evaluated as follows. First the cartesian product of the tables in the FROM clause is computed. Then the result is filtered by the predicate(s) of the WHERE clause. The last step is the display of the results which were requested through the SELECT clause.

### EXAMPLE.

```
SELECT plant, center
FROM production, demand
WHERE production.product = demand.product
```

An extension to this basic query is obtained by using the GROUP BY operator. The GROUP BY operator rearranges the FROM table into partitions or groups. Within any one

group all rows have the same value for the GROUP BY field. The SELECT clause is then applied to each group of the partitioned table rather than to each row of the original table.

EXAMPLE.

```
SELECT product, SUM(amount)
FROM demand
GROUP BY product;
```

The last SQL construct used in the design and implementation of our relational modeling system is the *view*. A view is simply a particular look at the database. Although a view is a table, it does not exist physically in the database as a table; no storage space or data are allocated for it. The CREATE VIEW statement is used to define a virtual table.

EXAMPLE.

```
CREATE VIEW largecap (plant, product, capacity) AS
SELECT plant, product, capacity
FROM production
WHERE capacity > 1000;
```

## 4. RELATIONAL MODELING

We will illustrate the basic concepts of relational modeling by means of an example. We consider a production distribution problem with single sourcing requirements (Mairs et al. 1978).

### 4.1. Problem Situation

The problem is to decide how much of each product to produce at plants, how to ship to warehouses and transship to demand centers, subject to the constraint that a warehouse has to ship all of the demand for all products to any demand center to which it ships. In other words, each demand center is assigned a single warehouse that must meet all its demand for the several products.

### 4.2. Instance Data

The data involved in this model are production cost per product per plant, production capacity per product per plant, shipping cost from plant to warehouse, shipping cost from warehouse to demand center, and demand per product per demand center. These data are assumed to be available in the database in user data tables: Production, Shipcost, Tranship, and Demand. An instance of the production-distribution problem is given by the following user data tables. This instance will be used throughout our discussion of the basic concepts of our approach.

Table Production:

PLANT	PRODUCT	CAPACITY	COST
topeka	chips	200	230
topeka	nachos	800	280
newyork	chips	600	255

Table Shipcost:

PLANT	WHSE	COST
topeka	topeka	1
topeka	newyork	45
newyork	topeka	45
newyork	newyork	2

Table Tranship:

WHSE	CENTER	COST
topeka	east	60
topeka	south	30
topeka	west	40
newyork	east	10
newyork	south	30
newyork	west	80

Table Demand:

CENTER	PRODUCT	AMOUNT
east	chips	200
east	nachos	50
south	chips	250
south	nachos	180
west	chips	150
west	nachos	300

### 4.3. Column and Row Strips

In an integer linear program, activities or decisions are modeled as variables, possibly with integrality restrictions on some of them, and restrictions and relations among the decisions are modeled as linear equations and inequalities in terms of the variables. Typically, variables in an integer linear program can be grouped into classes with similar characteristics, based on what they represent in the underlying problem situation. Similarly, the linear equations and inequalities, or constraints, can also be grouped into classes with similar characteristics. These classes of variables and classes of constraints can be used to construct a block-schematic view of the integer linear program, see for instance Welch (1987). In a block-schematic view, classes of variables are called *column strips*, classes of constraints are called *row strips*, and their intersections, where interactions occur, are called *blocks*.

There are three types of decisions (classes of variables) in our production distribution model:

- How much to produce of each product at a plant?
- How much of each product to ship from a plant to a warehouse?
- Which warehouse to assign to each demand center?

Each type of decision will be represented by a class of variables in the model and by a column strip in the block-schematic representation. Because there are three types of decisions, there will be three column strips: Produce, Ship, and Assign.

Because each decision is related to a specific subset of the data, e.g., we have to determine how much to produce for each combination of a plant and a product, we can think of column strips as selections of data, and we can thus define them using the SQL construct of a view.

Ultimately, a modeling system has to prepare a machine-readable form of an instance for an optimizer. Therefore, any modeling system has to create a representation of the coefficient matrix that is understood by the optimizer. This is typically done in the form of triplets, one for each nonzero coefficient, consisting of a column (or variable) index, a row (or constraint) index, and the value of the coefficient. To accommodate the generation of such a representation, we store an index with each variable and each constraint.

Because the relational database paradigm does not impose an ordering on the rows of a table, it is nontrivial to create such an index. Fortunately, most commercial implementations of SQL provide a pseudo-column that contains a number indicating the sequence in which a row was selected, and we will make use of this pseudo-column to create our index. In Oracle this pseudo-column is called `RowNum`, in Informix it is called `RowId`, whereas in Microsoft Access it is called `AutoNumber`. Because this is not a standard SQL feature and different vendors use different names for this feature, our relational modeling system has to be adapted depending on the database management system used for the implementation. In the remainder we assume that we are working with Oracle and use the `RowNum` feature to define the unique indices for each row and column strip.

Throughout the paper, we carry the indices in the tables explicitly, because they are important to convey the underlying concepts of our relational modeling system. However, in a commercial implementation based on the relational concepts we present, these indices can be hidden from the user because their generation and manipulation can be done automatically.

```
CREATE VIEW Produce (ix, plant, product) AS
SELECT RowNum, plant, product
FROM Production;
```

```
CREATE VIEW Ship (ix, plant, whse, product) AS
SELECT RowNum, plant, whse, product
FROM Production, Shipcost
WHERE Production.plant = Shipcost.plant;
```

```
CREATE VIEW Assign (ix, whse, center) AS
SELECT RowNum, whse, center
FROM Tranship;
```

The definition of `Ship`, for example, indicates that there will be a variable for each combination of a plant, a warehouse, and a product and that these combinations can be obtained from the user data tables `Production` and `Shipcost`. Based on the data tables of the instance specified above, the column strip `Ship` defines the following variables

plus associated indices.

IX	PLANT	WHSE	PRODUCT
1	topeka	topeka	chips
2	topeka	newyork	chips
3	topeka	topeka	nachos
4	topeka	newyork	nachos
5	newyork	topeka	chips
6	newyork	newyork	chips

Similarly, there are three types of relations and restrictions in the model:

- Production at a plant is linked to shipping from the plant to a warehouse, i.e., everything that is produced should be shipped to some warehouse.
- Enforcement of the product flow balance, i.e., the total amount of a product shipped from plants to a warehouse should equal the total amount of a product shipped to the centers.
- Enforcement of the single sourcing requirement, i.e., each center receives all its demand from a single warehouse.

Each class of constraints will be represented by a row strip in the block-schematic representation. Because there are three classes of constraints, there will be three row strips: `Prodrow`, `Shiprow`, and `Centrow`.

```
CREATE VIEW Prodrow (ix, plant, product) AS
SELECT RowNum, plant, product
FROM Production;
```

```
CREATE VIEW Shiprow (ix, whse, product) AS
SELECT RowNum, whse, product
FROM Tranship, Demand
WHERE Tranship.center = Demand.center
GROUP BY whse, product;
```

```
CREATE VIEW Centrow (ix, center) AS
SELECT RowNum, center
FROM Demand;
```

Based on the data tables of the instance specified above, the row strip `Shiprow` defines the following constraints plus associated indices:

IX	WAREHOUSE	PRODUCT
1	topeka	chips
2	topeka	nachos
3	newyork	chips
4	newyork	nachos

Observe that we do not specify the number of variables in a class or the number of constraints in a class. The size of an instance is not part of the model but is determined automatically by the number of records in the user data tables.

#### 4.4. Blocks

So far, we have defined the column strips and row strips of the matrix, i.e., the classes of variables and the classes of constraints of the model. Next, we have to determine

whether a class of variables interacts with a class of constraints, i.e., whether there are nonzero entries in the block defined by the associated column and row strips. This gives the blocks with the technological coefficients of the matrix.

```
CREATE VIEW Block11 (rowix, colix, coef) AS
SELECT Prodrow.ix, Produce.ix, -1
FROM Prodrow, Produce
WHERE Prodrow.product = Produce.product
AND Prodrow.plant = Produce.plant;
```

```
CREATE VIEW Block12 (rowix, colix, coef) AS
SELECT Prodrow.ix, Ship.ix, 1
FROM Prodrow, Ship
WHERE Prodrow.product = Ship.product
AND Prodrow.plant = Produce.plant;
```

```
CREATE VIEW Block22 (rowix, colix, coef) AS
SELECT Shiprow.ix, Ship.ix, -1
FROM Shiprow, Ship
WHERE Shiprow.product = Ship.product
AND Shiprow.whse = Ship.whse;
```

```
CREATE VIEW Block23 (rowix, colix, coef) AS
SELECT Shiprow.ix, Assign.ix, amount
FROM Shiprow, Assign, Demand
WHERE Shiprow.product = Demand.product
AND Shiprow.whse = Assign.whse
AND Assign.center = Demand.center;
```

```
CREATE VIEW Block33 (rowix, colix, coef) AS
SELECT Centrow.ix, Assign.ix, 1
FROM Centrow, Assign
WHERE Centrow.center = Assign.center;
```

The definition of Block23, for example, indicates that there will be a nonzero coefficient for each product that is shipped from a warehouse to a demand center and that the value of this coefficient is equal to the demand at this demand center, which can be found in the user data table Demand.

As an example, we show the intermediate table that is implicitly generated during the construction of the virtual table Block23 just before the final selection of rowix, colix, and coef is made.

WHSE	PRODUCT	CENTER	COEF	COLIX	ROWIX
topeka	chips	east	200	1	1
topeka	chips	south	250	2	1
topeka	chips	west	150	3	1
topeka	nachos	east	50	1	2
topeka	nachos	south	180	2	2
topeka	nachos	west	300	3	2
newyork	chips	east	200	4	3
newyork	chips	south	250	5	3
newyork	chips	west	150	6	3
newyork	nachos	east	50	4	4
newyork	nachos	south	180	5	4
newyork	nachos	west	300	6	4

Observe that each block defines a set of triplets specifying the nonzero coefficients of that block, and that all triplets are specified relative to that block. Therefore, to specify the

complete matrix all we have to do is impose an ordering on the column strips and row strips and add the appropriate offsets to the row and column indices appearing in the triplets.

It is convenient for us to consider information pertaining purely to a class of variables, such as objective coefficients, lower, and upper bounds, and information pertaining purely to a class of constraints, such as lower and upper bounds, as blocks as well. Because this type of information is typically referred to as belonging to the rim of the matrix, we will sometimes refer to these blocks as *rim blocks*. Note that we specify constraints using lower and upper bounds on the activity instead of using a sense and a right-hand side.

Below are the definitions of the rim blocks. Because these blocks will be part of the matrix description that will be input to an integer linear programming optimizer, we create triplets.

```
CREATE VIEW ProduceObj (rowix, colix, coef) AS
SELECT null, ix, cost
FROM Produce, Production
WHERE Produce.plant = Production.plant
AND Produce.product = Production.product;
```

```
CREATE VIEW ShipObj (rowix, colix, coef) AS
SELECT null, ix, cost
FROM Ship, Shipcost
WHERE Ship.plant = Shipcost.plant
AND Ship.whse = Shipcost.whse;
```

```
CREATE VIEW AssignObj (rowix, colix, coef) AS
SELECT null, ix, SUM(amount) * cost
FROM Assign, Demand, Tranship
WHERE Assign.center = Tranship.center
AND Assign.whse = Tranship.whse;
AND Assign.center = Demand.center;
GROUP BY cost;
```

```
CREATE VIEW ProduceUp (rowix, colix, coef) AS
SELECT ix, null, capacity
FROM Produce, Production
WHERE Produce.product = Production.product;
```

```
CREATE VIEW ProdrowLo (rowix, colix, coef) AS
SELECT ix, null, 0
FROM Prodrow;
```

```
CREATE VIEW ProdrowUp (rowix, colix, coef) AS
SELECT ix, null, 0
FROM Prodrow;
```

```
CREATE VIEW ShiprowLo (rowix, colix, coef) AS
SELECT ix, null, 0
FROM Shiprow;
```

```
CREATE VIEW ShiprowUp (rowix, colix, coef) AS
SELECT ix, null, 0
FROM Shiprow;
```

```
CREATE VIEW CentrowUp (rowix, colix, coef) AS
SELECT ix, null, 1
FROM Centrow;
```

```
CREATE VIEW CentrowLo (rowix, colix, coef) AS
SELECT ix, null, 1
FROM Centrow;
```

The definition of ProduceObj indicates that for each combination of a plant and a product defined in the column strip Produce, the objective coefficient can be found in the Production user data table in the field COST of the row that matches this particular combination. The definition of AssignObj shows that it is also possible to have computed objective coefficients. This completes the model description. A block-schematic view of the model is given here.

	RowLo	Produce	Ship	Assign	RowUp
Objective ColumnLo		ProduceObj 0	ShipObj 0	AssignObj 0	
ProdRow	0 ≤	Block11	Block12		≤ 0
ShipRow	0 ≤		Block22	Block23	≤ 0
CentRow	1 ≤			Block33	≤ 1
ColumnUp		ProduceUp	Inf	Inf	

Observe that the definition of column strips, row strips, and blocks depends only on the structure of the user data tables, *not* on the records contained in those tables. This ensures complete separation of model and data. It also means that the same model definition can handle instances with two plants, two products, two warehouses, and three demand centers, as well as instances with hundreds of plants, thousands of products, hundreds of warehouses, and millions of demand centers. Observe that an instance of a model exists as a collection of views, i.e., virtual tables. This is a major difference from systems in which instances are physically stored in a database.

In defining the column strips, row strips, and blocks of the production distribution model, we have presented the required SQL statements in full detail. This was done to convey how a relational modeling system can be implemented. However, a lot of information contained in the SQL statements does not pertain to the model, but is a result of the SQL syntax. In any commercial implementation of a modeling environment based on relational modeling concepts, a user interface needs to be developed that shields a user from the underlying SQL syntax and indexing, reducing the effort required to specify a model. In such a system, the user would specify a block simply by using the names of the column and row strips and the corresponding relationship between the attributes of the strips. The actual detailed SQL queries would be created automatically by the system.

#### 4.5. Ordered Domains

An important class of linear programming models involves multiperiod production planning. Such models typically contain a class of balancing constraints that ensure a proper transition from one period to the next, e.g., for every period except the first, the inventory at the start of period  $t - 1$  plus the production in period  $t - 1$  minus the sales in period  $t - 1$  has to equal the inventory at the start of period  $t$ . Such models pose a serious problem for the relational modeling approach because it relies on a natural ordering of the data, such as weeks, months, and years.

The relational model that forms the basis of relational database implementations does not support the concept of ordered domains. There are two ways to deal with this dilemma. First, commercial implementations of a relational database have special functions related to time and we could make use of these. Second, when building a model, we can use numerical representations of the ordered domains and use SQL constructs to implement ordering concepts such as “first,” “successor,” and “predecessor.”

As an example, consider the following two user data tables. The first table is not necessary but mainly serves as a table that can be used in the report generation phase.

Table Date:

NAME	PERIOD
February	2
April	4
June	6

Table Production:

PRODUCT	PERIOD	CAPACITY	COST
chips	2	2000	76
chips	4	1600	78
chips	6	2000	76
nachos	2	1200	82
nachos	4	1200	82
nachos	6	800	86

Consider the class of balancing constraints mentioned above. The column strip associated with the inventory variables can be defined as

```
CREATE VIEW Inventory (ix, product, period) AS
SELECT RowNum, product, period
FROM production;
```

We require a balance constraint for each product in each period except for the first. The row strip associated with the balancing constraints can be defined as

```
CREATE VIEW Balance (ix, product, period) AS
SELECT RowNum, product, period
FROM production
WHERE period > (SELECT min(period)
FROM production);
```

We have used a subquery to determine the first period appearing in the table production. For this class of constraints, there are two interactions between the column strip and the row strip — there is product flow “into” the period and “out” of the period. To accomplish this, we simply define the two matrix blocks:

```
CREATE VIEW Block_1 (rowix, colix, coef) AS
SELECT Balance.ix, Inventory.ix, -1
FROM Balance, Inventory
WHERE Balance.product = Inventory.product
AND Balance.period = Inventory.period;
```

```
CREATE VIEW Block_2 (rowix, colix, coef) AS
SELECT Balance.ix, Inventory.ix, 1
FROM Balance, Inventory
WHERE Balance.product = Inventory.product
AND Inventory.period =
  (SELECT max(period) FROM Inventory
   WHERE period < Balance.period);
```

Again, we have used a subquery, this time to determine the previous period appearing in the production table. This concludes our description of the basic concepts of relational modeling. In Appendix A, we provide examples of relational models for some well-known planning problems.

## 5. THE FLEET ASSIGNMENT PROBLEM

To provide the reader with a real-life application in which using the relational modeling scheme is a natural and convenient choice, we discuss an important planning problem faced by the airline companies. In the *fleet assignment problem*, we are given a flight schedule and a set of fleet (aircraft) types. The problem is to find a minimum cost assignment of the fleet types to the flight legs in the schedule (see Abara 1989 for a discussion and overview of the fleet assignment problem). The flight schedule (time table) of a medium to large airline is huge and is typically stored in a relational database. Below we show some of the typical information found in the database.

Table Schedule:

LEG	DEPSTA	DEPTIM	ARRSTA	ARRTIM	FLEET	COST	...
101	DFW	745	BOS	1055	734	8270	
101	DFW	745	BOS	1055	757	11088	
101	DFW	745	BOS	1055	767	12098	
102	BOS	1200	DFW	1500	734	8270	
102	BOS	1200	DFW	1500	757	11088	
102	BOS	1200	DFW	1500	767	12098	
201	DFW	745	SFO	1145	734	9653	
...							

In this particular example, we see that there is flight from Dallas/Fort Worth International Airport to Logan Airport in Boston departing at 7:45 AM and arriving at 10:55 AM and that this flight can be flown by either a Boeing 734 at a cost of 8270, a Boeing 757 at a cost of 11088, or a Boeing 767 at a cost of 12098.

The information on available fleet types is also stored in a relational database and looks somewhat like this:

Table Fleets:

FLEET	AVAIL	...
-----	-----	
734	6	...
757	2	...
767	9	...
...		

A feasible fleet assignment must assign a fleet type to each flight leg, cannot use more aircrafts of a given fleet type than are available, and must ensure that aircrafts that arrive at a station either depart or stay on the ground and similarly that aircrafts depart from a station where they landed earlier.

To model the fleet assignment problem, we introduce two classes of variables: (1) a class of binary variables indicating for each combination of flight leg and fleet type whether or not the fleet type is assigned to the flight leg, and (2) a class of integer variables counting the number of aircrafts of a specific fleet type on the ground at a particular station and time.

```
CREATE VIEW Assign (ix, leg, fleet) AS
SELECT RowNum, leg, fleet
FROM Schedule;
```

```
CREATE VIEW GroundArc (ix, fleet, station, time) AS
SELECT RowNum, Fleets.fleet, arrsta, arrtime
FROM Fleets, Schedule
UNION
SELECT RowNum, Fleets.fleet, depsta, deptime
FROM Fleets, Schedule;
```

Note that we create a so-called *ground arc* for each fleet type at each station and each event (arrival or departure) at that station. Ground arcs, as the name suggests are used to model aircrafts that stay on the ground at a station.

There are three classes of constraints. First, which is at the heart of the problem, we have to ensure that each flight leg is assigned exactly one fleet type.

```
CREATE VIEW Cover (ix, leg) AS
SELECT RowNum, leg
FROM Schedule
GROUP BY leg;
```

Second, we have to ensure that the fleet assignments are consistent, in the sense that there is flow balance for each fleet type at each station and each event occurring at that station. That is to say that for each fleet type, the number of aircrafts of that fleet type “arriving” at a station (either via an incoming flight leg or via a ground arc) is equal to the number of aircrafts “departing” from the station (either via an outgoing flight leg or a ground arc).

```
CREATE VIEW Balance (ix, fleet, station, time) AS
SELECT ALL
FROM GroundArc;
```

Finally, we have to ensure for each fleet type that we do not use more aircrafts than there are available.

```
CREATE VIEW AircraftCount (ix, fleet) AS
SELECT RowNum, fleet
FROM Fleets;
```

Now we define the blocks of the matrix. Because we have to assign exactly one fleet type to each flight leg, we have

the following block

```
CREATE VIEW CoverBlock (rowix, colix, coef) AS
SELECT Cover.ix, Assign.ix, 1
FROM Cover, Assign
WHERE Cover.leg = Assign.leg;
```

Next, we consider the blocks that define the flow balance constraints for each fleet type at each event occurring at a station. This involves selecting appropriate incoming and outgoing arcs. We create a block handling the outgoing flight legs first.

```
CREATE VIEW BalanceAssignDep (rowix, colix, coef) AS
SELECT Balance.ix, Assign.ix, 1
FROM Balance, Assign
WHERE Assign.leg = SELECT leg FROM Schedule
      WHERE Balance.fleet = Schedule.fleet
      AND Balance.station = Schedule.depsta
      AND Balance.time = Schedule.deptime;
```

Next, we create a block handling the incoming flight legs.

```
CREATE VIEW BalanceAssignArr (rowix, colix, coef) AS
SELECT Balance.ix, Assign.ix, -1
FROM Balance, Assign
WHERE Assign.leg = SELECT leg FROM Schedule
      WHERE Balance.fleet = Schedule.fleet
      AND Balance.station = Schedule.arrsta
      AND Balance.time = Schedule.arrrtime;
```

Now, we switch to ground arcs and start with outgoing ground arcs.

```
CREATE VIEW BalanceGroundOut (rowix, colix, coef) AS
SELECT Balance.ix, Ground.ix, 1
FROM Balance, Ground
WHERE Balance.fleet = Ground.fleet
AND Balance.station = Ground.station
AND Balance.time = Ground.time;
```

We follow with the incoming ground arcs. Selecting the incoming arcs is a bit more involved, since it requires the identification of the previous event at a station, which is done by means of a subquery. Furthermore, we need to distinguish the first event at a station from the other events because the previous event of the first event actually occurs as the last event occurring (of the previous day).

```
CREATE VIEW BalanceGroundIn (rowix, colix, coef) AS
SELECT Balance.ix, Ground.ix, -1
FROM Balance, Ground
WHERE Balance.fleet = Ground.fleet
AND Balance.station = Ground.station
AND Ground.time =
  SELECT max(Ground.time) FROM Ground
  WHERE Balance.fleet = Ground.fleet
  AND Balance.station = Ground.station
  AND Ground.time < Balance.time;
```

```
CREATE VIEW BalanceGroundInFirst (rowix, colix, coef) AS
SELECT Balance.ix, Ground.ix, -1
FROM Balance, Ground
WHERE Balance.fleet = Ground.fleet
AND Balance.station = Ground.station
AND (Balance.time =
  SELECT min(Balance.time) FROM Balance
  WHERE Balance.fleet = Ground.fleet
  AND Balance.station = Ground.station)
AND Ground.time =
  SELECT max(Ground.time) FROM Ground
  WHERE Balance.fleet = Ground.fleet
  AND Balance.station = Ground.station;
```

To ensure for each fleet type that we do not use more aircrafts than there are available, we take a snapshot at midnight and count all the aircrafts of a specific fleet type. Because the balancing constraints ensure that the flow is a circulation the number of aircrafts in use will be the same throughout the day and taking a snapshot at midnight suffices. There are two blocks: one to account for the “red eye” flights, i.e., flight legs corresponding to flights that are in the air at midnight, and one for the ground arcs that cross midnight.

```
CREATE VIEW RedEyeCount (rowix, colix, coef) AS
SELECT PCount.ix, Assign.ix, 1
FROM PCount, Assign
WHERE PCount.fleet = Assign.fleet
AND Assign.leg = SELECT leg FROM Schedule
      WHERE Schedule.arrrtime < Schedule.deptime;
CREATE VIEW GroundCount (rowix, colix, coef) AS
SELECT PCount.ix, Ground.ix, 1
FROM PCount, Ground
WHERE PCount.fleet = Ground.fleet
AND (PCount.station, Ground.time) IN
  SELECT station, MAX(time) FROM Ground
  GROUP BY station;
```

Specifying the rim blocks, i.e., objective function coefficients and lower and upper bounds on variables and constraints is straightforward.

## 6. A RELATIONAL MODELING SYSTEM

The preceding sections have shown the conceptual viability of modeling linear and integer programs using a relational scheme. In this section, we describe the design of a modeling environment that supports the relational modeling paradigm. More specifically, we show how to implement model management, instance management, and solver management using relational database tools. These management tasks are performed by the modeling environment and are typically invisible to a user of the system.

### 6.1. Model Management

In the block schematic approach, a mathematical programming model is specified entirely in terms of row strips, column strips, and matrix blocks. In a relational database environment we can conveniently manage this information

for many models. We create four “system” tables: `SysModels`, `SysRows`, `SysCols`, and `SysBlocks` that contain all the information about the views defining the various models.

The `SysModels` table contains the names of the models present in the system. It has attributes `Model` and `ObjSense`. The `Model` attribute is the unique name of a model and the `ObjSense` attribute is `MAX` or `MIN` indicating whether the specified model is a maximization or minimization problem.

The `SysRows` table contains the names of the row strips present in the system. It has attributes `Model` and `RowStrip`. The `Model` attribute is the unique name of a model and the `RowStrip` attribute is the unique name of a view defining a row strip of the specified model.

The `SysCols` table contains the names of the column strips present in the system. It has attributes `Model`, `RowStrip`, and `Type`. The `Model` attribute is the unique name of a model, the `ColStrip` attribute is the unique name of a view defining a column strip of the specified model, and the `Type` attribute is `CONTINUOUS`, `BINARY`, or `INTEGER` indicating the variable type associated with the specified column strip.

The `SysBlocks` table contains the names of the blocks present in the system. It has attributes `Model`, `Block`, `RowStrip`, `ColStrip`, and `Type`. The `Model` attribute is the unique name of a model, the `Block` attribute is the unique name of a matrix block of the specified model, the `RowStrip` attribute is the unique name of the view defining the row strip associated with the specified block, the `ColStrip` attribute is the unique name of the view defining the column strip associated with the specified block, and the `Type` attribute is `ROWLOWER`, `ROWUPPER`, `COLOBJ`, `COLLOWER`, `COLUPPER`, or `BLOCKDATA` indicating the type of the specified block.

The tables below show the relevant entries in the system tables pertaining to the production distribution model presented in the preceding sections.

Table `SysModels`:

Model	ObjSense
Prod-Dist	MIN
...	...

Table `SysRows`:

Model	RowStrip
Prod-Dist	Prodrow
Prod-Dist	Shiprow
Prod-Dist	Centrow
...	...

Table `SysCols`:

Model	RowStrip	Type
Prod-Dist	Produce	CONTINUOUS
Prod-Dist	Ship	CONTINUOUS
Prod-Dist	Assign	BINARY
...	...	...

Table `SysBlocks`:

Model	Block	RowStrip	ColStrip	Type
Prod-Dist	Block11	Prodrow	Produce	BLOCKDATA
Prod-Dist	Block12	Prodrow	Ship	BLOCKDATA
Prod-Dist	Block22	Shiprow	Ship	BLOCKDATA
Prod-Dist	Block23	Shiprow	Assign	BLOCKDATA
Prod-Dist	Block33	Centrow	Assign	BLOCKDATA
Prod-Dist	ProduceObj		Produce	COLOBJ
Prod-Dist	ShipObj		Ship	COLOBJ
Prod-Dist	AssignObj		Assign	COLOBJ
Prod-Dist	ProduceUp		Produce	COLUPPER
Prod-Dist	ProdrowUp	Prodrow		ROWUPPER
Prod-Dist	ProdrowUp	Prodrow		ROWLOWER
...	...	...	...	...

## 6.2. Instance Management

In the production distribution problem used to illustrate the relational modeling scheme, we have used specific data tables in the definition of the model—e.g., `Production`, `Shipcost`, `Tranship`, and `Demand`—even though we used only the structure of these tables. It is good practice, however, to define models completely independent of its instances. To do so, we make use of another feature of SQL called a *synonym*. A synonym is an alias assigned to a table or view that may thereafter be used to refer to it. For each data table required in the definition of a model, we introduce a synonym and all references to data tables are made through these synonyms. Then, to create a specific instance of a model, all that needs to be done is to update the synonyms so that they refer to the actual data tables defining the instance.

The relational database environment is well suited to manage many instances of the same model. We create two system tables: `SysDataTables` and `SysInstances`.

The `SysDataTables` table contains the names of the active user data tables for a model. It has attributes `Model`, `BaseTable`, `Syn`, and `ActiveTable`. The `Model` attribute is the unique name of a model, the `Block` attribute is the unique name of a matrix block, the `BaseTable` attribute is the unique name of a special data table, called base table, having the same structure as an instance data table required in the definition of the specified model, the `Syn` attribute is the synonym for the base table used in the model definition, and the `ActiveTable` attribute is the name of the current data table associated with the specified synonym.

As mentioned before, the definition of a model depends only on the structure of the user data tables, not on the records contained in those tables. Therefore, to completely separate model and data, we use artificial tables in the model definition. The artificial tables, which we call base tables, have the same structure as the user tables, but will always be empty. The use of base tables also gives the system a level of error-checking. When we attempt to associate a synonym with a user specified instance data table, we can check if this table has the proper structure by comparing it to the base table associated with the synonym.

In our production distribution example, instead of using the tables `Production`, `ShipCost`, `Tranship`, and `Demand` directly in the definition of the model, we create (empty) base tables `base_Production`, `base_ShipCost`, `base_Tranship`, and `base_Demand` (with the same structure) and synonyms `syn_Production`, `syn_ShipCost`, `syn_Tranship`, and `syn_Demand` (initially pointing to the base tables), and use the synonyms in the definition of the model. Then to instantiate the model, we let the synonyms point to the real data tables.

Table SysDataTables:

Model	BaseTable	Syn	ActiveTable
Prod-Dist	base_Production	syn_Production	Production
Prod-Dist	base_ShipCost	syn_ShipCost	ShipCost
Prod-Dist	base_Tranship	syn_Tranship	Tranship
Prod-Dist	base_Demand	syn_Demand	Demand

The `SysInstances` table contains the names of the (user data) tables of an instance. It has attributes `Model`, `Instance`, `Syn`, and `DataTable`. The `Model` attribute is the unique name of a model, the `Instance` attribute is a unique name of an instance of the specified model, the `Syn` attribute is the name of the synonym used in the definition of the specified model, and the `DataTable` attribute is the name of the data table associated with the specified synonym in the specified instance.

Table SysInstances:

Model	Instance	Syn	DataTable
Prod-Dist	PD_January	syn_Production	Jan_Production
Prod-Dist	PD_January	syn_ShipCost	ShipCost
Prod-Dist	PD_January	syn_Tranship	Tranship
Prod-Dist	PD_January	syn_Demand	Jan_Demand
Prod-Dist	PD_February	syn_Production	Feb_Production
Prod-Dist	PD_February	syn_ShipCost	ShipCost
Prod-Dist	PD_February	syn_Tranship	Tranship
Prod-Dist	PD_February	syn_Demand	Feb_Demand
...	...	...	...

Finally, there are two system tables that contain solution information: `SysRuns` and `SysSols`. The `SysRuns` table has attributes `Model`, `Instance`, `Solver`, `RunDate`, `RunTime`, `Obj`, and `CpuTime`. The `Model` attribute is the unique name of a model, the `Instance` attribute is a unique name of an instance of the specified model, the `Solver` attribute indicates the solver used for the run, the `RunDate` attribute is the date of a particular run of the specified model for the specified instance, the `RunTime` attribute is the time of day that the run started, the `Obj` attribute is the objective function value obtained in the run, and the `CpuTime` attribute is the CPU time for the run.

Table SysRuns:

MODEL	INSTANCE	SOLVER	RUNDATE	RUNTIME	OBJ	CPUTIME
Prod-Dist	PD_January	OSL	1/1/98	13:14:15	324130	00:02:14
...	...	...	...	...	...	...

The `SysSols` table has attributes `Model`, `Instance`, `RunDate`, `RunTime`, `StripName`, and `TableName`. The `Model` attribute is the unique name of a model, the `Instance` attribute is a unique name of an instance of the specified model, the `RunDate` attribute is the date of a particular run of the specified model for the specified instance, the `RunTime` attribute is the time of day of the run, the `StripName` attribute is a row or a column strip name of the specified model, and the `TableName` attribute is the name of the data table containing the solution information for the specified row or column strip obtained in the run.

Table SysSols:

Model	INSTANCE	RUNDATE	RUNTIME	STRIPNAME	TABLERNAME
Prod-Dist	PD_January	1/1/98	13:14:15	Ship	sol_Ship
...	...	...	...	...	...

Table sol\_Ship:

PLANT	WHSE	PRODUCT	VALUE
topeka	topeka	chips	200
topeka	newyork	chips	0
topeka	topeka	nachos	480
topeka	newyork	nachos	50
newyork	topeka	chips	200
newyork	newyork	chips	200

Observe that the solution is put in a collection of tables. It is also possible, and in fact very easy, to put the solution immediately into the appropriate user data tables. We have chosen for the above design because it is more flexible and puts control in the hands of the user.

When the values of the decision variables have been returned to the user data tables, SQL provides a convenient tool for viewing the results of the optimization. In particular, one can easily scan subsets of the solution which may be of interest. For example, the production facility manager in Topeka can easily determine his production requirements and the total production cost by the following two queries (where we assume that the levels of production determined by the optimizer, i.e., the values of `Produce`, have been put in an additional field `amount` in the data table `Production`).

```
SELECT product, amount
FROM production
WHERE plant = 'topeka';
```

```
SELECT SUM(cost*amount)
FROM production
WHERE plant = 'topeka';
```

### 6.3. Solver Management

Another feature of the relational modeling environment that is easily incorporated is to allow users to vary solver parameters by defining a table `SysParams` to hold these parameters.

The `SysParams` table has attributes `Solver`, `Parameter`, and `Value`. The `Solver` attribute is the name of a solver, the

Parameter attribute is the name of a parameter that can be set for the specified solver, and Value is the current value of the parameter.

Table SysParams:

Solver	Parameter	Value
CPLEX	CPX_PARAM_CLIQUES	1
CPLEX	CPX_PARAM_NODELIM	1000000
OSL	rtolpinf	0.00001
...	...	...

The system tables introduced above form the basis of the prototype relational modeling system ARMOS described in the next section. It should be noted that the system tables are created only once at the installation of the system and then used by the system as an internal database of existing models, instances, and solutions. Maintaining the system tables is a responsibility of the system, not of the user of the system.

#### 6.4. A Prototype

To demonstrate the viability of the ideas and concepts presented in the previous sections, we have developed a small prototype system called ARMOS, A Relational Modeling System. ARMOS offers a simple user interface that is coded in Embedded SQL (Oracle Corp. 1992) and that allows a user to list the models stored in system, to load a model, to list the instances stored in the system for the loaded model, to make an instance active, to optimize the active instance, and to display solution values. The user can also display the model's matrix block structure, and view the coefficients of any particular matrix block of an active instance. Currently, both linear and mixed integer linear programs can be solved. ARMOS is built on top of the commercial software package OSL [DSV]. A brief description of its functionality can be found in Appendix B.

As mentioned above, ARMOS is a prototype. It provides only the most basic functionality and it only has a simple text-based user instance. There is no dedicated graphical editor supporting model development. Our goal in developing ARMOS was to verify the viability of using the relational modeling scheme in an actual implementation.

### 7. DISCUSSION

This paper provides a "proof of concept" demonstration, showing that it is possible to develop a modeling environment for mathematical programming using a single paradigm: relational algebra. We feel there are several advantages to such an approach. It is often observed, see for example Hürlihan (1991) that despite recent developments, mathematical programming is still not fully exploited in practice. By designing a modeling environment centered around database management systems that are widely used in industry, we believe it is possible to make mathematical programming accessible to a wider audience. Furthermore, it is easy to set things up in such a way that

using a model has a "fill-in-the-blank" feel, where solution values are immediately imported into the appropriate data tables. This will be very appealing to end-users, and will increase their acceptance level. (We have all witnessed the acceptance of spread-sheet like interfaces!) It has also been observed, see for example Mitra et al. (1995), that the data in corporate information systems are often regularly revised and that it is therefore desirable that a decision-making system monitors changes in instance data, notifies users that the current solution values are out of date, and allows users to update that solution value with a single command. Because in the relational modeling system we propose an instance exists only through virtual links rather than physically, this feature is naturally available.

Finally, a few words on how the proposed relational modeling system compares to other systems based on the block-schematic model building paradigm, such as MIMI and MathPro. Obviously, both MIMI and MathPro are, at the moment, far easier to use than our prototype system because of their more sophisticated graphical user interfaces. MIMI also has a much larger functionality, because it includes an expert system component for rule-based model solution. On the other hand, MIMI supports only two-dimensional tables and therefore requires a hierarchy of tables to represent high-dimensional strips or blocks. Our prototype does not impose this restriction because SQL easily handles tables with multiple fields. Furthermore, MIMI requires all relevant data to be in its own private internal database, which requires copying/transferring data from the corporate database to MIMI's database. Because the relational modeling system is built on top of the corporate database, it eliminates the data transfers and storage duplication.

### Appendix

Appendices A and B can be found at the Operations Research Home Page: (<http://or.pubs.informs.org/ops-search.html>) in the Online Collection.

### ACKNOWLEDGEMENTS

The authors thank the area editor, the associate editor, and the referees for many helpful suggestions that significantly improved the quality and presentation of the paper. This research was supported, in part, by NSF Grant DMI-9700285 to the Georgia Institute of Technology.

### REFERENCES

- Abara, J. 1989. Applying integer linear programming to the fleet assignment problem. *Interfaces* **19** 20–28.
- Atamtürk, A., E. L. Johnson, J. T. Linderoth, M. W. P. Savelsbergh. 1996. Using ARMOS, a relational modeling system.
- ANSI. 1986. *Database Language SQL, Document ANSI X3.135-1986*. (Also available as ISO document ISO/TC97/SC21/WG3 N117.)

- Baker, T. E. 1983. RESULT: An interactive modeling systems for planning and scheduling. Presented at the ORSA/TIMS meeting, Chicago, IL.
- . 1992. *MIMI/LP User Manual*. Chesapeake Decision Science, Inc.
- Bisschop J., R. Entriken, 1993. *AIMMS The Modeling System*. Paragon Decision Technology.
- Brooke, A., D. Kendrick, A. Meeraus. 1988. *GAMS, A User's Guide*. The Scientific Press, Redwood City, CA.
- Choobineh, J. 1991. SQLMP: A data sublanguage for representation and formulation of linear mathematical models. *ORSA J. Comput.* **3**(4) 358–375.
- Date, C. 1987. *A Guide to the SQL Standard*. Addison/Wesley, Reading, MA.
- Dolk, D. R. 1988. Model management and structured modeling: The role of an information resource dictionary system. *Comm. ACM* **31**(6) 704–718.
- Druckerman, J., D. Silverman, K. Viaropulos. *Optimization Subroutine Library Release 2 Guide and Reference*. IBM.
- Elmasri, R., S. Navathe. 1994. *Fundamentals of Data Base Systems*, 2nd ed, Benjamin/Cummings.
- Fourer, R., D. M. Gay, B. W. Kernighan. 1993. *AMPL. A Modeling Language for Mathematical Programming*. The Scientific Press, Redwood City, CA.
- Geoffrion, A. M. 1987. An introduction to structured modeling. *Management Sci.* **35**(5) 547–588.
- Greenburg, H. I., F. H. Murphy. 1992. A comparison of mathematical programming modeling systems. *Ann. Oper. Res.* **38** 177–238.
- Hürlimann, T. 1991. Linear modeling tools. Working Paper 187, Institute for Automation and Operations Research, University of Fribourg, Switzerland.
- Johnson, E. L. 1989. Modeling and strong linear programs for mixed integer programming. S. W. Wallace, ed. *Algorithms and Model Formulations in Mathematical Programming*. Springer-Verlag, Berlin, NATO ASI Series, Vol. F51, 1–43.
- MathPro, Inc. 1989. *MathPro Usage Guide: Introduction and Reference*.
- Maximal Software. 1993. *MPL Modeling System*.
- Mitra, G. B., Kristjansson, C. Lucas, S. Moody. 1995. Sets and indices in linear programming modelling and their integration with relational data models. *Comput. Optim. Appl.* **4** 263–292.
- Mairs, T. G., G. W. Wakefield, E. L. Johnson, K. Speilbergh. 1978. On a production allocation and distribution problem. *Management Sci.* **24** 1622–1630.
- Oracle Corporation. 1992. *Programmer's Guide to the ORACLE Precompilers*, Version 1.5.
- Wagner, H. 1975. *Principles of Operations Research*. Prentice-Hall, Englewood Cliffs, NJ.
- Welch, J. S. 1987. PAM—A practitioners' approach to modeling. *Management Sci.* **33**(5) 610–625.