AMOS2 Introduction

Gustav Fahl and Tore Risch
EDSLAB
Department of Computer Science
Linköping University
Sweden
torri@ida.liu.se
14 juni 1999

This is an introduction and tutorial to the AMOS2 object-relational database system. AMOS2 is a descendant of AMOS, which has its roots in a main memory version of Hewlett Packard's DBMS IRIS [1]. The entire database is stored in main memory and is saved on disk only through explicit commands. AMOS2 can be used a single user database, a multi-user server, or as a collection of interacting AMOS2 multi-database servers. In this overview only the single user version is described. AMOS2 is an example of an object-relational DBMS having a declarative query language, AMOSQL. The overview does not cover every feature of AMOS2. For a complete description of AMOS2's functionality, see [2].

1. Introduction

To be able to run the examples in the overview you need a directory (called *amos directory*) containing the files *amos2.exe* and *amos2.dmp*. *amos2.exe* is the executable program and *amos2.dmp* is an empty database that contains only the data needed by the AMOS2 system itself. All the commands that are used in the introduction are in the file *tutorial.amosql*. In order to try out the queries at the end of this overview one also needs the file *wcdata.amosql* and *wcdata.amosql*.

AMOS2 is started from the MS-DOS command window by making *amos directory* the current directory and then giving the Windows command:

amos2

You will then enter the AMOS2 top loop where you have the prompt:

Amos 1>

You are assumed to have general knowledge about databases and Entity Relationship modeling.

1.1 Styles used in the examples

This introduction uses the following style for AMOSQL commands that are to be entered to the AMOS2 top loop in the tutorial:

```
Amos 4> create function name(country)->charstring as stored;
```

AMOSQL commands that are not be entered but only to serve as examples look like this:

```
create type tournament properties (year integer, host country);
```

Thus the AMOSQL commands that are supposed to be executed during the tutorial have the AMOS2 prompt first.

2. An example

As a running example throughout the overview we use a database to manage information about World Cup tournaments in soccer. We start with an Entity Relationship schema (ER schema) that describes the information to be represented in the database. The steps how to implement the database in AMOS2 are then gradually introduced as the database is built. The ER schema is shown in Figure 1. In Section 2.1the corresponding database description is made in natural language.

2.1 Database description

World Cup tournaments in soccer are arranged every forth year. Each tournament is arranged by some country. In each tournament a number of teams participate and they represent various countries. Each tournament contains a number of games (matches) each played between two teams. The team that made most goals wins. Each match is refereed by a referee and a certain number of spectators come to look at it. Each team contains a set of players. For each match played by a team of players a subset of these players are chosen to play in this particular match. A player can make a number of goals and a number of own goals in a certain game.

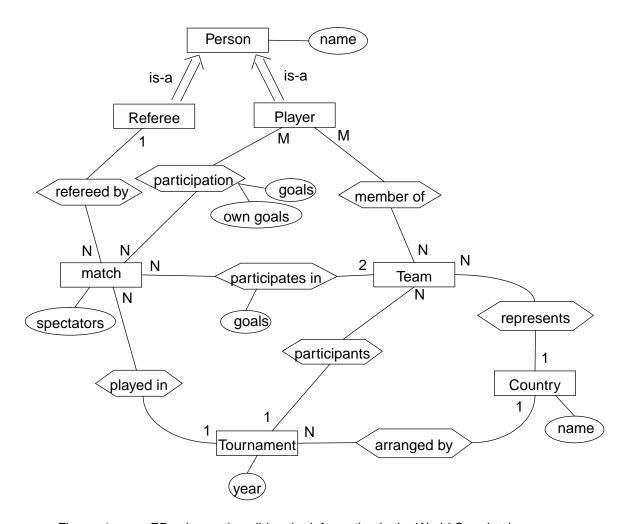


Figure 1 ER schema describing the information in the World Cup database.

3. Practical advises

3.1 Rollback

The database can be restored to an earlier state by the command:

```
rollback generation number;
```

For example, after the following command all changes from command 4 and forward are undone: rollback 4;

3.2 Reading commands from a file

AMOSQL commands can be read from a file and executed by the command

```
< `file name';
```

To read an execute all commands in the file tutorial.amosql, execute the command:

```
< 'tutorial.amosql';
```

In order to learn AMOSQL you are recommended to copy one command at the time from the file tutorial.amosql. Use cut and paste for this.

3.3 Saving a database

A database can be saved in the file named file name by the command:

```
save 'file name';
```

To start AMOS2 with the saved database, give the following command to the Windows command window:

```
amos2 file name
```

4. Types, functions, and objects

The three basic concepts in AMOS2 are *types*, *functions*, and *objects*. For those who are used to the terminology used with object oriented programming languages these concepts approximately correspond to *classes*, *methods*, and *instances*.

Types are used for classifying objects. Each object is an instance of a type. All properties of an object as well as relationships between objects are represented by functions.

4.1 Types

Entity types in an ER schema corresponds to *types* in AMOS2. We create types for the entity types *country* and *tournament* (see Figure 1) with the following AMOSQL commands:

```
AMOS 4>create type tournament;
AMOS 5>create type country;
```

4.2 Functions

The properties of objects and the relationships between objects are modeled as *functions*. In our example the entity type <code>tournament</code> has one attribute <code>year</code>. We model this as a function <code>year</code> that has a <code>tournament</code> as its argument and returns an integer:

```
AMOS 6>create function year(tournament)->integer as stored;
```

In the same way the attribute *name* of the entity type *country* becomes a function from countries to strings:

```
AMOS 7>create function name(country)->charstring as stored;
```

Between the entity types tournament and country there is the relationship arranged by. This is modeled by a function from tournaments to countries:

```
AMOS 8>create function host(tournament)->country as stored;
```

The above functions are defined as *stored functions* ('... as stored'). This means that the relationship between the argument and the result will be explicitly stored in the database. There are three kinds of functions in AMOS2. In addition to stored functions there are *derived functions* and *foreign functions*. Derived functions are introduced in Section 7.1. Foreign functions are defined in a conventional programming language, usually C or Lisp. Foreign functions are not discussed in this overview.

Types in AMOS2 can be *built in* or *user defined*. Examples of built in types are *integer*, *charstring*, and *real*.

User defined functions are defined with the *create type* statement, e.g.:

```
create type person;
```

Stored functions for a user defined type can be created together with the type definition. The commands 1-5 are equivalent to these commands:

```
create type country properties (name charstring);
create type tournament properties (year integer, host country);
```

4.3 Objects

We are now ready to create our first objects. The following command create three countries, i.e. three instances of the type *country*:

```
AMOS 9>create country instances :ita, :bra, :esp;
```

:ita, :bra and :esp are environment variables that are bound to the objects created. It is possible to refer to the objects through these variables throughout the AMOS2 session.

For the type *country* there is a function *name* defined. The following commands assign names to the three countries, i.e. defines the results of applying the function *name* on the three just created instances of type *country*:

```
AMOS 10>set name(:ita)='Italy';
AMOS 11>set name(:bra)='Brazil';
AMOS 12>set name(:esp)='Spain';
```

In the same way we create three instances of type tournament and define what values the functions year and host return when applied on the objects of type tournament:

```
AMOS 13>create tournament instances :t2, :t12, :t14;
AMOS 14>set year(:t2)=1934;
AMOS 15>set host(:t2)=:ita;
AMOS 16>set year(:t12)=1982;
AMOS 17>set host(:t12)=:esp;
AMOS 18>set year(:t14)=1990;
AMOS 19>set host(:t14)=:ita;
```

The results of the functions can also be defined directly when the objects are created. The commands 6-16 are equivalent with:

5. Query Language

Similarly to SQL, AMOSQL is a combination of a DDL (Data Definition Language) and a DML (Data Manipulation Language). The query part of the language is similar to SQL. The following command

answers the query 'What country arranged the World Cup 1982?':

An AMOSQL expression can be entered in one or several lines. The commands in AMOSQL are always terminated with semicolon (;).

The above query does not utilize the possibilities in AMOSQL to compose functions. The following command is equivalent to the previous one:

```
AMOS 21>select name(host(t))
from tournament t
where year(t)=1982;
```

The commands 17 and 18 actually answer the query 'What *name* did the country have that arranged the World Cup 1982?'. If you really want to retrieve the country object, not its name, the command would be:

```
AMOS 22>select host(t)

from tournament t

where year(t)=1982;
```

The following command answers the query 'What years did Italy arrange the World Cup?':

As the final query example the following command answers the query 'What years were the World Cup arranged and who were its hosts?':

```
AMOS 24>select year(t), name(host(t))
from tournament t;
```

6. More about types

6.1 Type Hierarchy

The types in AMOS2 are arranged in a *type hierarchy*. If a type A is defined as a subtype of a type B then the type A *inherits* all the functions defined on type B.

In our example the entity types referee and player are subtypes of the entity type person that has an attribute name.

```
AMOS 25>create type person;
AMOS 26>create type referee subtype of person;
AMOS 27>create type player subtype of person;
```

The attribute name is modeled as a function name from persons to strings:

```
AMOS 28>create function name(person)->charstring as stored;
```

The function name is now defined on the types person, referee, and player since referee

and player are subtypes of person.

If a type A is defined as a subtype of a type B it means that all instances of type A automatically become instances type B as well. To illustrate this, create the following instances of type person, referee, and player:

Then ask the following query and contemplate the answer:

```
AMOS 32>select name(p) from person p;
```

We now create the remaining types from Figure 1:

```
AMOS 33>create type team;
AMOS 34>create type match;
```

7. More about functions

7.1 Derived functions

A function can be defined in terms of other functions. Such a function is called a *derived function*. For example, if we often need to find out the name of the host country for a given year, we can define a derived function *host_name* for this:

When a query consists of a single function call, as the queries 33 and 34, the select can be omitted. The queries 33 and 34 can thus be written as:

```
host_name(1982);
host_name(1990);
```

7.2 Alternative function definitions

There are many different ways to model the information in an ER schema with functions. For example, consider the information that 'countries has names' in the ER schema of Figure 1 represented as an attribute *name* of the entity type *country*. One possibility is to model it as we did in command 4, i.e. as a function from countries to names (alternative 1):

```
create function name(country)->charstring as stored;
```

The database is then populated with the commands:

```
set name(:ita)='Italy';
set name(:bra)='Brazil';
```

```
set name(:esp)='Spain';
```

To get the name of a certain country (here the country object :ita) one makes the query:

```
name(:ita);
```

Another possibility is to model it as a function from names to countries (alternative 2):

```
create function country_named(charstring)->country as stored;
```

The database is populated with:

```
set country_named("Italy')=:ita;
set country_named("Brazil')=:bra;
set country_named("Spain')=:esp;
```

In order to get the name of the country object : ita one now has to ask the query:

```
select cn
from charstring cn
where country_named(cn)=:ita;
```

The way one chooses to model the information is always a judgement call. In this example alternative 1 is the most natural one but there are other cases where the choice is more arbitrary.

If one thinks that several functions are useful there is always the possibility to define one of them as a (stored) base function and then define the other as derived from the base function. In our example we can, e.g., define the function in alternative 1 as stored and then define the function in alternative 2 as derived:

```
create function country_named(charstring cn)->country c as
  select c where name(c)=cn;
```

7.3 More function definitions

We define some more functions from Figure 1:

```
AMOS 38>create function represents(team)->country as stored;
AMOS 39>create function refereed_by(match)->referee as stored;
AMOS 40>create function played_in(match)->tournament as stored;
AMOS 41>create function spectators(match)->integer as stored;
```

7.4 Set valued functions

We chose to model the relationship participants between the entity types team and tournament as a function from tournaments to teams:

```
create function participating_teams(tournament)-> team as stored;
```

participating_teams is an example of a function returning more than one value. In a tournament more than one team is participating. However, the definition above does not allow this.

When a stored function is defined AMOS2 by default assumes that the function returns *one* value, while the same result can be returned for different values of the argument. For example, the function <code>spectators(match)->integer</code> (command 68) it holds that one match can have one number as the number of spectators while several games can have the same number of spectators.

For the function <code>participating_teams</code> the opposite holds. For a certain value of the input parameter several values are returned (there are several teams participating in a tournament) while a certain result cannot be returned for more than one value of the input parameter (a team can only participate in one tournament). Such constraints on the cardinality of function results or arguments are called <code>cardinality constraints</code>.

In AMOS2 the cardinality constraints for a function are specified with the keywords *key* and *non-key*. The default for a stored function is that the first argument is specified as *key* while the remaining arguments and the result is *nonkey*¹. Command 38 is this equivalent with²:

```
create function spectators(match key)->integer nonkey as stored;
```

The correct definition of the function participating_teams is:

```
AMOS 42>create function participating_teams(tournament nonkey)-> team key as stored;
```

For set valued functions the command add is used for adding values to the results of function. The syntax is the same as for the set command:

To model the relationship player in between team and player (see Figure 1) we define a function from team to player. The relationship is many-to-many, i.e. a player can play in several teams and a team has several players. The function definition therefore becomes³:

```
AMOS 47>create function players(team nonkey)->player as stored;
AMOS 48>add players(:bra82)=:p1;
AMOS 49>add players(:bra82)=:p2;
AMOS 50>add players(:ita82)=:p3;
AMOS 51>name(players(:bra82));
```

7.5 Functions returning tuples

All functions so far have had a single result type. It is also possible to have functions that return tuples of values.

In command 25 the names of persons were defined by the function:

```
create function year(tournament)->integer key as stored;
create function name(country)->charstring key as stored;
to make it impossible to have two tournaments the same year or to have two countries with the same
```

^{1.} The cardinality constraints of the functions year(tournament) -> integer and name(country) -> charstring (defined with the commands 3 and 4) are not completely correct. They should actually have been defined as:

^{2.} The placement of *key* and *nonkey* may seem confusing. The easiest way to understand it is to regard the stored function as a table in a relational database. The column for match is then a primary key while the column for spectators is a non-key.

^{3.} The keyword *nonkey* is not needed after plays since *nonkey* is default for the result of a function.

```
create function name(person)->charstring as stored;
```

Another possibility could have been to separate the first and family names by defining a function that returns a *tuple* of two strings:

```
create function name2(person)-> <charstring, charstring> as stored;
```

A person can then be assigned a name with the command:

```
set name2(:p1)=<"Pavel', "Smerdjakov'>;
```

Functions for retrieving the first and family names, respectively, can then be defined as derived functions calling name 2:

```
create function first_name(person p)->charstring f as
  select f
  from charstring l
  where name2(p)=<f, l>;

create function last_name(person p)->charstring l as
  select l
  from charstring f
  where name2(p)=<f, l>;
```

7.6 ... and so the final function definitions

Now only two relationships from Figure 1 remain to be represented as functions: participation between player and match and participates in between team and match.

Let's start with the first relationship: 'Players are participants in games and makes a number goals and a number of own goals in them'.

The most interesting functions are 'What players participate in a certain match?', 'How many goals did a certain player make in a certain match?', and 'How many own goals did a certain player make in a certain match?':

```
create function participating_players(match m)->player p as ...
create function goals(player p, match m)->integer g as ...
create function own_goals(player p, match m)->integer og as ...
```

It is not recommended to define these functions as stored ones as it would create redundant data¹. We rather store the information about the participation of players in games in a function player_participation and define the three functions above as derived:

^{1.} For example the fact that a player participated in a certain match would have been stored in three different places.

^{2.} No variable names need to be specified after the types in the definitions of stored functions. However, for clarity we still specify the variable names goals and own_goals.

```
select g
    from integer og
    where player_participations(m)=<p, g, og>;

AMOS 55>create function own_goals(player p, match m)->integer og as
    select og
    from integer g
    where player_participations(m)=<p, g, og>;
```

We store new data in the database and make some queries:

The relationship 'Teams participate in games and make a number of goals in them' is implemented accordingly. The participation of teams in games is stored in the function <code>team_participation</code>. The two most interesting functions are 'What teams participate in a certain match?' and 'How many goals did a certain team make in a certain match?'. They are defined as derived functions:

7.7 Aggregation functions

Aggregation functions compute aggregate values over sets:

```
sum(...) - returns the sum of a number of values.

count(...) - counts the number of values.

maxagg(...) - returns the highest value.

minagg(...) - returns the lowest value.

some(...) - equivalent with count(...) > 0 but much faster.

notany(...) - equivalent with count(...) = 0 but much faster.
```

These functions have somewhat different semantics compared to 'ordinary' AMOS2 functions.

Consider the query 'What players participated in the tournament :t12?';

```
AMOS 65>players(participating_teams(:t12));
```

The function participating_teams is applied on the object :t12. With the data stored in the database at this point participating_teams returns two objects, :bra82 and :ita82.

Then the function players is applied on each one of the two objects returned from the call participating_teams, :bra82 and :ita82. players(:bra82) returns two objects, :p1 and :p2. players(:ita82) returns a single object, :p3.

Thus the result of the complete AMOSQL query above is three objects, :p1, :p2, and :p3.

Compare this with the query 'How many teams participated in tournament :t12?':

```
AMOS 66>count(participating_teams(:t12));
```

In this case the function *count* is applied once on the entire set of objects from *participating_teams*, not once per result from *participating_teams*. The reason is that *count* is defined to take a *bag of objects* as argument:

```
create function count(bag of object x)->integer r as ...
```

Nested queries can be used as arguments to aggregation functions. The following command answers the query 'How many own goals have been made?':

```
AMOS 67>sum((select own_goals(p, m) from player p, match m));
```

Notice that nested queries require an additional pair of parentheses.

8. Miscellaneous

8.1 System functions

Several useful functions are predefined in AMOS2. See [2] for a complete list of these. Here we only give some examples of how they can be used:

```
AMOS 68>plus(7,4);
AMOS 69>7+4;
```

Most arithmetic functions can be written either in prefix or infix notation. Other examples of arithmetic functions are minus (-), times (*), div (/), and mod.

```
AMOS 70 > max(7, 4);
```

max and min returns the largest and the smallest of two objects, respectively. This is different from maxagg and minagg that return the largest and smallest object from a bag of objects. Notice that comparison is defined not only for numbers, but actually for any kind of object. For example:

All comparison functions (<, >, =, !=, >=, and <=) are written in infix notation.

The comparison functions can be applied on numbers, strings, and any kind of objects. The result of comparing two arbitrary OIDs is, however, undefined (actually the internal OID numbers are then compared).

```
AMOS 72>typenamed('player');

returns the type having the name 'player'.

AMOS 73>cardinality(typenamed('player'));

returns the number of objects of the type named 'player'.

AMOS 74>allfunctions();

returns all currently defined functions.
```

9. More about the query language

9.1 Environment variables

It is possible to assign the value of a query to an environment variable.

```
AMOS 75>select c into :c
     from country c, tournament t
     where c=host(t)
         and year(t)=1982;
AMOS 76>name(:c);
```

9.2 Examples of queries

In the file 'wcdata.amosql' there are AMOSQL commands that populate the database. Read and execute these commands with these commands:

```
AMOS 77>< 'wcdata.amosql';
```

Next follows some examples of somewhat more complicated queries:

1. 'In what tournaments did Sweden participate?'

2. 'How many games have been played by Sweden?'

```
AMOS 79>count((select m from match m where name(represents(participating_teams(m)))='Sweden'));
```

3. 'What is the total number of goals made by Sweden?'

4. 'In what games were most goals made?'

A good start is to define a derived function goals that returns the total number of goals in a match. We make a first attempt:

```
AMOS 81>create function goals1(match m)->integer as
```

```
select goals(t1, m) + goals(t2, m)
from team t1, team t2;
AMOS 82>goals1(:m1);
```

In the match :m1 Italy made three goals and Brazil two. The answer to the query thus ought to be '5'. As the function is currently define t1 and t2 can also be bound to the same team and therefore the erroneous answers '4' and '6' are returned. Thus we have to specify that t1 and t2 are different teams:

The function is no longer returning illegal answers but instead it returns the correct answer twice. In one case t1 is bound to Italy and t2 to Brazil. In the other case t2 is bound to Italy and t1 to Brazil. To only use one of these cases we can use '<' rather than '!='. The comparison function '<' can be used for comparing any kind of object and it thus can order any objects.

We can now specify the original query:

```
AMOS 87>select m from match m where goals(m)=maxagg((select goals(m1) from match m1))
```

The answer is not very informative. We therefore define the function matchinfo that, given a certain match, returns a tuple of each participating team, how many goals the team made, and what year the match was played. Then we ask the query once more:

5. 'In what matches were more than ten goals made?'

```
AMOS 90>select matchinfo(m) from match m where goals(m)>=10;
```

10. Repetition exercises

If you typed something illegal or if there is some error you can recreate the correct database with the Windows command:

amos2 wc.dmp

In Table 1 there is an overview of the types and functions that are defined.

- 1. How many spectators did the match :m1 have?
- 2. What was the name of the referee in the match :m1?
- 3. Make a derived function ref_name(match)->charstring that, given a certain match, returns the name of the referee that refereed the match. Then use the derived function to answer query 2.
- 4. Which years were World Cup tournaments played?
- 5. A team is trained by a trainer. A trainer has a name and a salary. Create the types and functions needed to model this. Create a trainer object too. Store name and salary for the created trainer. Specify thereafter an AMOSQL command showing the name and salary for all trainers.
- 6. Who were the referees in the games where Tomas Brolin played?
- 7. Create a derived function referees(charstring)->charstring that, given the name of a certain player, returns the names of the referees that refereed games where that player participated. Then use the derived function to answer query 6.
- 8. How many times did Mexico arrange the World Cup?
- 9. What match had the most spectators? Assign the environment variable :m to the answer. Then use :m to retrieve the number of spectators and what teams that played in that match (use the function matchinfo).
- 10. What games did Sweden win?
- 11. How many goals did Kurt Hamrin make?

Types	Functions
tournament	year(tournament) -> integer
	host(tournament) -> country
	participating_teams(tournament) -> team
country	name(country) -> charstring
person	name(person) -> charstring
referee (subtype of person)	
player (subtype of person)	goals(player, match) -> integer
	own_goals(player, match) -> integer
team	represents(team) -> country
	players(team) -> player
	goals(team, match) -> integer
match	refereed_by(match) -> referee
	played_in(match) -> tournament
	spectators(match) -> integer
	player_participations(match) ->
	<player, goals,="" integer="" own_goals=""></player,>
	participating_players(match) -> player

Types	Functions
	team_participations(match) -> <team, goals="" integer=""></team,>
	participating_teams(match m) -> team
	goals(match) -> integer
	matchinfo(match) ->
	<charstring c1,="" c2,="" charstring="" g1,="" g2,integer="" integer="" y=""></charstring>

Table 1: Type and function definitions

11. Referenses

- 1 D.H.Fishman. et al., 'Overview of the Iris DBMS', Object-Oriented Concepts, Databases and Applications, W. Kim, F.H. Lochovsky (eds.), ACM Press, Addison-Wesley, 1989.
- 2 S.Flodin, T.Risch, M.Sköld: 'AMOS2 User's Guide', http://www.ida.liu.se/~torri/amos2/amosdoc.html.