

Part 2: Introduction to the Relational Model and SQL

References:

- Elmasri/Navathe: Fundamentals of Database Systems, 3rd Edition, 1999.
Section 7.1, "Relational Model Concepts"
Section 8.2, "Basic Queries in SQL"
- Kemper/Eickler: Datenbanksysteme (in German), 3rd Edition, 1999.
Section 3.1, "Definition des relationalen Modells" ("Definition of the Relational Model")
Section 4.6, "Einfache SQL-Anfragen" ("Simple SQL Queries")

Objectives

- Basic notions of the relational model.
- Simple SQL queries.
- Introduction to using Oracle SQL*Plus.
- Readiness for practical homework assignments.

Overview

1. The Relational Model, Example Database

2. Using SQL*Plus: First Demonstration

3. Simple SQL Queries

4. Historical Remarks

The Relational Model (1)

- The relational model structures data in table form, i.e. a relational DB is a set of named tables.
- A relational database schema defines:
 - Names of tables in the database.
 - The columns of each table, i.e. the names of the columns and the data types of the column entries.

The columns have a sequence (first column, second column, etc.). Each column can store only data of a particular type, e.g. strings, numbers of a certain length and precision, date values, etc.

- Integrity constraints (see a later chapter).

Integrity constraints are conditions that the data must satisfy.

The Relational Model (2)

- For instance, Oracle comes with an example database that consists of three tables:
 - EMP: Information about employees.
 - DEPT: Information about departments.
 - SALGRADE: Information about salary ranges for different levels/grades.

This table is not used in the following example queries.

- Depending on the version of the example database, there might be further tables.

The Relational Model (3)

- The Table DEPT has the following columns:
 - DEPTNO: Department Number,
 - DNAME: Department Name,
 - LOC: Location.

DEPT		
DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

The Relational Model (4)

The columns have the following data types:

- DEPTNO has the data type NUMERIC(2),
i.e. can be a two-digit integer from -99 to +99.

Obviously, one does not want negative department numbers.
They can be excluded by means of constraints.

- DNAME has the type VARCHAR(14), i.e. is a character string of variable length that consists of at most 14 characters.
- LOC has the type VARCHAR(13).

The data dictionary (system catalog) actually lists types NUMBER(2), VARCHAR2(14), and VARCHAR2(13). These are Oracle-specific names for the types listed above.

The Relational Model (5)

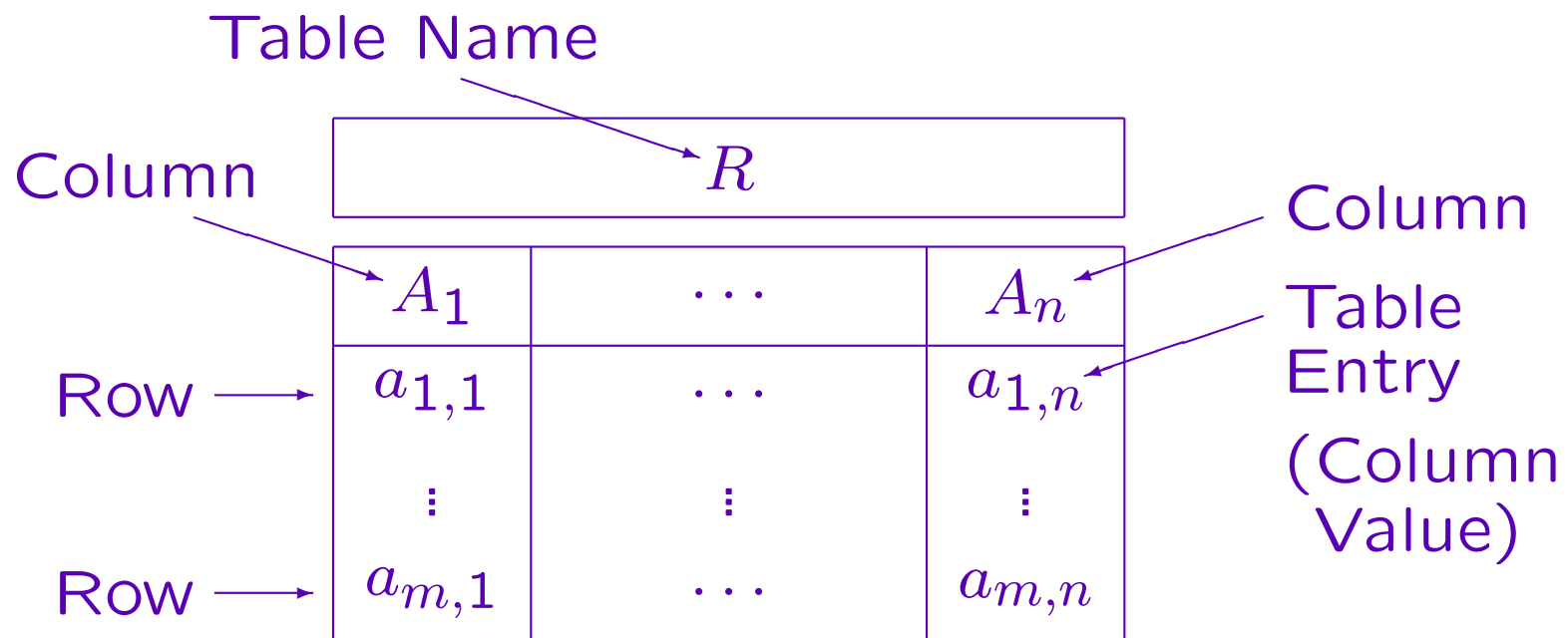
- A relational database state (instance of a given schema) defines for each table a set of rows.
- The example table “DEPT” has currently four rows.
- The relational model does not define any particular order of the rows (e.g. which is first, second, etc.).

But rows can be sorted for output.

- Each row specifies values for the columns of the table.

E.g. one row above has the value 10 for the column DEPTNO, the value 'ACCOUNTING' for DNAME, and 'NEW YORK' for LOC.

Summary (1)



Summary (2)

- A more theoretically oriented person would use the following synonyms:

- Relation instead of table.

A table is formally a subset of the cartesian product of the domains of the columns, and that is called a relation in mathematics. Cartesian coordinates are (X, Y) -pairs of real numbers, i.e. elements of $\mathbb{R} \times \mathbb{R}$. The relation $<$ can also be understood as a subset of $\mathbb{R} \times \mathbb{R}$ (e.g. $(1, 2)$ is contained in the relation, and $(2, 1)$ is not contained in the relation). However, database relations are always finite and they may have more than two columns.

- Tuple instead of row.
- Attribute instead of column.

Summary (3)

- Old-style practical people might say
 - record instead of row,
 - field instead of column,
 - field value instead of table entry,
 - file instead of table.

That should be avoided since it is confusing: Modern DBMS might store many tables in the same operating system file, and they may also split the same table over different files.

More RM Concepts (1)

EMP (data about employees) has the following columns:

- EMPNO: A unique number for every employee.
- ENAME: Employee name.
- JOB: Employee position (e.g. CLERK).
- MGR: Direct supervisor of this employee.
- HIREDATE: Employee hire date.
- SAL: Employee salary.
- COMM: Commission (only for salespeople).
- DEPTNO: Department where this employee works.

More RM Concepts (2)

EMP							
EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	7902	17-DEC-80	800		20
7499	ALLEN	SALESMAN	7698	20-FEB-81	1600	300	30
7521	WARD	SALESMAN	7698	22-FEB-81	1250	500	30
7566	JONES	MANAGER	7839	02-APR-81	2975		20
7654	MARTIN	SALESMAN	7698	28-SEP-81	1250	1400	30
7698	BLAKE	MANAGER	7839	01-MAY-81	2850		30
7782	CLARK	MANAGER	7839	09-JUN-81	2450		10
7788	SCOTT	ANALYST	7566	09-DEC-82	3000		20
7839	KING	PRESIDENT		17-NOV-81	5000		10
7844	TURNER	SALESMAN	7698	08-SEP-81	1500	0	30
7876	ADAMS	CLERK	7788	12-JAN-83	1100		20
7900	JAMES	CLERK	7698	03-DEC-81	950		30
7902	FORD	ANALYST	7566	03-DEC-81	3000		20
7934	MILLER	CLERK	7782	23-JAN-82	1300		10

More RM Concepts (3)

References (Foreign Keys):

- The relational model has no physical pointers.
- But it allows comparison of column values in different rows.
- E.g., the column DEPTNO in the EMP table refers to the table DEPT (DEPTNO uniquely identifies rows in DEPT).
- MGR even refers to another row in the same table.

It contains the employee number of the direct supervisor.

- So this is some kind of “logical pointer”.

More RM Concepts (4)

Null Values:

- The relational model allows table entries to remain empty (contain a “null value”).
- In the example table: Only salespeople have a commission, the company president has no supervisor.
- In the schema declaration, one can specify for each column whether it accepts null values or not.
- The null value is treated specially in comparisons.

See a later chapter.

Overview

1. The Relational Model, Example Database

2. Using SQL*Plus: First Demonstration

3. Simple SQL Queries

4. Historical Remarks

Oracle and SQL*Plus (1)

- SQL is the standard database language for relational database management systems (RDBMSs).

Oracle supports SQL, as do all other modern RDBMS (e.g. DB2).
The systems differ in many small details, see below.

- SQL*Plus is Oracle's basic interface to the DBMS.

The DBMS itself (DB server) runs as a set of background processes,
maybe on another machine in the network.

- E.g. C programs with embedded SQL also communicate with the Oracle server in SQL, but do not use SQL*Plus.

Oracle and SQL*Plus (2)

- The main task of SQL*Plus is
 - to allow the user to enter SQL statements,
It provides some editing functions, e.g. to correct errors.
 - send the query to the server, fetch the result,
 - and print the output table.
It provides some control over output formatting.
Simpler reports can be developed entirely in SQL*Plus.
- SQL*Plus also can process batch files (with several commands). This is also supported by a simple variable replacement mechanism.

Basic Use of SQL*Plus (1)

- Under UNIX, enter the command “sqlplus” .
Environment variables must be set, see Appendix B.
- Under Windows, select “Start → Programs → Oracle-OraHome81 → Application Development → SQL Plus” .
- SQL*Plus then asks for username and password.
The “host string” field in the login box can usually be left empty. It allows the user to select a remote server.
- Many Oracle installations have a guest user “scott” with password “tiger” .

Basic Use of SQL*Plus (2)

- The example tables should already be installed under the guest account.
- To install the tables under another account, execute
 - “demobld” under UNIX from the UNIX prompt or
 - “@C:\Oracle\Ora81\DBS\demobld.sql” under Windows from the SQL*Plus prompt.
- In Oracle, every user has his/her own DB schema.

I.e. after the demobld command you have your own copy of the example data and are free to modify the tables. You should not modify the tables under the guest account.

Basic Use of SQL*Plus (3)

- The SQL*Plus prompt is “SQL>” .
Or “2”, “3”, etc. for continuation lines.
- In SQL*Plus, SQL statements must end with “;” .
Alternative: “/” as the only contents of the next line.
- SQL statements can extend over several lines, and SQL*Plus needs to know when the user is finished.
The “;” is not part of SQL. In DB2/SQL server, the user must instead click on “execute” when the query is complete.
- SQL is not case sensitive (except in strings).

Basic Use of SQL*Plus (4)

- Leave SQL*Plus with “exit” or “quit”.

In contrast to SQL statements, such control commands need no semi-colon at the end. If they extend over multiple lines, each line (except the last) must end with a hyphen “-”.

- You can list the contents of a table with, e.g.:

```
SELECT * FROM EMP;
```

- You can list all tables in your schema with:

```
SELECT * FROM CAT;
```

- You can list the columns of a table with, e.g.:

```
DESCRIBE EMP
```

Basic Use of SQL*Plus (5)

```
UNIX> sqlplus
```

```
. . . (Version and copyright information for SQL*Plus)
```

```
Enter user-name: scott
```

```
Enter password: tiger (not visible)
```

```
. . . (Version information for the database server)
```

```
SQL> select * from dept;
```

DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

Basic Use of SQL*Plus (6)

- The last SQL command is still stored in a buffer. You can change and re-execute it.

E.g. in case of a syntax error.

- `c/old/new` replaces the first occurrence of “old” by “new” .

In the current line (line in which the error was detected).

- `l (list)` shows the contents of the buffer.

`l3` shows only line 3 and makes it the current one.

- `r (run)` executes the contents of the buffer.

Basic Use of SQL*Plus (7)

- `edit` writes the contents of the buffer to a file and calls an editor.

The editor can be selected by the user, e.g. `define _editor = vi`. This command is only valid for the current session, but one can write it into `login.sql` (in the current directory).

- The user can also write the SQL query into a file, e.g. “`x.sql`”, and execute the file with `@x`.

E.g. you can keep an editor open in one window, write to the file, and execute the file in the other window, where you have SQL*Plus open.

Basic Use of SQL*Plus (8)

```
SQL> select *  
      2  frm cat;  
frm cat  
*  
ERROR at line 2:  
ORA-00923: FROM keyword not found where expected  
SQL> c/frm/from/  
      2* from cat  
  
SQL> r  
      1  select *  
      2* from cat  
  
(shows the catalog, i.e. lists tables/views owned by current user)
```

Basic Use of SQL*Plus (9)

- If the first keyword is misspelled, the command is not stored in the buffer, and must be retyped.

Only SQL statements (recognized by first keyword) are stored in the buffer, not SQL*Plus control commands.

- Oracle reports only one error in the SQL command.

There might be more errors. The position which Oracle marks is not necessarily the first position where a parser could find an error (e.g. quotes are missing around strings):

```
SQL> insert into dept values(40, WWW, DALLAS);
insert into dept values(40, WWW, DALLAS)
                                *
ERROR at line 1:
ORA-00984: column not allowed here
```

Overview

1. The Relational Model, Example Database
2. Using SQL*Plus: First Demonstration
3. Simple SQL Queries
4. Historical Remarks

Simple SQL Queries (1)

- Simple SQL queries have the structure

```
SELECT ... FROM ... WHERE ...
```

- After FROM list the table from which to extract data.

More than one table can be listed, see below.

- After WHERE specify conditions for the rows to be selected.

The WHERE-clause can be missing, then all rows are selected.

- After SELECT define which columns to print.

“*” prints all columns.

Simple SQL Queries (2)

- Show the entire department table:

```
SELECT * FROM DEPT
```

DEPTNO	DNAME	LOC
10	ACCOUNTING	NEW YORK
20	RESEARCH	DALLAS
30	SALES	CHICAGO
40	OPERATIONS	BOSTON

- Equivalent Solution:

```
SELECT DEPTNO, DNAME, LOC FROM DEPT
```

Simple SQL Queries (3)

- SQL is not case-sensitive, except inside strings.

All characters in an SQL command are turned into uppercase before they are processed (except inside quotes).

To get a table or column name containing lower case letters, it must be enclosed in " (should be avoided).

- SQL is format-free (like Pascal, C, Java, etc):

Line breaks, spaces, and tabulator characters can be inserted between the words (tokens) of an SQL command.

- Thus, the following query is the same as the previous one:

```
select deptno,  dname ,   loc
from    dept
```

Using Conditions (1)

- To get all data about the department in “DALLAS”:

```
SELECT * FROM DEPT WHERE LOC = 'DALLAS'
```

DEPTNO	DNAME	LOC
20	RESEARCH	DALLAS

- String constants are marked by single quotes.

Double quotes are only used for delimited identifiers (column names containing lowercase characters etc.).

- Inside string constants, SQL is case-sensitive.

So the following will select 0 rows (empty result):

```
SELECT * FROM DEPT WHERE LOC='Dallas'
```

Using Conditions (2)

- Print the name, job, and salary of all employees who earn at least \$2500:

```
SELECT ENAME, JOB, SAL
FROM EMP
WHERE SAL >= 2500
```

ENAME	JOB	SAL
JONES	MANAGER	2975
BLAKE	MANAGER	2850
SCOTT	ANALYST	3000
KING	PRESIDENT	5000
FORD	ANALYST	3000

Using Conditions (3)

- It is not necessary to output columns which are used in the conditions. `WHERE` is evaluated before `SELECT`.
- E.g. print the employee number and name of all managers:

```
SELECT EMPNO, ENAME
FROM   EMP
WHERE  JOB = 'MANAGER'
```

EMPNO	ENAME
7566	JONES
7698	BLAKE
7782	CLARK

Pattern Matching (1)

- SQL also has an operator for “pattern matching” of strings (allowing the use of “wildcards”).

- E.g. print number and name of all managers:

```
SELECT EMPNO, ENAME
FROM EMP
WHERE JOB LIKE 'MANA%'
```

- “%” matches any sequence of arbitrary characters,
“_” matches any single character.

So “%” corresponds to the “*” in the shell, and “_” to “?”.

Pattern Matching (2)

- LIKE must be used for pattern matching.
The equals sign only tests for literal equality.
Even if the comparison string contains “%” or “_”.
- If one wants to do pattern matching, but also needs the meta-characters “%” and “_” literally, one must declare an “escape character” and prefix that character to the occurrences without special meaning:

```
SELECT COLUMN_NAME
FROM   USER_TAB_COLUMNS
WHERE  COLUMN_NAME LIKE '%\_%' ESCAPE '\'
```

Arithmetic Expressions

- Standard arithmetic expressions are allowed in SQL.
- E.g. print all employees who earn less than \$15000 per year:

```
SELECT ENAME
FROM EMP
WHERE SAL < 15000 / 12
```

ENAME
SMITH
ADAMS
JAMES

- The condition “ $SAL * 12 < 15000$ ” is equivalent.

Renaming Output Columns

- E.g. print the yearly salary of all managers:

```
SELECT ENAME, SAL * 12
FROM EMP
WHERE JOB = 'MANAGER'
```

ENAME	SAL * 12
JONES	35700
BLAKE	34200
CLARK	29400

- To print the column heading "YEARLY SALARY":

```
SELECT ENAME, SAL * 12 "YEARLY SALARY" ...
or SELECT ENAME, SAL * 12 AS "YEARLY SALARY" ...
```

Logical Connectives (1)

- AND, OR, NOT and parentheses “(”, “)” can be used to construct more complicated conditions.
- E.g. print name and salary of all managers and the president:

```
SELECT ENAME, SAL
FROM EMP
WHERE JOB = 'MANAGER' OR JOB = 'PRESIDENT'
```

- The query would not work with AND instead of OR.

The result would be empty (“0 rows selected”), since the JOB cannot be “MANAGER” and “PRESIDENT” at the same time. The WHERE-condition is conceptionally evaluated for every row of EMP. AND is true if both parts are true, OR is already true if one part is true.

Logical Connectives (2)

C1	C2	C1 AND C2	C1 OR C2	NOT C1
False	False	False	False	True
False	True	False	True	True
True	False	False	True	False
True	True	True	True	False

- E.g. C1 is JOB = 'MANAGER', C2 is JOB = 'PRESIDENT'.
- For every given row, only one or the other can be true.

Or both can be false. I.e. only the first three cases occur.

- Thus, "C1 AND C2" will be false.

Logical Connectives (3)

- Without parentheses, AND binds more strongly than OR (and NOT binds even more strongly than AND).

- E.g. consider this query:

```
SELECT ENAME, SAL      Wrong!
FROM   EMP
WHERE  JOB = 'MANAGER' OR JOB = 'PRESIDENT'
AND    SAL >= 3000
```

- The system will understand the condition as:

```
WHERE JOB = 'MANAGER'
OR     (JOB = 'PRESIDENT' AND SAL >= 3000)
```

Removing Duplicates (1)

- Queries can produce duplicate rows.
- E.g. list all jobs:

```
SELECT JOB FROM EMP
```

- Since this query is processed by a loop over the employee table, in which the job of every employee is printed, it will output the same job multiple times.
- Duplicate elimination can be requested by adding the keyword `DISTINCT` after `SELECT`:

```
SELECT DISTINCT JOB FROM EMP
```

Removing Duplicates (2)

- `DISTINCT` works on output rows, not single columns. Specify it only once, even for multiple columns:

```
SELECT DISTINCT JOB, MGR
FROM EMP
```

- There is no way to print each job only once and for every job the set of managers — that would be a nested table (supported only in advanced NF² or object-relational DBMS).

In standard relational DBMS, each table entry is atomic. But output formatting can give something similar. E.g. in SQL*Plus, it is possible to suppress a column value if it is the same as in the preceding row.

Sorting Output Rows (1)

- The sequence, in which the resulting rows are printed, is not predictable unless one requests sorting.
- E.g. print employee names and their salary, ordered alphabetically by employee name:

```
SELECT  ENAME, SAL
FROM    EMP
ORDER BY ENAME
```

- The “ORDER BY” clause is purely cosmetic:
It does not change the query result in any way,
it only prints the result in a more readable fashion.

Sorting Output Rows (2)

- One can also specify multiple sorting criteria.

Only useful if the main column for sorting contains duplicate values.

- E.g. print the employees who earn at least 2000 dollars, ordered by department number. For equal department numbers, employees should be ordered by descending salaries:

```
SELECT  DEPTNO, ENAME, SAL
FROM    EMP
WHERE   SAL >= 2000
ORDER  BY DEPTNO, SAL DESC
```

Outlook: Joining Tables

- Data from different tables can be combined.
E.g. print employees in the “RESEARCH” department:

```
SELECT ENAME  
FROM EMP, DEPT  
WHERE EMP.DEPTNO = DEPT.DEPTNO  
AND DNAME = 'RESEARCH'
```

- Conceptually, the WHERE-condition will be evaluated for every combination of one row from each table.
- Since “DEPTNO” appears in both tables, and both tables are referenced in this query, the column name must be made unique by prefixing the table table.

Exercises (1)

Please formulate the following queries in SQL:

- Who has employee number 7839 (King) as direct supervisor?

- Who has a salary between \$1000 and \$2000?

Print name and salary and order the result by names.

“Between” is meant as including the two boundaries.

- Which employee names consist of exactly four characters?

Exercises (2)

- Print name, salary, and department of all employees who work in department 10 or 30 and earn less than \$1500.

Make sure that both conditions are really satisfied.

- Which jobs occur in which departments? I.e. print every combination of department number and job which occurs in the EMP table, and print each such combination only once.

Overview

1. The Relational Model, Example Database
2. Using SQL*Plus: First Demonstration
3. Simple SQL Queries
4. Historical Remarks

Relational Model: History

- The RM was proposed by E. F. Codd (1970).
- It was the first data model that was theoretically defined prior to implementation.
- First Implementations (1976):
 - System R (IBM)
 - Ingres (Stonebraker, UC Berkeley).
- First commercial systems: Oracle (1979), Ingres (1980?) and IBM SQL/DS (1981).
- Today “state of the art” in industry.

Reasons for Success (1)

- Much simpler than earlier data models.
Only one concept: Finite relation (set of tuples).
- Easily understandable, even for non specialists:
Relations correspond to tables.
- Abstraction of known “files of records” .
- The relational model has set-oriented operations.
In earlier models, one had to navigate from one record to the next.

Reasons for Success (2)

- Declarative query language:

No need to think about efficient evaluation.

One only writes conditions for the required data. The DBMS contains a “query optimizer” that finds an efficient query evaluation plan (i.e. generates a good imperative program for evaluating the query). In earlier models, programmers had to think about the use of indexes (access paths) and many other details.

- The relational model has a solid theoretical foundation. It is tightly connected to first-order logic.

SQL

- Today, SQL is the only database language for relational DBMSs (industry standard).
- SQL is used for:
 - Interactive “ad-hoc” commands and
 - application program development (embedded into other languages like C, Java, HTML, PL/SQL).
- SQL is based on a variant of first order logic called tuple calculus.

But includes elements from relational algebra, too (e.g. UNION).
It tries to be relatively near to natural language.

History

- SEQUEL, an earlier version of SQL, was designed by Chamberlin, Boyce et al. at IBM Research, San Jose (1974).

SEQUEL stands for “Structured English Query Language”. Some people pronounce SQL this way. Others use “ess-cue-ell”. The name was changed for legal reasons (SEQUEL was a registered trademark). Codd was also in San Jose when he invented the relational model in 1970.

- SQL was the language of System/R (1976/77).

System/R was a very influential research prototype.

- First commercial systems supporting SQL were Oracle (1979) and IBM SQL/DS (1981).

Standards

- First Standard 1986/87 (ANSI/ISO).

This was very late as there were already several SQL systems on the market. The standard was the “smallest common denominator”. It contains only the common features of the existing implementations.

- Extension for foreign keys etc. in 1989 (SQL-89).

This version is called also SQL-1. All commercial implementations today support this standard, but each have significant extensions.

- Major Extension: SQL-2 or SQL-92 (1992).

Upward compatible to SQL-1. The standard defines three levels: “entry”, “intermediate”, “full”. Oracle 8.0 and SQL Server 7.0 have only entry level conformance, but many extensions. SQL-92 is still the yardstick for RDBMSs.

Future

- Current Standard: SQL-99.

SQL-99 is a preliminary version of the SQL-3 standard. Until 12/2000, the volumes 1–5 and 10 of the SQL-99 standard appeared. They have together 2355 pages. The SQL-2 standard, which is not yet completely implemented, had only 587 pages.

- Some Features of SQL-3:

- User-defined data types, type constructors.

E.g. “LIST”, “SET”, “ROW” for structured attribute values.

- OO-Features (e.g. inheritance/subtables).

- Recursive queries.

- Triggers, Persistent Stored Modules.