Chapter 5. Advanced SQL

5.1 Accessing SQL from a Programming Language

Remark. Some queries or nondeclarative actions cannot be expressed in SQL.

ex) Very complicated queries. User interactions. Printing, Scanning. ...

i) Dynamic SQL

A general-purpose program communicating with a database server with a collection of functions. ex) ODBC. JDBC.

Remark. Prepared statements allow more efficient execution, and robust to malicious attacks like SQL injection.

ii) Embedded SQL.

SQL are embedded at the source for general-purpose program. Preprocessor submits embedded SQL to a database server for precompilation and optimization, and embeds the result with appropriate target language. to be compiled locally.

5.2 Functions and Procedures.

Remark. These allow business logic to be stored within a database.

- Useful for special data types, like images.
- It allow multiple applications to access it, so it allows single point of change.

Example. CREATE OR REPLACE FUNCTION func-name(param type, ...)
RETURNS type or table
↳ "table functions"
can be thought as "parameterized views".

Def. Persistent Storage Module (PSM)
is a language construct for procedures and functions.

Compound Statement → BEGIN ... END
Loops → FOR ... AS ... DO ... END FOR.
Variables → SET ...
Branches → IF ... THEN ... ELSEIF ... ELSE ... END IF
Parameters → IN name type, OUT name type

	Language.	Safety
Persistent Storage Module	SQL Standard. But most DBMSs implement their own nonstandard versions.	Syntax and safety checks are done within DBMS.
External Language Routine.	C. Java. C#. VB. Python. Perl. ... Depends on implementation.	"Safe languages" executes code in a sandbox within the query execution process. Not possible for some languages like C.

5.3 Triggers.

Def. A **trigger** is a statement that the system executes automatically as a side-effect of a modification to a database.

It consists of 3 components: event, condition, and actions.

Remark. Triggers can be activated before or after the event.

Remark. Need for triggers.

- Complex integrity constraints
- Monitoring or Tasks.
- Delta relations.

Remark. Hazards of triggers

- Runtime failure of the triggering event.
- Chain of triggers.

5.4 Recursive Queries

Example How recursive query works

i) Run base query

ii) Run recursive query

iii) Repeat until no new tuples are added.

```
with recursive c_prereq(course_id, prereq_id) as (
        select course_id, prereq_id
        from prereq                          } Base query
    union
        select prereq.prereq_id, c_prereq.course_id
        from prereq, c_prereq                } Recursive query.
        where prereq.course_id = c_prereq.prereq_id
    )
select *
from c_prereq;
```

Remark. It is an "iterative" process.

Def. Query or View is **monotonic**

if its result R_1, R_2 on view relation V_1, V_2 satisfies $R_1 \subset R_2$ if $V_1 \subset V_2$.

Remark. Recursive query has to be monotonic.

Therefore, aggregation, NOT EXISTS, EXCEPT (set difference) on right-hand side is not permitted, as it may result in $R_1 \supset R_2$.

Remark. SQL also allows creation of recursive views.

5.5 Advanced Aggregation Feature.

i) Ranking

- RANK: 1, 2, 2, 4, 4, 4, 7
- DENSE_RANK: 1, 2, 2, 3, 3, 3, 7
- CENT_RANK: 0/7, 1/7, 1/7, 3/7, 3/7, 3/7, 6/7 (RANK−1 / TOTAL−1)
- CUME_DIST: 1/8, 2/8, 2/8, 3/8, 3/8, 3/8, 1/8. (COUNT / TOTAL)
- ROW_NUMBER: 1, 2, 3, 4, 5, 6, 7 (2, 3 / 4, 5, 6 is arbitrary)
- NTILE(n): n buckets.
- NULLS FIRST / LAST.

Remark These can be implement using nested queries, but system's implementation of various ranking may speed up the query.

ii) Windowing.

Example. SELECT AVG(money) OVER (ORDER BY year ?)

i) ROWS n PRECEDING : for 2000, AVG(money) for year 1998, 1999, 2000

ii) ROWS n FOLLOWING : for 2000, AVG(money) for year 2000, 2001, 2002

iii) ROWS UNBOUNDED PRECEDING / FOLLOWING : All the years.

iv) RANGE BETWEEN year−4 and year : for 2000, 1996, 1997, 1998, 1999, 2000.

5.6 OLAP

Def. An online analytical processing (OLAP) is an interactive system that permits an analyst to view different summaries of multidimensional data, within a few seconds of latency per response.

Def. Multidimensional data are data that can be modelled as dimension attributes and measure attributes.

Def. Cross-tabulation, or pivot-table, is a table derived from a relation R where values from one attribute from R forms a row, and another attribute from R forms a column.

Def. The operation of changing the dimensions used in a cross-tab is called pivoting.

Remark. The different level of detail for an attribute can be organized into a hierarchy.

Def. The operation of making a pivot-table with some attributes fixed to some value is called slicing or dicing.

Def. The operation of moving from finer-granularity data to a coarser one (by aggregation) is called a rollup and the opposite drill down.

Example. A cross-tab of sales with hierarchy on item name.

Slicing (or dicing) by clothes_size.

clothes_size: all

hierarchical attributes

category	item_name	color: dark	color: pastel	color: white	total	
womenswear	skirt	8	8	10	53	
	dress	20	20	5	35	
	subtotal	28	28	15		88
menswear	pants	14	14	28	49	
	shirt	20	20	5	27	
	subtotal	34	34	33		76
total		62	62	48		164

Row

roll-up

drill-down.

Column