SQL FOR DB2

Chapter 1
Introduction to DB2

Chapter Objectives

- Explain what a DBMS is
- Explain the hierarchy of database, tables, rows, and columns
- Describe the different SQL data types
- Explain the EBCDIC representation for storing data

Database Management Systems (DBMSs)

- A Database Management System (DBMS) is a set of software programs that control the organization, storage, and retrieval of a database
- Structured Query Language (SQL), is a language used by DBMS
- Common DBMS that include SQL are DB2 and MySQL
- DBMS:
  - Manages data, organizes it, modifies and retrieves data
  - Lets one or more users create and access data
  - Ensures the integrity of the data and provides security

DB2 Universal Database

- DB2 Universal Database or DB2 UDB, refers to the entire family of IBM databases
- This book used the DB2 UDB for i, which is referred to simply as DB2
- DB2 is an integrated part of the IBM i system's OS; it is not purchased as a separate software product

SQL vs Traditional Terminology

SQL Term	Non-SQL Terms
Schema. Consists of a library, a journal, a journal receiver, an SQL catalog, and optionally a data dictionary. A schema groups related objects and allows you to find the objects by name.	**Library**. Groups related objects and allows you to find the objects by name.
Table. A set of columns and rows.	**Physical file**. A set of records.
Row. The horizontal part of a table containing a serial set of columns.	**Record**. A set of fields.
Column. The vertical part of a table of one data type.	**Field**. One or more characters of related information of one data type.
View. A subset of columns and rows of one or more tables.	**Logical file**. A subset of fields and records of one or more physical files.

Hierarchy of Data

- Table(File)
  - Collection of data about a given kind of entity or object
- Row(Record)
  - Collection of data about one specific instance of the table
- Column(Field)
  - One piece of data on the row

Internal Binary Representation of data

- Internally, data are stored differently depending on the computer system used, PC's use ASCII and the IBM System i uses EBCDIC
- It is difficult to understand data in its binary form, so it is converted to a 2-character hexadecimal value
- Hex characters can range from 0 (all 4 bits off=0000) to F (all 4 bits on=1111)
- In EBCDIC each storage position, or byte, consists of eight bits. The high-order four bits are used to specify zone portion, and the low-order four bits are used to specify the digit portion
- The zone portion is used to indicate whether the value stored is a letter, positive number, negative number or special character
- The digit portion is used to indicate whether the value stored is a number

EBDIC Codes

EBCDIC											
	Binary				Binary				Binary		
Char	Zone	Digit	Hex	Char	Zone	Digit	Hex	Char	Zone	Digit	Hex
blank	0100	0000	40	u	1010	0100	A4	P	1101	0111	D7
a	1000	0001	81	v	1010	0101	A5	Q	1101	1000	D8
b	1000	0010	82	w	1010	0110	A6	R	1101	1001	D9
c	1000	0011	83	x	1010	0111	A7	S	1110	0010	E2
d	1000	0100	84	y	1010	1000	A8	T	1110	0011	E3
e	1000	0101	85	z	1010	1001	A9	U	1110	0100	E4
f	1000	0110	86	A	1100	0001	C1	V	1110	0101	E5
g	1000	0111	87	B	1100	0010	C2	W	1110	0110	E6
h	1000	1000	88	C	1100	0011	C3	X	1110	0111	E7
i	1000	1001	89	D	1100	0100	C4	Y	1110	1000	E8
j	1001	0001	91	E	1100	0101	C5	Z	1110	1001	E9
k	1001	0010	92	F	1100	0110	C6	0	1111	0000	F0
l	1001	0011	93	G	1100	0111	C7	1	1111	0001	F1
m	1001	0100	94	H	1100	1000	C8	2	1111	0010	F2
n	1001	0101	95	I	1100	1001	C9	3	1111	0011	F3
o	1001	0110	96	J	1101	0001	D1	4	1111	0100	F4
p	1001	0111	97	K	1101	0010	D2	5	1111	0101	F5
q	1001	1000	98	L	1101	0011	D3	6	1111	0110	F6
r	1001	1001	99	M	1101	0100	D4	7	1111	0111	F7
s	1010	0010	A2	N	1101	0101	D5	8	1111	1000	F8
t	1010	0011	A3	O	1101	0110	D6	9	1111	1001	F9

Data Types

- Every variable must have a data type
- A data type identifies what type of data can be stored and how large
- Common data types:
  - Character
  - Decimal
  - Integer
  - Date

Character

- A character data type contains any combination of letters, digits and special characters, such as $, %, @, or & any printable character
- A variable defined as a character cannot be used in an arithmetic operations even thought the field may contain only numeric digits
- Each byte of a character data type is divided into two portions: the high-order, 4 bit zone portion and the low-order, 4-bit digit portion
- The following example shows how a 1-byte character variable containing the value “J” is represented J=1101 0001

Zone	Digit	
D	1	**Hex value**
1101	0001	**Binary value**

Decimal

- Eight bytes (15 digits) in length
- Two numeric digits are stored in each byte, except for the rightmost byte, which also contains the sign
- The low-order four bits (digit portion) of the rightmost byte of a Decimal data type contain the sign (F=positive, D=Negative)
- All other zones are stripped, and two digits are packed into a single byte

DECIMAL data type containing the numeric value 68254

Zone	Digit	Zone	Digit	Zone	Digit	
6	8	2	5	4	F	**Hex**
0110	1000	0010	0101	0100	1111	**Binary**

Integer

- An integer data type is store in binary, in which the bits are directly interpreted as a number
- For example, the number 14 is 1110 when represented in binary
- The leftmost bit of an integer field is used to keep track of the sign, 0 for positive and 1 for negative
- 2257 stored in a SMALLINT data type

Byte 1		Byte 2	
0000	1000	1101	0001
0	2048	208	1

Date

- A Date data type is a variable that contains a valid date
- Date variables have a predetermined size of 10 bytes and predetermined format
- No data type or length is specified for date variables
- The default internal format is a 10-byte field in the format yyyy-mm-dd
- A Timestamp data type is a combination of the date, time as well as fraction of a second

Sample Data for one Record

Field Description	Data Type	Size	Bytes	Location	Sample Data
Employee Number (PK)	INTEGER	9,0	4	1-4	244575315
Store Number	SMALLINT	4,0	2	5-6	2257
First Name	CHARACTER	20	20	7-26	Janice
Middle Initial	CHARACTER	1	1	27-27	H
Last Name	CHARACTER	20	20	28-47	Matheson
Hire Date	DATE	10	10	48-57	2009/01/15
Department Number	SMALLINT	3,0	2	58-59	111
Hourly Rate	DECIMAL	5,2	3	60-62	015^70
Hours Worked	DECIMAL	3,1	2	63-64	22^5
Sales	DECIMAL	9,2	5	65-69	00056.75-

		1	2	3	4	5	6
		12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789					
Char:		******Janice			HMatheson		2009-01-15**
Hex.	Zone	09E50DD898884444444444444444CD8A88A9944444444444FFFF6FF6FF060502500065					
	Digit	E3C381115935000000000000008413852650000000000002O90010150F17F2F0057D					

Chapter Summary

- SQL:
  - Stands for Structured Query Language
  - Is a language used to define, retrieve, and manipulate databases
- DBMS:
  - Is a set of software programs that controls the organization, storage, and retrieval of data in a database
  - Used a standard method of cataloging, retrieving, and running queries on data
  - Manages incoming data, organizes it, and provides ways for the data to be modified or extracted by users or other programs