

Relational Database DB2 Overview for Java Developers

Relational Database

Chapter 1:

Objectives

You will learn:

- DB2 objects.
- The role and utilization of indexes, keys, constraints, and objects.
- Storage structures - table space and index space.
- Purpose of sequences.
- Routines - functions and stored procedures.

Relational Database

DB2 Overview for Java Developers

Objectives

You will learn:

- Aliases.
- MQT: Materialized Query Tables.
- Data types in DB2.
- Derived columns.
- Aggregation.

SYS-ED/Computer Education Techniques, Inc.

1: 3

Tables

- Tables are logical structures maintained by DB2.
 - Tables are made up of columns and rows.
 - There is no inherent order of the rows within a table.
 - At the intersection of every column and row there is a specific data item called a value.
 - A column is a set of values of the same type.
 - A row is a sequence of values such that the nth value is a value of the nth column of the table.
 - Every table must have one or more columns, but the number of rows can be zero.

SYS-ED/Computer Education Techniques, Inc.

1: 4

Relational Database

DB2 Overview for Java Developers

Schemas

- A schema is a collection of named objects.
- The objects that a schema can contain include distinct types, functions, stored procedures, sequences, and triggers.
- An object is assigned to a schema when it is created.

SYS-ED/Computer Education Techniques, Inc.

1: 5

Tables

- **base table**
 - A table created with the SQL statement CREATE TABLE and used to hold persistent user data.
- **auxiliary table**
 - A table created with the SQL statement CREATE AUXILIARY TABLE and used to hold the data for a column that is defined in a base table.
- **temporary table**
 - A table defined by either the SQL statement CREATE GLOBAL TEMPORARY TABLE or DECLARE GLOBAL TEMPORARY TABLE and used to hold data temporarily, such as the intermediate results of SQL transactions.
- **materialized query table**
 - A table created with the SQL statement CREATE TABLE and used to contain materialized data that is derived from one or more source tables specified by a fullselect.
 - A source table is a base table, view, table expression, or user-defined table function.
- **result table**
 - A set of rows that DB2 selects or generates from one or more tables or views.

SYS-ED/Computer Education Techniques, Inc.

1: 6

Relational Database

DB2 Overview for Java Developers

Indexes

- An index is an ordered set of pointers to rows of a base table or an auxiliary table.
 - Each index is based on the values of data in one or more columns.
 - An index is an object that is separate from the data in the table.
- When an index is defined using the CREATE INDEX statement, DB2 builds this structure and maintains it automatically.

SYS-ED/Computer Education Techniques, Inc.

1: 7

Indexes

- An index is an ordered set of pointers to rows of a base table.
 - Each index is based on the values of data in one or more table columns.
 - An index is an object that is separate from the data in the table.
 - When an index is created, the database manager builds this structure and maintains it automatically.

SYS-ED/Computer Education Techniques, Inc.

1: 8

Relational Database

DB2 Overview for Java Developers

Indexes

- Indexes are used by the database manager to:
 - Improve performance.
 - In most cases, access to data is faster than without an index.
 - Ensure uniqueness.
 - A table with a unique index cannot have rows with identical keys.

SYS-ED/Computer Education Techniques, Inc.

1: 9

Keys

- A key is one or more columns that are identified as such in the description of a table, an index, or a referential constraint.
- There are:
 - unique keys
 - primary keys
 - parent keys
 - foreign keys

SYS-ED/Computer Education Techniques, Inc.

1: 10

Relational Database

DB2 Overview for Java Developers

Constraints

- A unique constraint is a rule that prevents duplicate values in one or more columns in a table.
- A referential constraint is a rule about values in one or more columns in one or more tables.
- A check constraint sets restrictions on data added to a specific table.

SYS-ED/Computer Education Techniques, Inc.

1: 11

Triggers

- A trigger defines a set of actions that are executed when a delete, insert, or update operation occurs on a specified table.
- When such an SQL operation is executed, the trigger is said to be activated.

SYS-ED/Computer Education Techniques, Inc.

1: 12

Relational Database

DB2 Overview for Java Developers

Storage Structures

- **table space**
 - A table space can hold one or more base tables, or one auxiliary table.
 - All tables are kept in table spaces.
 - A table space can be defined using the CREATE TABLESPACE statement.
- **index space**
 - An index space contains a single index.
 - An index space is defined when the index is defined using the CREATE INDEX statement.

SYS-ED/Computer Education Techniques, Inc.

1: 13

Databases

- In DB2, a database is a set of table spaces and index spaces.
 - The index spaces contain indexes on the tables in the table spaces of the same database.
- Databases are defined using the CREATE DATABASE statement and are primarily used for administration.

SYS-ED/Computer Education Techniques, Inc.

1: 14

Relational Database

DB2 Overview for Java Developers

Catalog

- DB2 maintains a set of tables that contain information about the data under its control.
 - These tables are collectively known as the catalog.
 - The catalog tables contain information about DB2 objects such as tables, views, and indexes.

SYS-ED/Computer Education Techniques, Inc.

1: 15

Views

- A view provides an alternative way of looking at the data in one or more tables.
- A view is a named specification of a result table.
 - The specification is an SQL SELECT statement that is effectively executed whenever the view is referenced in an SQL statement.
 - At any time, the view consists of the rows that would result if the fullselect were executed.

SYS-ED/Computer Education Techniques, Inc.

1: 16

Relational Database

DB2 Overview for Java Developers

Sequences

- A sequence provides a way to have the database manager automatically generate unique integer primary keys and to coordinate keys across multiple rows and tables.
 - A sequence is a stored object that generates a sequence of numbers in a monotonically ascending or descending order.
 - A sequence eliminates the serialization which results from the programmatic generation of unique numbers by locking the most recently used value and then incrementing it.

SYS-ED/Computer Education Techniques, Inc.

1: 17

Routines

- A routine is an executable SQL object.
- The two types of routines are functions and stored procedures.
 - **Functions**
 - A function is a routine that can be invoked from within other SQL statements and returns a value or a table.
 - **Stored procedures**
 - A stored procedure, also referred to as a procedure, is a routine that can be called to perform operations which include both host language and SQL statements.

SYS-ED/Computer Education Techniques, Inc.

1: 18

Relational Database

DB2 Overview for Java Developers

Aliases

- An alias is an alternate name for a table or view. It can be used to reference a table or view in those cases where an existing table or view can be referenced.
 - As with tables and views, an alias may be created, dropped, and have comments associated with it.
 - Unlike tables, aliases may refer to each other in a process called chaining.
- Aliases are publicly referenced names therefore no special authority or privilege is required to use an alias.
 - Access to the tables and views referred to by the alias, however, still require the appropriate authorization for the current context.
- In addition to table aliases, there are other types of aliases such as database and network aliases.
 - Aliases can also be created for nicknames.

SYS-ED/Computer Education Techniques, Inc.

1: 19

MQT: Materialized Query Tables

- MQTs cache results for faster access.
- DB2 automatically:
 - Caches data from other sources.
 - Reroutes queries to MQTs.
 - Maintains MQTs if the base data changes.

SYS-ED/Computer Education Techniques, Inc.

1: 20

Relational Database

DB2 Overview for Java Developers

Data Types

- When two data types can equally support the attribute and the requirements of the application, choose the data type with the least cost in term of performance.
- The cost of a data type depends on two factors:
 - The CPU required to manipulate the data.
 - The disk space required for storing the data.

SYS-ED/Computer Education Techniques, Inc.

1: 21

Data Types in DB2

- **char(n)**
 - Fixed length character string, with user-specified length n.
- **varchar(n)**
 - Variable length character strings, with user-specified maximum length n.
- **Int**
 - Integer, which is a finite subset of the integers that is machine-dependent.
- **Smallint**
 - Small integer, which is a machine-dependent subset of the integer domain type.
- **numeric(p,d)**
 - Fixed point number, with user-specified precision of p digits, with n digits to the right of decimal point.
- **real, double precision**
 - Floating point and double-precision floating point numbers, with machine-dependent precision.
- **float(n)**
 - Floating point number, with user-specified precision of at least n digits.

SYS-ED/Computer Education Techniques, Inc.

1: 22

Relational Database

DB2 Overview for Java Developers

Data Types

- DB2 data types in terms of their relative CPU cost for SQL processing is:
 - Least Expensive
 - SMALLINT and INTEGER
 - FLOAT
 - CHAR
 - VARCHAR
 - DECIMAL
 - DATE and TIME
 - TIMESTAMP
 - Most Expensive
- This is a rule-of-thumb ranking; since there can be differences resulting from input or output, local or distributed, and so on.

SYS-ED/Computer Education Techniques, Inc.

1: 23

VARCHAR

- The longest fixed length character data type supported by DB2 is CHAR(255).
 - While it is possible to specify CHAR(300), any character strings longer than 255 are stored using a variable length data type, such as VARCHAR.
- When the maximum length of a character attribute is fewer than 256 bytes, there are several factors to consider when deciding whether to implement a CHAR or VARCHAR data type.
 - Each variable length column includes a two byte length indicator in addition to the data string that is being stored.
- From a performance perspective, the savings from using variable length columns must overcome:
 - The disk space cost of the additional two bytes per column.
 - The CPU cost of handling it.
 - The potential additional logging when the row is updated.

SYS-ED/Computer Education Techniques, Inc.

1: 24

Relational Database

DB2 Overview for Java Developers

VARCHAR

- There may be an additional performance impact of implementing a variable length row, either through the use of variable length columns or compression.
 - Whenever a variable length row is updated, the new row length could be longer than the original row length.
 - There may not be sufficient space on the page for the updated row.
- When this happens, DB2 relocates the row on another page and places a pointer to it on the original page in the RID: Record ID.

SYS-ED/Computer Education Techniques, Inc.

1: 25

VARCHAR

- This results in an additional getpage and possibly an additional I/O for any reference to this row.
 - This row remains relocated until it is updated to a size that fits on the original destination page, or until the table space is reorganized.
- Increasing the PCTFREE parameter of the table space can reduce the chance of relocated rows.
 - However, it also reduces the disk savings of VARCHAR if this is the only reason for the increased free space.

SYS-ED/Computer Education Techniques, Inc.

1: 26

Relational Database

DB2 Overview for Java Developers

NULL

- All data types include the null value.
 - Distinct from all not null values, the null value denotes the absence of a value.
 - A DB2 column allows nulls by default if the NOT NULL clause is omitted.
 - Logical attributes that are mandatory should always be implemented as columns that specify NOT NULL or NOT NULL WITH DEFAULT.
- Allowing null values in a column does not save space.
 - To the contrary, a one byte null indicator is added to the row for each nullable column.
- As a consequence of the additional coding required to handle nulls in the application, they should not be used indiscriminately.

SYS-ED/Computer Education Techniques, Inc.

1: 27

Derived Columns

- A derived column is a column whose value can be derived using the value from other columns in the same or related rows.
- The typical reasons for using derived columns are to improve the performance of application processes in the following ways:
 - Reduce the number of tables or rows being accessed.
 - Improve access path selection and the performance of data retrieval.
 - Achieve greater control of the point in time when data is derived.
- Any time a derived column is used, there is a possibility that a derived column can become out of sync with the source data from which it is derived.

SYS-ED/Computer Education Techniques, Inc.

1: 28

Relational Database

DB2 Overview for Java Developers

Aggregation

- Aggregation is the process of summarizing data from multiple rows and possibly multiple tables into a single value or set of values.
 - By design, DB2 can perform aggregation functions through the power of SQL.
 - However, the cost in terms of CPU consumption and elapsed time of these aggregation functions may be prohibitive.
 - In some cases, there may not be sufficient work space available to complete the query.
 - It may be necessary to add derived columns or tables to the physical design in order to store aggregated values.
 - Another option is the use of MQT: materialized query tables for supporting an application's aggregation requirements.