WIKIPEDIA

Cardinality (SQL statements)

In SQL (Structured Query Language), the term **cardinality** refers to the uniqueness of data values contained in a particular column (attribute) of a database table. The lower the cardinality, the more duplicated elements in a column. Thus, a column with the lowest possible cardinality would have the same value for every row. SQL databases use cardinality to help determine the optimal query plan for a given query. [1]

Values of cardinality

When dealing with columnar value sets, there are three types of cardinality: high-cardinality, normal-cardinality, and low-cardinality.

High-cardinality refers to columns with values that are very uncommon or unique. High-cardinality column values are typically identification numbers, email addresses, or user names. An example of a data table column with high-cardinality would be a USERS table with a column named USER_ID. This column would contain unique values of 1-n. Each time a new user is created in the USERS table, a new number would be created in the USER_ID column to identify them uniquely. Since the values held in the USER_ID column are unique, this column's cardinality type would be referred to as high-cardinality.

Normal-cardinality refers to columns with values that are somewhat uncommon. Normal-cardinality column values are typically names, street addresses, or vehicle types. An example of a data table column with normal-cardinality would be a CUSTOMER table with a column named LAST_NAME, containing the last names of customers. While some people have common last names, such as Smith, others have uncommon last names. Therefore, an examination of all of the values held in the LAST_NAME column would show "clumps" of names in some places (e.g. a lot of Smiths) surrounded on both sides by a long series of unique values. Since there is a variety of possible values held in this column, its cardinality type would be referred to as normal-cardinality.

Low-cardinality refers to columns with few unique values. Low-cardinality column values are typically status flags, Boolean values, or major classifications such as gender. An example of a data table column with low-cardinality would be a CUSTOMER table with a column named NEW_CUSTOMER. This column would contain only two distinct values: Y or N, denoting whether the customer was new or not. Since there are only two possible values held in this column, its cardinality type would be referred to as low-cardinality. [2]

See also

- Cardinality (mathematics)

References

1. https://docs.oracle.com/cd/B28359_01/server.111/b28286/functions015.htm#SQLRF06305
2. https://www.ibm.com/developerworks/data/library/techarticle/dm-1309cardinal/

Retrieved from "https://en.wikipedia.org/w/index.php?title=Cardinality_(SQL_statements)&oldid=819678221"

This page was last edited on 10 January 2018, at 18:32.