The TABLESAMPLE Clause: Excerpts From SQL:2003

Concepts

In a `<table reference>`, `<sample clause>` can be specified to return a subset of result rows depending on the `<sample method>` and `<sample percentage>`. If the `<sample clause>` contains `<repeatable clause>`, then repeated executions of that `<table reference>` return a result table with identical rows for a given `<repeat argument>`, provided certain implementation-defined conditions are satisfied.

Syntax

```
<table reference> ::= <table factor> | <joined table>
<table factor> ::= <table primary> [ <sample clause> ]
table primary> ::= <table or query name> [ [ AS ] <correlation name> ]
<sample clause> ::= TABLESAMPLE <sample method> <left paren> <sample percentage> <right paren>
[ <repeatable clause> ]
<sample method> ::= BERNOULLI | SYSTEM
<repeatable clause> ::= REPEATABLE <left paren> <repeat argument> <right paren>
<sample percentage> ::= <numeric value expression>
<repeat argument> ::= <numeric value expression>
```

General Rules

Let `TP` be the `<table primary>` immediately contained in a `<table factor>` `TF`. Let `RT` be the result of `TP`. Case:

1. If `<sample clause>` is specified, then:
   (a) Let `N` be the number of rows in `RT` and let `S` be the value of `<sample percentage>`.
   (b) If `S` is the null value or if `S < 0` (zero) or if `S > 100`, then an exception condition is raised: “data exception — invalid sample size”.
   (c) If `<repeatable clause>` is specified, then let `RPT` be the value of `<repeat argument>`. If `RPT` is the null value, then an exception condition is raised: “data exception — invalid repeat argument in a sample clause”.
   (d) Case:
      i. If `<sample method>` specifies `BERNOULLI`, then the result of `TF` is a table containing approximately `(N * S/100)` rows of `RT`. The probability of a row of `RT` being included in result of `TF` is `S/100`. Further, whether a given row of `RT` is included in result of `TF` is independent of whether other rows of `RT` are included in result of `TF`.
      ii. Otherwise, result of `TF` is a table containing approximately `(N * S/100)` rows of `RT`. The probability of a row of `RT` being included in result of `TF` is `S/100`.
   (e) If `TF` contains outer references, then a table with identical rows is generated every time `TF` is evaluated with a given set of values for outer references.

2. Otherwise, result of `TF` is `RT`. 