Appendix

Estimating Disk Space Requirements

Objectives
In this appendix you will learn:

- How to estimate disk space requirements for Oracle.

In Step 4.4 of the physical database design methodology presented in Chapter 17, we stated that it may be a requirement that the physical database implementation can be handled by the current hardware configuration. Even if this is not the case, the designer still has to estimate the amount of disk space required to store the database, in the event that new hardware has to be procured. Estimating disk usage is highly dependent on the target DBMS and the hardware used to support the database. In general, the estimate is based on the size of each row and the number of rows in each table. The latter estimate should be a maximum number, but it may also be worth considering how each table will grow, and modifying the resulting disk size by this growth factor to determine the potential size of the database in the future.

In this appendix, we illustrate this process for estimating the size of non-clustered tables created in Oracle9i on an NT platform (we examined Oracle in some detail in Section 8.2). The formulae are based on tables when users are not deleting or updating rows, and there is no concurrency. In general, calculations for the amount of space required by a table involve multiplying the number of rows by the size of an average row. The size of an average row is a sum of the average column sizes plus any overhead introduced by the system. The size of a column is partly determined by the size and type specified by the user.

Estimating the space for non-clustered tables
There are four steps to calculating the space required by a non-clustered table:

1. calculate the total block header size;
2. calculate the available space per data block;
3. calculate the space used per row;
4. calculate the total number of rows that will fit in a data block.
(1) Calculate total block header size
The first step is to calculate the size of the block header:

\[
\text{totalBlockHeaderSize} = \text{fixedHeaderSize} + \text{fixedTransactionHeader} + \text{variableTransactionHeader} + \text{dataHeader}
\]

\[
\text{fixedHeaderSize} = \text{KCBH} + \text{UB4}
\]

\[
\text{fixedTransactionHeader} = \text{KTBBH}
\]

\[
\text{variableTransactionHeader} = \text{KTBIT} \times (\text{INITTRANS} - 1)
\]

\[
\text{dataHeader} = \text{KDBH}
\]

where the parameters KCBH, UB4, KTBBH, KTBIT, KDBH can be obtained from the system table v$type_size, and INITTRANS is the initial number of transaction entries per object. For a non-clustered table on an NT platform with INITTRANS = 1, we would get:

\[
\text{totalBlockHeaderSize} = (20 + 4) + 48 + 14 = 86
\]

(2) Calculate available data space per data block
Next, we calculate the amount of space available within a block for data as:

\[
\text{availableDataSpace} = \text{ROUNDUP}((\text{blockSize} - \text{totalBlockHeaderSize}) \times (1 - \text{PCTFREE}/100) - \text{KDBT})
\]

where PCTFREE is the percentage of space reserved in a block for updates (see Section 8.2.2). For a non-clustered table on an NT platform with PCTFREE = 10, we would get:

\[
\text{availableDataSpace} = (2048 - 86) \times (1 - 10/100) - 4 = 1766 - 4 = 1762
\]

(3) Calculate space used per row
We now calculate the combined data space required for an average row. This depends on:

- the number of columns in the table definition;
- the data types used for each column.

The total column size, including byte lengths, is calculated as:

\[
\text{totalColumnSize} = \text{columnSize} + (1, \text{if column size} < 250, \text{else} \ 3)
\]

The row size is then:

\[
\text{totalRowSize} = \text{rowHeaderSize} + \sum \text{totalColumnSize}
\]

where rowHeaderSize = 3 bytes per row of a non-clustered table (4 bytes per row of a clustered table). The minimum row size for a non-clustered table is 9 bytes; Therefore, if the calculated value is less than the absolute minimum row size, use the minimum value as the row size. For example, for the following table:

```
CREATE TABLE table1 (a char(5), b DATE, c CHAR(10));
```

we would get:

\[
\text{totalRowSize} = 3 + \sum (6 + 8 + 11) = 28
\]
(4) Calculate total number of rows per data block

We can calculate the total number of rows that will fit into a data block as:

\[
\text{noRowsPerBlock} = \text{ROUNDDOWN}(\frac{\text{availableDataSpace}}{\text{totalRowSize}})
\]

In this case, we get:

\[
\text{noRowsPerBlock} = \text{ROUNDDOWN}(\frac{1762}{28}) = 62
\]

Finally, we calculate the space required for a table by multiplying this value by the estimated number of rows in the table. It should be remembered that this is an approximation, so it is recommended allowing a 10–20\% margin of error. A similar procedure can be used to estimate the size of clustered tables and indexes. The interested reader is referred to the Oracle documentation set.