October, 1996

Advisor Answers

Q: I recently had to deal with a large set of imported data. All the records were supposed to be unique, but when I did a SELECT DISTINCT on the ID field, I came up with a few less records than there were in the original table. That told me I had some duplicate ID's, but how could I identify them? I ended up indexing on the ID field, and then writing a rather kludgy SCAN routine that compared each record with its immediate neighbors. It worked, but surely there's a more elegant way to do this.

–Stein Goering (via the Internet)

A: You can get SELECT to do the whole job for you here. Rather than SELECT DISTINCT, use the GROUP BY clause to find IDs that appear more than once. The basic idea is to use a query like:

```
SELECT Id, CNT(*) ;
   FROM MyTable ;
   GROUP BY Id ;
   HAVING CNT(*) > 1 ;
   INTO CURSOR Dups
```

The resulting cursor contains the ID and the count for every ID that appears more than once in the original table. To find the duplicate records, you can set a relation from the cursor back into the original table, like:

```
SELECT MyTable
SET ORDER TO Id

SELECT Dups
SET RELATION TO Id INTO MyTable
```

Now you can BROWSE the duplicated records and figure out what went wrong.

–Tamar