SQL in Spreadsheets

Query
We have three columns of data with headers in a spreadsheet and wish to compute a query:

![Image of a spreadsheet with headers and data]

Figure 1: The query and the desired spreadsheet

Why do we want this?
- Cheap, simple alternative to real databases.
- Already supported in cloud computing.
- Familiar to spreadsheet users — no entry barrier.

What about other spreadsheets?

It’s doable, immediately, at no cost!
Columns of identical formulas, which compute SQL queries.

Even more
- Spreadsheet formulas can implement all of SQL (including joins)
- No macros necessary
- Works in Excel, OpenOffice.org, gnumeric, Google docs, etc.
- Sufficiently effective for a few thousands tuples

Spreadsheet stores data and executes queries. It is a database engine!

Google docs QUERY and FILTER solution

We achieve our goal in two steps.

**QUERY**

![Image of a spreadsheet using Google Docs query]

Figure 2: QUERY in Google spreadsheet. HAVING clause not allowed.

**FILTER**

![Image of a spreadsheet using Google Docs filter]

Figure 3: FILTER expresses HAVING clause.

No join, no union, no difference in the language.

Performance

Horizontally tuples in thousands. Vertically time in seconds.

![Graph showing performance results]

Figure 5: Many-to-many equijoin in Microsoft Excel 2010 beta.
Six experiments, 10000 or 20000 rows of formulas, joined tables from 300 to 1000 tuples.

Figure 6: Many-to-one equijoin in Microsoft Excel 2010 beta.
50000 rows of formulas, foreign key table with 5000 rows (solid) and 10% of the large table (broken).

Figure 7: Average income query in Microsoft Excel 2010 beta.
10000 rows of formulas.

Acknowledgments: Many thanks to Jan Van den Bussche and Krzysztof Stencel.