

## Query Design

The runners use a MySQL statement to query the database. When designing these query statements, the most important aspect is accuracy. To make sure we are getting the correct data, we built tests that use a custom database. This database contains a small amount of known data for which we know the expected result that the query will return. When we get all these tests to pass, we know that the query is returning accurate results.

Another important aspect of the query design is performance. If we made a page that returned accurate results, but it took five minutes to run, then it would not be acceptable. We found that the best place to increase performance was at the query level. Two queries may return the same exact same result, but one may be much slower because of how the result is assembled in the query. This led us to develop multiple versions with the same output data. We then tested these queries individually on a large amount of data, and chose which one to use based on which query performed the fastest. We time the queries using human judgment, as we are concerned with perceptible improvements, not improvements on the scale of a few seconds.

Following is a simple example of the performance improvement process. Let's say that we have data on whether some students are sleeping during our final presentation. Also let's say that we want these results grouped by which group the students are in (Spatial or Toilers2). We might come up with an expected result in a format like what is shown in Figure 5.

<b>Group</b>	<b>Time</b>	<b>Sleeping?</b>
Spatial	<b>Error! Reference source not found.</b>	No
Toilers2	<b>Error! Reference source not found.</b>	yes

Figure 1

In our MySQL tables we might have data like what is shown in Figure 6. The numbers in the names and group column are references to other tables in order to minimize the storage of duplicate names or groups.

<b>Name</b>	<b>Time</b>	<b>Sleeping?</b>	<b>Group</b>
1	<b>Error! Reference source not found.</b>	No	1
2	<b>Error! Reference source not found.</b>	No	1

3	<b>Error! Reference source not found.</b>	yes	2
4	<b>Error! Reference source not found.</b>	no	2

Figure 2

From this point you could formulate your result in two different ways. You could make a simple query to retrieve the group names. The names are then grouped as shown in Figure 7. You could also make a slightly more complicated query and group by the group reference numbers first, and then retrieve the group names later as shown in Figure 8.

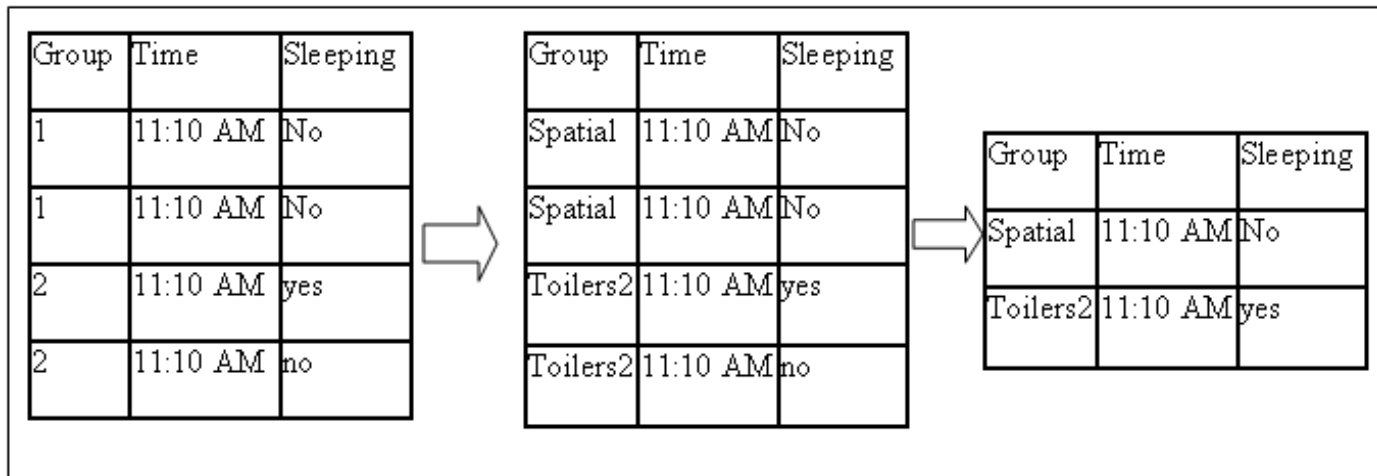


Figure 3

If the group names are retrieved like in Figure 8, the number of group name retrievals that are performed is cut in half, which in theory, as well as in practice, performed much faster than the query shown in Figure 5.

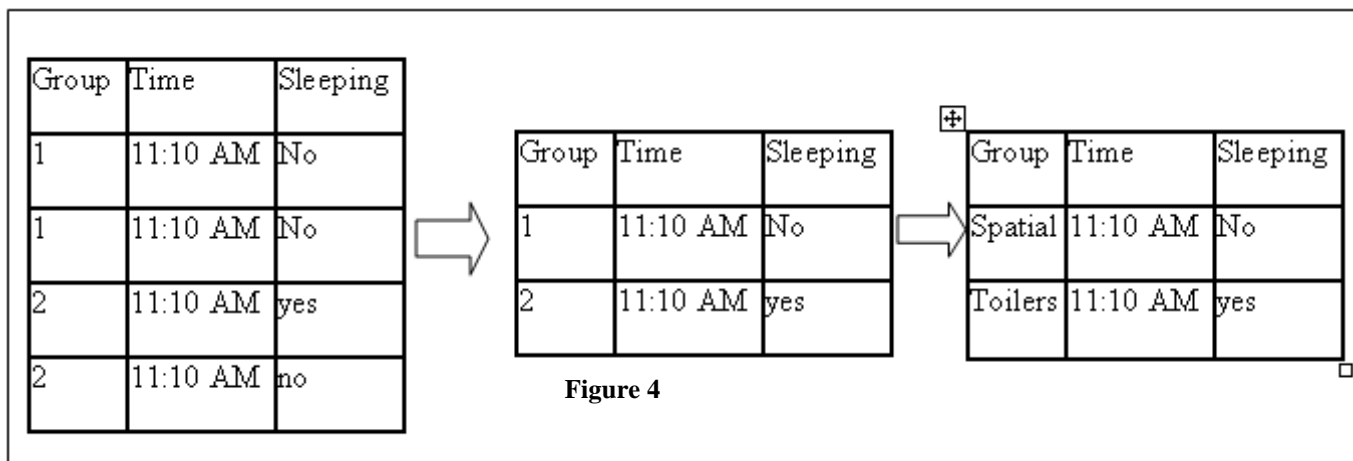


Figure 4