

You can use backups for things other than restoring

 devblogs.microsoft.com/oldnewthing/20120110-00

January 10, 2012



Raymond Chen

A customer wanted to know the internal file format of Visual SourceSafe databases. (That wasn't the actual question, but I've translated it into something equivalent but which requires less explanation.) They explained why they wanted this information:

We are doing some code engineering analysis on our project, so we need to extract data about every single commit to the project since its creation. Things like who did the commit, the number of lines of code changed, the time of day... We can then crank on all this data to determine things like *What time of day are most bugs introduced?* and possibly even try identify bug farms. Since our project is quite large, we found that generating all these queries against the database creates high load on the server. To reduce the load on the server, we'd like to just access the database files directly, but in order to do that, we need to know the file format.

Oh great, directly accessing a program's internal databases while they're live. What could possibly go wrong?

I proposed an alternative:

Take a recent backup of your project and mount it on a temporary server as read-only. Run your data collection scripts against the temporary server. This will spike the load on the temporary server, but who cares? You're the only person using the temporary server; the main server is unaffected. After you collect all your data from the temporary server, you can then perform a much smaller number of queries against the live server to get data on the commits that took place since the last backup.

[Raymond Chen](#)

Follow

