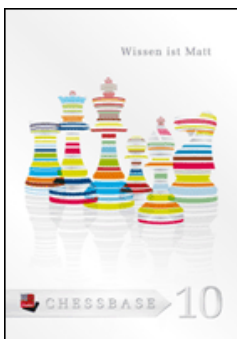




COLUMNISTS

ChessBase Cafe

Steve Lopez



CHESS THEATRE

Play through and download the games from [ChessCafe.com](#) in the [DGT Game Viewer](#).

[The Complete DGT Product Line](#)

One Database or Many?

Before we continue with an examination of ChessBase 10 features, we should take a brief detour and address a question that every chess database user will ask at some point in their chess career: “Should I maintain a single master database or should I have several databases instead?”

It’s a valid question, and one that doesn’t have a single “right” or “wrong” answer. It’s all a matter of personal preference, and the decision may even be somewhat skewed by the user’s specific computer hardware.

We’ll start with a little bit of history. I’ve been employed by various ChessBase vendors since 1992, the days of DOS 5.0 and ChessBase 4. The only playing program ChessBase offered at that time was Fritz1, then called Knightstalker. The Internet was available, of course, but access to it was fairly pricey for many people, and this was in the days when a 14.4kps modem was still considered adequate. Desktop computers had 5.25” floppy drives as standard equipment, and hard drives were typically in the 100Mb to 250Mb storage range; most laptops still had 40Mb hard drives as their standard.



Advertisement for Fritz 1

It wasn’t a time when people could download thousands of games every week. In fact, the way most people received their chess data was through subscription services such as *ChessBase Magazine* or the thrice-yearly *Informant* disks. The largest ChessBase 4 starter package came complete with 33,000 games.

That’s right – 33,000 games. I mention this to illustrate how drastically things have changed between then and now.

In those days, it was pretty common for players to maintain a single master database of games. Each time they received more games, be it through a subscription service or because they’d purchased a floppy disk containing several thousand games of a particular opening or from a certain historical period, they’d just merge them into their master database. Sure, they’d sometimes find that there were duplicate games in

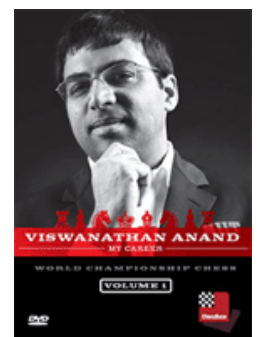
Check out these bestselling titles from [USCFSales.com](#):



Fritz 11



ChessBase Magazine



Viswanathan Anand

[My Career, Vol. 1](#)

by Viswanathan Anand

it, but they could run a utility on the database to remove them and usually have the process complete overnight on a 386 or 486 computer.

Faster computers, larger hard drives, and Internet access changed everything. By the end of the decade, Internet users could watch games live as they were being played. Chess players no longer had to wait months for the latest theoretical novelty to be published in the *Chess Informant*; TN's were widely available within hours or even faster. Online publications such as *The Week in Chess* not only furnished news reports, but allowed readers to download more games faster than ever before.

Chess databases stopped being measured in the tens of thousands of games. By the start of the twenty-first century, half-million game databases had become standard in playing programs, even in mass-market chess software aimed at casual players. The advanced database programs were shipping with databases containing millions of games.

There was an interesting side effect to this "information explosion": even though computer processors had become far faster over those years, the sheer number of available games and the consequent increasing size of chess databases continued to outstrip the ability of processors and software to quickly identify and eliminate duplicate games from large databases.

In short, even though computers are lightning-fast these days, removing duplicates from a large database containing millions of games can still take hours, even becoming an overnight (or still longer) process.

That's the primary reason why a ChessBase user has to make the fundamental decision of whether to maintain a single master database or several smaller databases. There are, however, other considerations as well.

One of the reasons a user might prefer multiple databases is because annual data integration is easier when one can just replace one database with another. For example, if one doesn't add games or additional notes to, say, the *ChessBase Opening Encyclopedia*, one can simply delete the old version when a new version is purchased. The same idea applies to the annual databases such as the [Big Database](#), [Mega Database](#) or the [Correspondence Database](#).

It's possible to search multiple databases at a single pass in ChessBase (by holding down the CTRL key while single-clicking the icons for the databases one wishes to search). For example, if you're researching a particular opening, you could highlight multiple databases and perform the search, then copy the results into a single new database devoted to the games of that opening. You could then run the "Kill Doubles" function on just that small subset of games. This takes just a few moments, instead of the hours required for a huge master database. In this case you're sure to not miss any relevant information, since several databases can be searched at one time, and you can copy it to a single location, but at the cost of additional disk space utilized.

That's the downside of this approach, occurring when hard disk space is at a premium. Although ChessBase data is remarkably compact, you still might not want to waste disk space on even a small number of duplicate games, thus making the use of a single master database desirable (even at the cost of the extra time required to maintain it).

Another argument in favor of a single master database is the presence of several ChessBase features that provide statistical information or other useful data based on the contents of a *single* master "reference" database. These features include the Opening Report and the "Reference" tab in the game notation pane of a game window. Using these "one click" data collators makes it desirable to have all of one's data stored in a single

database.

Flipping the coin over once again, though, a single large database can be time-consuming to maintain. Anytime you add data that is not brand-spanking new to a master reference database requires you to run the “Kill Doubles” function on that database or risk having statistical results skewed (however slightly) by the presence of duplicate games. And, as we’ve already discussed, this function may require hours to complete, even on a very fast home computer, when the total database size is in the four to five million game range.

Whenever I bring this subject up, I’m invariably asked which approach I prefer. Personally, I’ve always kept multiple smaller databases for the reasons mentioned earlier, as well for the ability to segregate games by source (that is, where the data came from, whether a commercial vendor or a particular website or e-publication, etc). However, this is a *personal preference*, not a definitive answer that “must” be followed by everyone.

The choice, ultimately, rests with you. The purpose of this column has been to provide food for thought to aid you in making an informed decision.

Until next month, have fun!

All the software described in this column, as well as many more programs, are available in the [USCFSales Online Catalog](#).

ChessBase Cafe welcomes readers’ questions. Send one along and perhaps it will be answered in an upcoming column. Please include your name and country of residence.

[Yes, I have a question for Steve!](#)

© 2008 Steve Lopez. All Rights Reserved.

 [TOP OF PAGE](#)

 [HOME](#)

 [COLUMNS](#)

 [LINKS](#)

 [ARCHIVES](#)

 [ABOUT THE
CHESS CAFE](#)

[\[ChessCafe Home Page\]](#) [\[Book Review\]](#) [\[Columnists\]](#)
[\[Endgame Study\]](#) [\[The Skittles Room\]](#) [\[Archives\]](#)
[\[Links\]](#) [\[Online Bookstore\]](#) [\[About ChessCafe.com\]](#) [\[Contact Us\]](#)

© 2008 CyberCafes, LLC. All Rights Reserved.
"ChessCafe.com®" is a registered trademark of Russell Enterprises, Inc.