

# DBToaster breaks up data jams in server farms

September 21 2012, by Lionel Pousaz

---



In gigantic server farms around the world, billions of database entries are queried every second. EPFL researchers have developed a system that drastically improves the circulation of this flow of information. The economic and environmental benefits are considerable. Credit: EPFL

Databases have revolutionized the business world. Every bottle of shampoo you buy, every purchase you make, is just one more data point sent out to your bank's and your supermarket's servers. This enormous quantity of detailed information allows merchants to optimize their inventories and displays and bankers to optimize the flow of money. Gigantic farms of servers are deployed in an effort to keep up with this breakneck pace of information storage and transfer. Researchers in EPFL's DATA Laboratory have developed DBToaster, a system that speeds up the pace of operations by a factor of 100 – 10,000. The latest version has just been made available on [www.dbtoaster.org](http://www.dbtoaster.org).

"Ten years ago, [CERN](#) set up one of the world's largest databases," explains EPFL professor Christoph Koch, DBToaster's creator. "Today, your average supermarket has a bigger system." This inflation has escalated dramatically, to the point that optimizing databases has become an environmental issue. In the U.S., electricity use by [server farms](#) is growing exponentially, currently representing 2% of total electricity consumption.

## **Avoiding data jams by accelerating the flow of data**

In a classic database, data are handled in a series of successive packets. For example, say a bank wants a list of all its clients who live in Zurich who have a balance of at least 5,000 francs. The user queries the database by selecting certain criteria. This request is translated into a series of [mathematical operations](#). Because every banking transaction results in a separate database entry, the amount of information that must be sorted is phenomenal - the first operation has to search through billions of entries. The resulting data set is then sorted by the second operator, and so on, until the list is reduced to the clients desired.

The data are so vast that often the server's RAM is not large enough to temporarily store initial results, causing a data jam. The server must temporarily store intermediate results on the hard disk before sending them on to the next operator. This slows things down considerably, because accessing the hard disk is 10,000 times slower than accessing RAM. It also requires much more electricity.

The EPFL scientists were able to get their system to compile successive operators as one single operator. This extremely complex operation makes it possible to store huge intermediate results. In doing so, DBToaster is able to efficiently prevent data jams.

## Keeping queries in memory so you don't have to reinvent the wheel

DBToaster has a second innovation, as well. The researchers took into account the fact that queries are often repetitive. "In general, the same operator is used many times within brief periods of time," explains Koch. Rather than having to recalculate everything each time, the system keeps the preceding result in memory and merges it with new entries. "The big innovation with DBToaster is its ability to generate efficient code that manages to figure out how previous queries should be changed in order to be updated." In this way, only recently entered data has to be queried, rather than billions of entries.

DBToaster is available online for no charge. Financial institutions, in particular, are enthusiastic about the system. According to Koch, DBToaster "enables analytical processing in real time, which financial institutions need to perform automated trading or to enforce regulatory compliance – for instance to detect patterns of money laundering in their streams of financial transactions." But the benefits go farther than this. As data processing consumes escalating amounts of power, DBToaster is a solution that can be easily deployed on existing servers to reduce their [electricity consumption](#) and mitigate their impact on the environment.

Provided by Ecole Polytechnique Federale de Lausanne

Citation: DBToaster breaks up data jams in server farms (2012, September 21) retrieved 24 April 2024 from <https://phys.org/news/2012-09-dbtoaster-server-farms.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--