

IBExpert Benchmark: more astonishing results

A performance comparison Firebird 3 and Firebird 4

Holger Klemt, November 2021

Many readers were a little confused after we reported that our standard benchmark test for Firebird was showing slower figures for Firebird 4 compared to Firebird 3.

As we wrote in our recent performance White Paper, we were also astonished, but the numbers didn't lie.

Thanks to the Firebird core developer team, who were able to reproduce what our benchmark does, we received their feedback. So thanks to Vlad, Pavel and Alexey, we were also able to understand why the new Firebird 4 architecture will by default have several problems with our test procedure.

After reviewing our test database architecture (which was last changed in 2013, but so far has always performed correctly for Firebird <=3), we decided to change it as recommended, at least the important parts, by the core developers' document.

The trigger that creates multiple locks, by simply updating the number of items on stock on product record level, is in fact a method that I would never recommend in real world multi-threaded applications, but it was written that way years ago and as every new Firebird version, especially with the step to Firebird 3, showed better results than before, we had no reason not to use it, since - as we have always stated - it is designed to evaluate the speed of one Firebird server hardware/OS with another Firebird server hardware/OS using the same Firebird version. If you compare a Celeron HDD-based machine with a i9 NVMe SSD-based machine, you can directly see a huge difference in the results. It is much more complicated to compare a Xeon-based virtual host using a specific external NAS with another almost similar Xeon-based virtual host using a different external NAS.

This is what the benchmark is designed for. If internally it perhaps does some strange things, it still easily shows the better solution by a result in percent, which can convince management to invest the money in a more powerful solution.

We were astonished that Firebird 4 was significantly slower than Firebird 3, and since we understood that there was a big change in the basic architecture, we were a little afraid that Firebird 4 was more or less a step back rather than a step forward.

We saw no direct problem with our way of testing, since it has been in use for almost 20 years and, so far, has always done what it was intended for, i.e. compare the speed and find the best setup.

After reading the valuable answer given by the core developers, we now view it differently. And as we all know that software is a dynamic business, we decided to change the logic inside our benchmark database to use an insert trigger, in order to have the real quantity of stock available only by reading a sum on all records for a specific product. We accept that this method is considerably more powerful in Firebird 4, but now also use it for any other benchmark result, so even a test in Firebird 2.5 and Firebird 3.0 now does exactly the same as with Firebird 4.0.



Our free Benchmark Tool will continue to only use the non-modified default firebird.conf file. Anyone who wishes to change this can do so. We also distribute all 3 Firebird versions as the newest 64-bit versions. If you download this newest version, you will see *IBExpert Benchmark 2.0* in the application header. And the version included in the current IBExpert IDE, which can be used also for testing remote machines with Windows or Linux, now also displays that header.

Our next steps are now a complete retesting of all the tests that were documented in our initial test comparison White Paper, and then publication of these in a new document.

And as a short preview: Firebird 4 generally does a great job with the new *IBExpert Benchmark 2.0* and mostly offers better results when compared to any older Firebird version (although Firebird 3 is still stronger in some cases), so the core developers were right to not accept our results as generally valid.

They were valid for our test procedures at the time, proven by what we know from empirical databases using a similar implementation, but in fact, when we are reviewing customer database structures and find this implementation (multiple updates on single records from multiple connections), we will always recommend changing it as we have done in our own test database: avoiding multi-threaded single record blocking updates by replacing them with non-blocking inserts, and checking results using sum functions, and then later compressing the multiple records to new single records when there is no longer a lock problem, is a much better design.

Perhaps the public discussion also found more readers than the internal dialogue; now we definitely no longer have a reason not to recommend a change to Firebird 4, and we thank the developer team for their work, not only on clarifying this issue. This can now be a good starting point for all application developers, who complain about the speed of their application with Firebird databases.

The current downloadable version of IBExpert and the Benchmark Tool are already based on the new 2.0 database and now generally show better results for Firebird 4, not because it was specifically designed for Firebird 4, but because the overall architecture of the test database has been improved for all Firebird versions.