IBExpert White Paper

Firebird White Paper

Minimizing Database size, Should you store data files in the file system?

Holger Klemt, February 2018

One of the significant Firebird features since years is the effective management of blob information.

The Movie title "The BLOB" with the subtitle "Indescribable, indestructible, nothing can stop it" is not very helpful, luckily it has nothing todo with Blobs in database systems. The "Thing" from the movie with Steve McQueen is some kind of killer amoeba from space that eats and digests anything that comes in its way.

A blob in a firebird database can be better understand as a "Binary Large Object". That's in the spirit of data management everything that falls under the concept of a data file. Every smaller or larger size of binary data chain can be stored on the file system for later usage. Simple text files, pictures, videos and much more are suitable for this kind of storage.

IBExpert White Paper

For documentation in ERP (Enterprise Resource Planning) and CRM (Customer Relationship Management) systems, more and more people are switching to not just printing the documents directly to a printer, but also storing them as PDF files in order to send them for example as an email attachment.

A lot of data can come together at regular day time operations. For a customer with approximately 40 employees who uses our BRP software for order processing and production planning, at least 500 PDF documents per day need to be sufficiently. These documents can contain several individual pages including photos to document production their projects.

The BRP-Software is at this client since 6 years in production and over this timeframe the average document size is determined to be 500 KB. As of today approximately 60000 documents with average size of 0.5MB are stored in the archive.

Many clients still think that there are advantages to store such data elements in the file system. Our experience with Firebird shows us that the data storage in the database will bring significant advantages.

1. Try to store 1 million files in a folder on a NTFS-Filesystem. You will quickly find out that tools like windows explorer are pretty useless because the folder barely opens.
2. Do you prevent a file from being overwritten or deleted by a user even though the user has write permission to the folder. All changes should be transactionally logged and undone, if needed. The recycle bin is overwhelmed very quickly because of the file size and emptied quickly or on the network drive not even available.
3. Do you keep multiple version of the same file and keep track of who did what and when? You will find out that especially in the network problems are created.

4. Despite appropriate write permissions, do you prevent files from being moved to other folders. Your order will break faster than you think and nobody knows who it was.
5. If a user accidentally starts an email attachment with one of the not so rare encryption trojans, you probably will not notice that this has already partially encrypted all important files on reachable file shares in the network and thus made them useless.
6. Prevent an employee from simply copying and taking home all the documents he has network access to on a local drive. Perhaps as a reminder of this former client relationships that he would like to use with his new employer…. Difficult as Edward Snowden and other have shown.
7. Try to search for a file in the file system. In the age of modern SSD's, the response time is not quite as catastrophic as it is without it, but if you also search for content, you should plan minutes or up to hours, because functional indexes are limited.

There are many other arguments against file system storage, so we think that storing files in the file system is the wrong way. Certainly there are possibilities in active directory or Linux file systems to limit the basic problems by restrictions. But protection agains the above mentioned problems is hardly feasible.

In our IBExpert Firebird 3 Bootcamps, we will present a combination of the following technologies and you will see that storing millions of PDF's or images in Firebird databases enables a very powerful integration of DMS (Data Management Technologies) in your application.

• How do I read and write BLOB data with calculated columns and updatable views
• How do I save and read the BLOB data in a separate database, without changes in my application and avoiding a too large production database

IBExpert White Paper

- How to I move the BLOB data from my BLOB database to a read- only history database on one or more different servers in an annual or monthly process?
- How do I process PDF files without any Adobe tools to preview the first or all pages? How do I extract the text from a PDF file without Adobe tools?
- How can I view PDF data in my windows application without using Adobe tools? How do I integrate a powerful OCR engine to create text, images, or non-text image PDF's?
- How to I care a powerful full text search engine based on these techniques?

If your time does not allow you to visit our IBExpert Firebird 3 Bootcamp, we offer a 12 hour remote education and project support by phone and TeamViewer/PCVISIT/GotoMeeting. The training takes place in 3 daily sessions of 4 hours each. The price for distance education is $2480 USD.

At the end…

Steve McQueen was offered $ 2500 or 10% of the profits. He took the $ 2500 because it was not expected that this movie would earn much. In the end, over $ 4 million was earned. Producer Jack H. Harris told film historian Tom Weaver in an interview that the movie eventually grossed $ 40 million.

…So you should not underestimate the value of BLOB's

Contact:
IBExpert
202 River Bend CT
32779 Longwood FL.
USA

Telephone: 407-462-7283 Monday to Friday 9am-6pm
Email: contact@ibexpert.com