



Firebird White Paper

Datenbankgröße oder: Sollte man Dateien nicht besser im Filesystem speichern?

Holger Klemt, Februar 2018

Eines der Merkmale, die Firebird schon seit Jahren sehr leistungsfähig macht, ist die sehr effektive Verwaltung von BLOB Daten.



Der Titel für den gleichnamigen Film "THE BLOB" von 1958 mit den Überschriften "Unbeschreibbar, Unzerstörbar, Nichts kann es stoppen" ist dabei wenig hilfreich, hat aber auch zum Glück gar nichts mit dem BLOB in Datenbanken zu tun.

Das Ding aus dem Film mit Steve McQueen ist eine Art Killeramöbe aus dem Weltall, die alles Mögliche auffrisst und verdaut.

Ein BLOB in einer Firebird Datenbank ist besser zu verstehen als Akronym für "Binary Large Object". Das ist im Sinne der Datenverarbeitung also alles, was man auch unter dem Begriff Datei kennt. Jede mehr oder weniger große binäre Datenkette kann in den meisten Betriebssystemen als Datei auf einem Datenträger abgelegt und bei Bedarf neu gelesen werden. Neben einfachen Textdateien sind auch Bilder, Videos und vieles mehr dafür geeignet.

Zur Dokumentation in ERP- und CRM-Systemen geht man immer mehr dazu über, die Dokumente nicht mehr nur direkt auf einem Drucker auszugeben, sondern auch als PDF-Datei abzulegen, um diese zum Beispiel als E-Mail Anhang zu versenden.

Im Tagesbetrieb kommt schon einiges an Daten zusammen. Bei einem Kunden mit ca. 40 Mitarbeitern, der unsere BRP-Software für Auftragsbearbeitung und Produktionsplanung einsetzt, sind das an jedem Arbeitstag mindestens 500 PDF-Dokumente. Diese Dokumente können, je nach Inhalt, auch mehrere Seiten umfassen und mit Fotos Besonderheiten für die Produktion dokumentieren.



Die BRP-Software ist bei diesem Kunden seit ca. 6 Jahren im Produktiveinsatz und über die Jahre hinweg hat sich ein Durchschnitt von 500KB pro Dokument als realistischer Wert ergeben. Aktuell belegen also etwa 600000 Dokumente zu je 0,5MB das Archiv. Durch Kopfrechnen kommt man dabei auf 600*0,5GB oder ca. 300GB Gesamtplatz, den das Archiv belegt.

Viele Kunden denken immer noch, dass es Vorteile hätte, derartige Dateiinhalte im Filesystem zu speichern. Unsere Erfahrung mit Firebird zeigt uns aber, dass die Speicherung in Datenbanken erhebliche Vorteile bringt.

1. Versuchen Sie doch einmal, einen Dateiordner in einem Windows NTFS-Filesystem mit einer Millionen Dateien zu füllen. Sie werden schnell feststellen, dass Werkzeuge, wie der Windows Explorer, dabei ziemlich unbrauchbar sind, weil sich der Ordner kaum öffnen lässt.

2. Verhindern Sie, dass eine Datei von einem Benutzer überschrieben oder gelöscht werden kann, obwohl der Benutzer Schreibrechte an diesem Ordner hat. Sämtliche Änderungen sollen transaktionsecht protokolliert und bei Bedarf rückgängig gemacht werden, wenn es erforderlich ist. Der Papierkorb ist dabei sehr schnell überfordert aufgrund der Dateigröße und schnell geleert oder auf dem Netzwerklaufwerk gar nicht erst verfügbar.

3. Bewahren Sie mehrere Versionen der gleichen Datei unter einem Dateinamen auf und protokollieren Sie, wer was wann gemacht hat. Sie werden feststellen, dass insbesondere im Netzwerk einiges an Problemen entsteht.

4. Verhindern Sie trotz entsprechender Schreibrechte das Verschieben von Dateien in andere Ordner. Ihre Ordnung ist schneller dahin als Sie denken, und keiner weiß, wer es war.

5. Sollte ein Anwender mal versehentlich einen E-Mail Anhang mit einem der mittlerweile nicht mehr so seltenen Encryption Trojaner starten, werden Sie vermutlich gar nicht merken, dass dieser schon alle wichtigen Dateien auf erreichbaren Dateifreigaben auch im Netzwerk teilweise verschlüsselt und somit unbrauchbar gemacht hat.

6. Verhindern Sie, dass ein Mitarbeiter einfach alle Dokumente, auf die er Netzwerkzugriff hat, auf ein lokales Laufwerk kopiert und mit nach Hause nimmt, - vielleicht als Gedächtnisstütze für seine ehemaligen Kundenbeziehungen, die er bei seinem neuen Arbeitgeber gerne mal nutzen möchte... schwierig, wie es schon Edward Snowden und andere gezeigt haben.

7. Suchen Sie einfach mal eine Datei im Dateisystem. Im Zeitalter moderner SSDs ist die Antwortzeit nicht mehr ganz so katastrophal wie ohne, aber wenn man dabei auch noch Inhalte durchsucht, sollte man durchaus Minuten bis Stunden einplanen, weil funktionale Indizes dafür nur begrenzt verfügbar sind.

Es gibt noch weitere Argumente gegen eine Speicherung im Filesystem, daher halten wir den Weg über das Filesystem für den falschen Weg. Sicherlich gibt es im Active Directory oder Linux Filesystemen Möglichkeiten, die grundlegenden Probleme durch Restriktionen zu begrenzen, aber ein Schutz gegen die oben genannten Probleme ist auch dadurch kaum realisierbar.

In unserem IBExpert Firebird 3 Bootcamp stellen wir eine Kombination der folgenden Technologien vor und Sie werden sehen, dass die Speicherung von Millionen von PDFs oder Bildern in Firebird-Datenbanken eine sehr leistungsfähige Integration von DMS-Technologien in Ihrer Anwendung ermöglicht.

- Wie lese und schreibe ich BLOB-Daten mit berechneten Spalten und aktualisierbaren Ansichten?



- Wie speichere und lese ich die BLOB-Daten in einer separaten Datenbank, ohne Änderungen in meiner Anwendung und unter Vermeidung einer zu großen Produktionsdatenbank?
- Wie verschiebe ich die BLOB-Daten aus meiner BLOB-Datenbank zu einer Read-Only History-Datenbank auf einem oder mehreren verschiedenen Servern in einem jährlichen oder monatlichen Prozess?
- Wie verarbeite ich PDF-Dateien ohne jegliche Adobe-Werkzeuge, um eine Vorschau der ersten oder aller Seiten zu erzeugen?
- Wie extrahiere ich den Text aus einer PDF-Datei ohne Adobe-Werkzeuge?
- Wie kann ich PDF-Daten in meiner Windows-Anwendung ohne Adobe-Werkzeuge anzeigen?
- Wie integriere ich eine leistungsstarke OCR-Engine, um Text, Bilder oder Nicht-Text-Bild PDFs zu erstellen?
- Wie erzeuge ich eine leistungsstarke Volltextsuchmaschine auf der Grundlage dieser Techniken?

Wenn Ihre Zeit keinen Besuch unseres IBExpert Firebird 3 Bootcamps erlaubt, bieten wir eine 12-stündige Fernausbildung und Projektunterstützung per Telefon und TeamViewer/pcvisit/GoToMeeting an. Die Ausbildung findet in 3 täglichen Sitzungen von jeweils 4 Stunden statt. Der Preis für die Fernausbildung beträgt 1990 €.

Zum Schluss ...

Steve McQueen wurden \$ 2500 oder 10% der Gewinne angeboten. Er nahm die \$ 2500, weil nicht erwartet wurde, dass dieser Film viel einspielen würde. Am Ende wurden über 4 Millionen Dollar verdient. Produzent Jack H. Harris ließ Filmhistoriker Tom Weaver in einem Interview wissen, dass der Film letztendlich 40 Millionen Dollar einspielte.

... man sollte also den Wert von BLOBs nicht unterschätzen!

Kontakt:

**IBExpert GmbH
Oldenburger Straße 233
26203 Wardenburg
Deutschland**

Telefon: **+49 (0) 4407 3148770** Montag bis Donnerstag von 9.00-17.00 Uhr; Freitag 9.00-14.30 Uhr
Fax: **+49 (0) 4407 3148770**

E-Mail: **sales@ibexpert.biz**