12 Common Mistakes while Backing Up Databases

This article was initially intended for Firebird DBMS developers and administrators, but contacts with administrators of other databases made it clear that most mistakes are common among them as well and literally everyone stumbles over nearly the same stones. If you can add something to this list (even something specific to a particular DBMS), contact us via our e-mail or post it in the comments.

Our company (www.ib-aid.com) supplies technical support and produces tools for restoring, backing up, optimizing and supporting DBMSs (mainly Firebird, but we also deal with MSSQL, PostgreSQL, InterBase, etc.) and we have accumulated a collection of mistakes related to the backup process as a result of numerous repairs and audits. All the issues described below are based on real-life cases with corrupted databases, lost and corrupted backup copies and disks, failed servers and other nightmares of administrators. We offer a general recommendation and a recommendation specific to Firebird for every issue.
We would like to let it be known so that administrators and developers could change their approaches to backup management and prevent possible problems.
So, let us get down to it.

1. Deleting the previous backup copy before a new backup copy is created

This mistake is most common among newbies who do not realize that the main purpose of a database backup copy is not just to create a database copy, but to make the downtime of an information system (an important part of which is the database) as short as possible.
As a result, the system remains unprotected from the moment the latest backup copy is deleted and to the moment the new one is created because the database does not have a single backup copy during this period. Since creating a backup copy may take quite a while, it is the perfect time for Murphy's law to take effect. This approach works especially well when it is combined with Issue 7 (see below).
Recommendations: do not delete the previous backup copy before the new one is created! (and do not make a new backup copy into an existing file).

Recommendation for Firebird: The FBDataGuard tool included in HQbird (an advanced distribution package of Firebird) deletes the earliest backup copy in the history only after a new one is created.

2. Overwriting an existing database while restoring it from a backup copy

This mistake is less common though the results may be much worse. If the backup copy has not been verified and turns out to be corrupted (see Issue 6), you will have neither the previous copy of the database nor a valid backup copy.
A mess like this usually happens on a Friday evening when things get hectic and when directions from the management get kind of contradictory. A bit of bad luck and a languorous weekend in the server room is there for you.
Firebird has kind of protection against this mistake – it will not be possible to restore a database from a backup copy with the help of the gbak utility if its default –create switch is on and if the specified file name points to an existing database. Unfortunately, there is a way to get around this protection: the –rep switch still allows you to overwrite the existing file.

Recommendation: never overwrite the file of a working database without a written directive from your management.
Recommendation for Firebird: Use FBDataGuard because it never overwrites the database file.

3. Using one-step backup/restore without using an intermediate backup file

Standard input/output streams make it possible to do a funny trick with a lot of DBMSs (including Firebird): implement a streaming backup with restoring the database from it at once. No intermediate backup file is created as a result. It is convenient for routine maintenance and for running a test restore operation (provided there is another backup copy available), but you must not use it for automatic backup!
For instance, if a serious disk failure occurs during this backup/restore process, the initial database may become damaged while no new database has yet been created. Of course, if you take into account Issue 1 and there is a database copy from the previous attempt, only the data created or updated in the database after that copy was created will be lost.
Recommendations: do not use one-step backup/restore in the automatic mode and always check the availability of a copy that is up-to-date enough in the manual mode.

4. Storing backup copies and the database on one and the same physical device

Many of you may find it funny that the advice we give is kind of childish – the ABC of backup. Right, that is true, but the database and the disk may end up being stored on one data storage system due to the popularity of virtual environments. And it will certainly fail at the most inappropriate moment. Plus there are still people who believe that nothing can happen with their data if they use RAID arrays (version 1 or higher :)). Besides, there are people who believe that some "brand" servers are failproof, but that is a special case.
Recommendations: do not store backup copies and the database on one device no matter how reliable it may seem.

5. No control over the successful completion of the backup process

It is a rather common mistake among both administrators and the heads of IT departments. If you do not check the results of the backup process, you might as well not perform it at all. You must receive notifications about the successfully completed backup process by e-mail or, even better, via texting as well. And the absence of such notifications is a sign of a problem!
An attentive reader who reached this point in our article (though it is too early to give a prize for that so far) may ask: 'But what does it have to do with the management?' Here is what - the administrator usually configures the backup process, but he finds it too boring to check notifications especially when they are stored in a separate folder so it is never too much to request additional reports about the status of the process. It is regarding the question who is to blame when it seems as if backup copies were there, but they are actually not there the moment you need them :)
! once combined with Issue 2, we have neither the database nor its backup copy.
Recommendations: use backup automation tools that can monitor successful and unsuccessful backup processes, notify users about problems and offer summary control tools (it is especially relevant when you need to control dozens and hundreds of backup processes on different servers).
Recommendation for Firebird: FBDataGuard checks whether the backup process has been completed and sends the corresponding notification. For systems with a lot of databases there

is second-level summary monitoring with the help of the Control Center tool that allows you to see the statuses of all monitored servers and databases on one page.

6. No backup validation

The fact that backup copies are stored somewhere does not mean that they can be read from there.
That is why you must regularly verify backup copies you create in order to make sure that they are not corrupted or copied to /dev/null.
Recommendation for Firebird: you can automate backup validation with the help of FBDataGuard.

7. No database health checks while using unverified backup copies

Usually, databases use several types of backup – dumps, regular backup copies, etc. Without going into details, we can single out two categories: verified and unverified. In case of Firebird, it is gbak and nbackup.
Gbak reads the entire database on the level of records in order to create a backup file and creates a database by inserting records into a new database thus verifying the backup copy (there are ways for errors to sneak into the restored copy, but that is another way for the database administrator to mess things up related to poorly organized migration) and the database itself (if it can be read from beginning to end, it is most likely not corrupted).
Nbackup (aka incremental backup) temporarily locks the main database file for updates (in the consistent state) and makes it possible to quickly copy the database file (fully or partially/incrementally).
In case of large Firebird databases (larger than 500 GB), it is advisable to use nbackup in order not to slow down user operations, but at the same time it is necessary to validate the database because unverified backup copies it creates are database page copies and if an error resides on the level of records (due to a RAM failure) or on the logical level, an unverified backup copy will contain it as well as the original database.
To avoid it, you should use online validation for the original database (online validation with the help of gfix is available starting from Firebird version 2.5.4 while our FBDataGuard tool supports online database validation for versions 1.5-2.5).
Also, it is advisable to perform verified backup once in a while (once a week, for instance) in addition to unverified backup.
Recommendation for Firebird: besides online health check, FBDataGuard allows you to test the backup restore process in the automatic mode.

8. No control over free space for backup copies

Actually, it is a classic mistake: if there is not enough space, backup copies occupy all the free space and the process ends with an error. Storing backup copies on one disk with the database may lead to an interruption in the operation of the database and storing them on the system disk may result in a system failure.
In combination with Issue 4, the best possible outcome will be the one when the system stops functioning because the database also needs free space, but it is occupied by backup copies. As to combination with Issues 5 and 2, it leaves us with neither the database nor its backup copy again.
Recommendations: use backup tools that predict the backup size and warn you about the possible lack of free space.
Recommendation for Firebird: FBDataGuard controls the size of free space for backup purposes and also the size of free space on the disk with databases as well as on the system disk.

9. No control over the time it takes to create a backup copy

The backup process took 40 minutes literally half a year ago and then suddenly it takes three hours already – why is that? The size of the database may have increased or a disk may have dropped out from your RAID array resulting in considerably slower write performance and all your backup copies may be about to depart from this world. Or a good colleague of yours may have run one more backup system at the same time (by the way, Firebird allows you to run several backup processes at once though it is not quite clear why one may need it at all).
If you do not control the time it takes to make a backup copy, you may overlook a newly emerged problem and miss the chance to fix it before it gets massive.
Besides, if the backup system does not monitor the statuses of backup tasks and runs them just according to schedule, you may easily "jump the gun", which means the situation when the system starts a new backup process while the previous one is not over yet.
Recommendations: use tools that control the time the backup process takes!
Recommendation for Firebird: FBDataGuard controls the time the backup process takes.

10. Backing up the database while operating system updates are being applied

It is a very common problem especially in combination with Issue 9 and enabled automatic Windows updates (by default, updates are applied at 3 a.m.). It lead to a slowdown at best, but if the operating system is restarted in order to apply the updates, the backup copy will be damaged. At least, the good news is that the operating system is not updated every day.
Recommendations: schedule operating system updates when they do not interfere with the backup process.

11. Backing up the database with the help of file backup tools or virtual machine backup tools while the database server is running

Many administrators forget that any DBMS has an active and complex cache that contains data being read and written while database files themselves are opened in the random access mode. That is why it is necessary to use special backup types instead of the mere file backup (including just copying database files) or the virtual machine backup. File backup tools read the database sequentially and it may take quite a long time, especially in case of large databases, so it is impossible to guarantee the integrity of the created backup copy.
Virtual machines may use the mechanisms of snapshots and Changed Block Tracking, but it is necessary to synchronize the created backup copies in order to get a consistent database backup copy because the backup copy will be inconsistent in case of any active write operations with the database at the moment of sorting out the collection of changed blocks.
For those who wish to back up their databases with the help of file or virtual machine backup tools, we can offer two methods:
1) completely shut down DBMS services and processes so that there is nothing in the cache,
2) use agents and/or scripts that switch the database to a special mode that makes it safe to copy the database file sequentially. For instance, there is a mechanism called VSS writer for MSSQL databases. On request, it switches the database to the snapshot-friendly mode at the moment when a snapshot is made. If you use mechanisms based on Changed Block Tracking, you yourself should make sure that the database is consistent at the moment of synchronization.
If you do not switch the database to the backup-friendly mode, the resulting database copy will look as if a hard reset (for instance, a power outage) occurred on the host computer. This level of reliability is absolutely insufficient for most businesses.

You can learn more about this in the article "Peculiarities of Working with Databases on Virtual Machines".
For Firebird, it is necessary to lock the main file of the database with the help of nbackup before the backup process starts and unlock it after the process is over. For other DBMSs there are similar tools for switching the corresponding modes on/off.
Some database administrators are sure that they can safely back up their databases with the help of standard file backup tools if the DBMS has a transaction log because only this log will be corrupted at most. It is a dangerous misconception that DBMS developers do not support.
The roots of this misconception are clear: aggressive advertising by the developers of virtual machines and backup tools usually fails to mention that databases as well as other intensively updated files require advanced configuration. Do not believe the hype - not all yogurts have equal benefits.
Recommendations: do not use file and VM backup tools without the corresponding automation tools.
Recommendation for Firebird: use FBDataGuard (from the HQbird distribution package), it provides integration with VSS-aware backup tools.

12. Replacing backup with replication

Data backup and data replication are used to increase the reliability and to prevent data loss, but still they are quite different.
Everyone loves replication for the ability to synchronize data on another server with the minimum delay, but backup has some undisputed advantages as well. For instance, in case of accidental (or intentional) data deletion, replication will quickly and imperturbably send the changes to the replica while backup (especially with copies on read-only media) is immune to such operations. It takes certain effort to configure both replication and backup correctly and still the possibility of errors exists anyway.
Recommendations: If you have replication configured, do not neglect backup copies, use both.
Recommendation for Firebird: use the HQbird Enterprise distribution package, it includes both backup and replication tools.

Summary

It is not that easy to configure backup for your favourite DBMS so usually database administrators from organizations where they value their data usually use professional backup tools that allow them to take into account the issues mentioned above and prevent the problems.

For Firebird (pardon for advertising) there is a package called HQbird that includes FBDataGuard.
Also, our company provide complete backup and maintenance support for Firebird and other databases, this is a good choice for those who are not aware about all technical details of backups.

And, of course, keep cherishing your admin paranoia, for instance, get up and check your backup copies right now :)

Contacts

Please feel free to ask any questions: support@ib-aid.com.