



Open Source  
Data Warehousing

## How To Backup and Restore in Infobright

“The ability to quote is a serviceable substitute for wit.” *Somerset Maugham*

*Susan Bantin, Infobright, 2009-07-05*

**Infobright, Inc.**  
**47 Colborne Street, Suite 403**  
**Toronto, Ontario M5E 1P8**  
**Canada**  
[www.infobright.com](http://www.infobright.com)  
[www.infobright.org](http://www.infobright.org)

### **Table of Contents**

<b>Synopsis</b>	<b>3</b>
<b>Introduction</b>	<b>3</b>
<b>Methodology</b>	<b>4</b>
<b>Example</b>	<b>7</b>
<b>Summary</b>	<b>8</b>

### **Synopsis**

A regularly scheduled backup procedure is essential in ensuring system reliability. The following document describes how Infobright stores database and Knowledge Grid files and how to successfully backup and restore your data warehouse.

Infobright uses the native file system to store all data files, therefore any backup tool can be used to back up the data. Data files are stored in a compressed format and so backup times are considerably faster than other data warehouse solutions.

During regular data warehouse access (read and write), the tables are locked, and so caution is required to ensure that loads and queries are not occurring during the backup window.

### **Introduction**

Backup of Infobright is straightforward and consists of making a copy of the compressed data warehouse files and associated Knowledge Grid files. Infobright uses the native files system, and as such, does not need a dedicated agent to perform backups. For a full backup, simply backup the directory containing the database.

For an incremental backup, it is sufficient to backup the newly created files and any modified files; new data is added in 2GB files. Ensure that the Knowledge Grid files are always backed up, as changes can occur to the metadata during query operations without the addition of new data. This procedure is also supported by any backup tool.

When restoring your data warehouse, we recommend a full restore of all database files including the Knowledge Grid files.

### **Methodology**

#### Backup Procedure

To back up the Infobright databases, copy the entire directory containing the Infobright databases, including the Knowledge Grid. This is usually the data subdirectory in your Infobright installation directory.

The safest methods of ensuring a complete backup of the database is:

1. Shut down the database before making a copy or,
2. Lock the tables and take a snapshot

You can take advantage of incremental backups, since only some of the database files are updated when new data is imported. Be sure to do a full backup occasionally.

Important: Regarding the Knowledge Grid, some files in the KNFolder are updated when queries (using JOIN) are run so be sure to back up the KNFolder on a regular basis, even when making incremental backups.

#### Restore Procedure

To restore the Infobright databases from a backup copy, do the following:

1. Replace the entire data directory with the backup copy. This is usually the data subdirectory in your Infobright installation directory.
2. Replace the KNFolder with the backup copy (if the KNFolder is not inside the data directory).

Important: Do not manually modify database files or move them from one active database to another — this may lead to data corruption and unpredictable results.

#### Archived Instances

If you want to set up a fully archived instance of your Infobright data warehouse, it is necessary to install Infobright to another location with another instance using different port and file directories. The original data must be transferred by exporting it using `SELECT ... INTO OUTFILE` in binary or text format and then

loading the data into the second instance using LOAD DATA INFILE. Currently, there is not a method of transferring the data without using a decompression and load.

### **Can I restore a single database table?**

When restoring tables, it is important to ensure that the Knowledge Grid is up-to-date, therefore a full restore of the entire instance is recommended, rather than just a single table. For this reason, when doing either full or incremental backups, the Knowledge Grid should always be updated as well.

### **Can I rename the data file folder?**

Infobright tables are globally numbered in order to identify Knowledge Node files. Therefore, while you can rename the entire database by renaming the folder on disk, you should not copy a database folder from one active instance to another, or within the same active instance (e.g. in an effort to make a backup). Copying database folders within one instance may result in different tables with the same globally assigned number, which may lead to errors in query results or an unstable environment. A backup of the whole database folder including the Knowledge Grid is recommended.

Note: the “briighthouse.seq” file is used to store the largest table number used, and is modified when CREATE TABLE is used within the Infobright storage engine. Editing it may allow for copying a database from one active instance to another safely, but it is not recommended.

### **Do the Knowledge Grid files need to be backed up separately?**

Even if there are no data changes, the pack-to-pack nodes within the Knowledge Grid are constantly updated to reflect relationships found during query operations. A Knowledge Grid backup would provide the ability to restore pack-to-pack nodes and is recommended.

Within one instance of Infobright, all Knowledge Grid files are stored together for all databases. It is not currently possible to distinguish specific Knowledge Grid files associated with a specific database. All Knowledge Grid files within the KNFolder should be backed up every time.

### **How does Infobright manage database locks?**

Our locking model follows the standard MySQL model for managing transactions. MySQL does have commands to explicitly lock and unlock tables. It also locks during an update and will automatically unlock the table after a commit (commit may be automated depending on the value of the autocommit variable). If there is a lock against a table the next operation will queue up according to the following priorities:

1. When a WRITE lock is issued. If there are no locks currently on the table, the WRITE lock is granted without queuing. Otherwise, the lock is put into the WRITE lock queue.
2. When a READ lock is issued. If the table has no WRITE locks on it, the READ lock is granted without queuing. Otherwise, the lock request is put into the READ lock queue.

Whenever a lock is released, threads in the WRITE locks queue are given priority over those in the READ queue. Therefore, if a thread is requesting a WRITE lock, it will get the lock with minimal delay.

In the event of frequent table access, for example loads scheduled every 5 minutes, the time to backup may exceed the time available. In this event it is critical that a snapshot of the filesystem be taken instead of simply copying the data files.

Note: Not every type of filesystem supports snapshots. A few that do are: Sun ZFS, and OES-Linux (uses SUSE Linux Enterprise Server). ZFS is available for Linux as well. See also Zmanda "(it can use Snapshots for instant full backups if LVM, ZFS, NetApp or VxFS is being used)."

### **Example**

The fastest way to backup the database is to backup the data directory using typical tools available within your filesystem. The location of the KFolder is available from the brighthouse.ini. By default it is a directory named BH\_RSI\_Repository in your datadir. If it has not been changed specifically, it will read:

```
KNFolder = BH_RSI_Repository
```

To archive the database to another instance of Infobright, it is necessary to export and load the data as follows:

To export a table, use the `select ... into outfile` command. For a quicker recovery, use the binary format:

```
set @bh_dataformat = 'binary';

select * from mytable into outfile '/tmp/mytable.bu' fields
terminated by '\t';
```

To restore, use the `load data infile` command:

```
set @bh_dataformat = 'binary';

load data infile '/tmp/mytable/bu' into table mytable fields
terminated by '\t';
```

Unfortunately 'binary' data format is not available in ICE and you will need to export/import via a text file.

### **Summary**

Backing up an instance of Infobright is as straightforward as taking a snapshot or copy of the data files, including all Knowledge Grid files, and then restoring all files to the original instance. In the case of creating an archived instance, it is necessary to export and re-load the data into a second installation of Infobright.

In all cases, whether doing a full or incremental backup, it is necessary to backup the KNFolder to ensure consistency between the Knowledge Grid and the database files at all times.

And because Infobright uses the native file system to store all data files, any backup tool can be used to back up the data.