

UTF-8 Support Now in ICE!

Infobright Community Edition 3.3.2 Beta
March 10, 2010

About ICE

ICE is an open source software solution designed to deliver a scalable data warehouse optimized for analytic queries. The architecture delivers the following key benefits:

- Ideal for data volumes up to 50TB
- Market-leading data compression (from 10:1 to over 40:1), which drastically reduces I/O (improving query performance) and results in significantly less storage than alternative solutions
- No licensing fees
- Fast response times for ad hoc analytic queries
- Query and load performance remains constant as the size of the database grows
- No requirement for specific schemas, e.g. Star schema
- No requirement for materialized views, complex data partitioning strategies, or indexing
- Simple to implement and manage, requiring little administration
- Reduction in data warehouse capital and operational expenses by reducing the number of servers, the amount of storage needed and their associated maintenance costs, and a significant reduction in administrative costs
- Runs on low cost, off-the-shelf hardware
- Is compatible with major Business Intelligence tools such as Pentaho, JasperSoft, Cognos, Business Objects, and others

ICE is designed to run high-performance ad hoc, complex queries on very large data sets without the need of manual tuning, data partitioning, or index creation.

New Features

Infobright Community Edition 3.3.2 Beta adds support for UTF-8!

- **UTF-8 Support**
 - Ability to load UTF-8 data via the Infobright high-speed loader
 - Querying against all UTF-8 languages (approximately 650) and foreign characters (stored in 1 to 3 bytes) in all text-based operations, both input and SQL statement (i.e., function, constant, min/max, etc.)
 - Ability to convert current ASCII tables to UTF-8

- **Supported for all UTF-8 Collation Types Supported by MySQL:**
 - utf8_bin
 - utf8_czech_ci
 - utf8_danish_ci
 - utf8_esperanto_ci
 - utf8_estonian_ci
 - utf8_general_ci (default)
 - utf8_hungarian_ci
 - utf8_icelandic_ci
 - utf8_latvian_ci
 - utf8_lithuanian_ci
 - utf8_persian_ci
 - utf8_polish_ci
 - utf8_roman_ci
 - utf8_romanian_ci
 - utf8_slovak_ci
 - utf8_slovenian_ci
 - utf8_spanish2_ci
 - utf8_spanish_ci
 - utf8_swedish_ci
 - utf8_turkish_ci
 - utf8_unicode_ci (includes French and German)

Previously, ICE 3.3.1 migrated the version of MySQL that is embedded inside Infobright. Previously, ICE shipped with MySQL 5.1.14, this release updates that to MySQL 5.1.40 GA. The update fixes some previously reported issues with ICE, as well as improves some performance and query syntax support as a result of improvements made to the MySQL engine.

Enabling UTF-8 Support

Infobright 3.3.2 Beta ships with a flag set in the brighthouse.ini file:

```
UseCharset = 0
```

If you are loading new data to a new installation of Infobright and wish to enable UTF-8 support, then before loading data change this flag to:

```
UseCharset = 1
```

Please note: see the User Guide addendum (pages 8-9) for instructions on converting a current table to UTF-8 structure.

Upgrade Issues

Existing users may have charsets and collations defined for tables/columns that do not match the actual Infobright storage of ascii charset and ascii_bin collation. When you enable UTF-8 support (by setting `usecharset = 1` in the brighthouse.ini file), these settings will be respected. To avoid any conflict between the setting and the actual data structures, please run the **Charset Migration Tool** to remove any conflicts. Please see the instructions for the Charset Migration Tool in the User Guide addendum on pages 8-9.

If you are upgrading from ICE 3.3 or earlier, due to the migration to the MySQL 5.1.40 GA version (introduced in 3.3.1), **it is necessary to run a MySQL Updater program after the upgrade** to update the Infobright data files to the new compatible structure.

ICE 3.3.2 Beta Release Notes

The upgrade instructions are included below, and available online at:

Windows: http://www.infobright.org/wiki/Upgrade_Guide_for_Windows/
Linux: http://www.infobright.org/wiki/Upgrade_Guide/

Windows Upgrade Instructions

To upgrade IEE on Windows, please follow the Windows Installation Instructions. The Install Wizard automatically detects a previous version of IEE and upgrades your IEE installation while preserving your data and configuration settings.

After installation, run the MySQL Upgrade utility from the Windows command line:

```
cd "C:\Program Files\Infobright"  
  
.\bin\mysql_upgrade.exe --defaults-file="c:\Program Files\my-  
ib.ini" -uroot --tmpdir=c:\tmp
```

Important: The MySQL Upgrade utility may display several errors regarding the use of locks with log tables and errors requiring table upgrades. The errors are handled automatically by Infobright or the upgrade utility and can be ignored.

Stop and start the Infobright server from the Start Menu items.

Linux Upgrade Instructions

These instructions are for upgrading using the rpm or deb package. To upgrade a tar install, please see:

http://www.infobright.org/wiki/Tar_Upgrade_Guide/

Run either:

```
rpm -U infobright-version-platform.rpm  
dpkg -i infobright-version-platform.deb
```

Start the Infobright server:

```
/etc/init.d/mysqld-ib start
```

Run mysql upgrade tool to upgrade data folder:

```
cd /usr/local/inforbright  
  
./bin/mysql_upgrade --defaults-file=/etc/my-ib.cnf --user=root --  
tmpdir=/tmp
```

Important: The MySQL Upgrade utility may display several errors regarding the use of locks with log tables and errors requiring table upgrades. The errors are all handled automatically by Infobright and/or the upgrade utility and can be ignored.

Re-start the Infobright server:

```
/etc/init.d/mysqld-ib restart
```

ICE 3.3.2 Beta Release Notes

Supported Platforms

ICE 3.3.2 Beta is available on the following core platforms for 64-bit Intel or 64-bit AMD:

- Red Hat Enterprise Linux 5
- CentOS 5.2

ICE is also available on the following platforms for 32-bit Intel or 32-bit AMD: (for solution testing purposes only, not recommended for performance testing or production deployments):

- Windows XP

ICE 3.3.1 is also available on:

- Windows Server 2003/2008 (64-bit)
- Ubuntu 8.04 (32-bit)
- Fedora 9 (32-bit)

System Requirements

For Personal Evaluation and/or Application Development

CPU Speed	32-bit:	1.6GHz minimum 2.0GHz or faster dual or quad core recommended
	64-bit:	1.8GHz minimum 2.0GHz or faster dual or quad core recommended
Memory	32-bit:	1GB minimum 2GB or more recommended
	64-bit:	2GB minimum 4GB or more recommended

Production Deployment

CPU Speed	64-bit:	2.0GHz minimum 2.0GHz or faster dual or quad core recommended
Memory	64-bit:	4GB minimum 16GB or more recommended (and at least 2GB per core)

ICE 3.3.2 Beta Release Notes



Download ICE

<http://www.infobright.org/Downloads/ICE/>

Installation Instructions

Windows: http://www.infobright.org/wiki/Install_Guide_for_Windows/
Linux: http://www.infobright.org/wiki/Install_Guide_for_Linux/
MacOS (VM): http://www.infobright.org/wiki/Install_Guide_for_Mac_OS/

Upgrade Instructions

Windows: http://www.infobright.org/wiki/Upgrade_Guide_for_Windows/
Linux: http://www.infobright.org/wiki/Upgrade_Guide/

Issues Fixed

The following issues have been fixed and verified since the release of ICE 3.3.1. For more details on these, please visit <http://bugs.infobright.org/>

Ticket Description

- #1010 Server crashes if, for some reason, one of the DPN files doesn't exist or not accessible
- #1022 Install shell scripts should detect 32/64 bit of OS, and give nice message
- #1272 ICE 3.1 for windows 32, startup error "mysqld-nt: unknown option '--skip-federated'"
- #1309 Compare date / datetime / timestamp column to value '0000-00-00' / '0000-00-00 00:00:00' return error.
- #1705 Broken results when enclosing known (but unimplemented) mysql functions in other functions
- #1708 select with range test on negative double does not return expected result
- #1809 CHAR and VARCHAR sizes aren't handled correctly

Known Issues

For the current list of known and closed issues, please visit <http://bugs.infobright.org/>

Ticket Description

- #848 Knowledge Node files intermittently become unavailable to queries, resulting in slower queries
- #870 Infobright can't find numbers in numeric text; inconsistent with MyISAM's treatment
- #875 Infobright engine appears as transactional - cosmetic only; no impact on performance
- #888 installer hardening: quotes in path (and spaces) not handled by installer
- #891 Default MySQL tests included in Infobright contain tests that will not run
- #916 Use of constants in select causes switch to MySQL Path (and slow performance)



ICE 3.3.2 Beta Release Notes

- #918 IB Loader wrongly calculate size of the file whether there is an escape character present in the data.
- #933 Results of queries executed by IB engine aren't cached
- #935 Usability: "Cannot start bhloader" needs improved error message
- #960 escaped N should be loaded as NULL
- #980 ICE 32: Illegal mix of collations error on select
- #987 Cannot support Debian Etch because fdopendir is available in glibc from version 2.4
- #997 Embedded enclosure character aborts load without error or warning
- #1019 mysqld crashes when killing mysql client during load
- #1026 ICE Src: ability to set build comment from commandline
- #1030 ICE: make 'txt_variable' as the default dataformat for export
- #1031 ICE: delete with condition doesn't report error
- #1039 Untarring the 3.0.2 tarball will overwrite contents in previous 3.0 install - it does not create new 3.0.2 dir
- #1042 Queries with join on Infobright and myisam table do not return warning that is executed by mysql.
- #1072 "LINES TERMINATED BY" ignored on export
- #1073 FF: Query using views with a condition in the WHERE clause using function/complex expression on column from views switches to mysql engine.
- #1115 text load skips last line if no newline present
- #1141 ICE: compile warning
- #1161 ICE: Misleading error message about termination of session
- #1172 Dropping a table while it is currently loading hangs rather than returning an error
- #1174 LOCK TABLES does not work as expected.
- #1177 rename source tar to not have hardware (i686 or x86_64 or win32)
- #1179 Permission issue in Infobright binary causes the installation to fail for /home or /root folder
- #1188 load of a lookup column slows a lot for larger number of distinct values
- #1235 ICE32 VMWare: export file gets newline character M at the end of each line
- #1275 TPC-H Q22 wrong results
- #1282 A variable declaration as a result of an aggregation may switch to mysql query path
- #1310 Wrong conversion value '0000-00-00 00:00:00' in timestamp datatype column.
- #1315 "CHARACTER SET" not recognized in LOAD DATA INFILE
- #1338 Loader - enhancement to allow for rejected records option
- #1344 Problems related to character_set_server
- #1418 rpm install: force installation or upgrade doesn't set the owner on data directory
- #1452 dump/load is inconsistent
- #1453 inconsistent dump/load, continued
- #1454 corrupt dump/load when ordering data
- #1469 SHOW xxx returns BINARY
- #1470 --debug-info broken in mysql-ib
- #1479 security issue in InfoBright? engine
- #1484 LOAD DATA INFILE ignores "LINES TERMINATED BY"

ICE 3.3.2 Beta Release Notes

- #1601 DATE_FORMAT returns a binary result
- #1627 No ability to change name/location of my-ib.cnf with RPM/DEB
- #1649 bugs in build system
- #1657 non-database folders and files in data directory
- #1662 "Numeric string" cannot be inserted into numeric columns, inconsistent with standard MySQL treatment
- #1718 importing escaped backslashes fails
- #1720 Add brighthouse.ini variable to make bh_dataformat="txt_variable" default
- #1730 mysqld crashes when process killed
- #1737 query keep generating duplicate bh.err message for 10+ hours
- #1740 Talend infobright connector problem
- #1758 Bug when connecting remotely to IB on 64bits
- #1766 STDDEV wrong result
- #1770 crash when with select and subquery and <> any
- #1774 joining large tables without WHERE conditions very slow on rough level
- #1780 Hash-based authentication
- #1785 bhloader fails and warns about wrong data/column on a unexisting row
- #1786 select format(count(*),0) from ... crash server.
- #1789 rpm install with prefix parameter doesn't work
- #1801 ICE importer ignores IGNORE LINES
- #1804 RENAME TABLE causes the table to be inaccessible

ICE Resources

- Community Website <http://www.infobright.org/>
- Community Wiki <http://www.infobright.org/wiki>
- Community Forum <http://www.infobright.org/Forums>
- FAQ <http://www.infobright.org/Resources/FAQ/>
- Bugs/Requests <http://bugs.infobright.org/>

ICE 3.3.2 Beta Release Notes

Addendum to ICE 3.3.1 User Guide

Existing users may have charsets and collations defined for tables/columns that do not match the actual Infobright storage of ascii charset and ascii_bin collation. When you enable UTF-8 support (by setting `UseCharset = 1` in the `brighthouse.ini` file), these settings will be respected.

Previously, any odd, extra or garbage data in columns would have been treated as binary or ignored. In 3.3.2 Beta (with `UseCharset = 1` flag) this data will be treated as UTF-8. To avoid any conflict between the setting and the actual data structures, please run the **Charset Migration Tool** to remove any conflicts.

Charset Migration Tool

Infobright includes a standalone application to adapt existing tables created prior to 3.3.1 to UTF-8 capable structures. The Charset Migration Tool (**CHMT**) is in the Infobright bin directory.

IMPORTANT: You should backup your data before running the CHMT tool.

CHMT requires a text file containing a mapping between collations used for conversion.

```
chmt --help // help message
```

Executing CHMT:

```
chmt --datadir=/absolute/path/to/data/directory [other parameters]
```

List of parameters:

Parameter	Type	Description	Remarks
<code>datadir</code>	mandatory	absolute path to data directory	
<code>conv-map</code>	optional	absolute path to file with collations conversions;	if not specified CHMT would try to use file: <code>chmt-binary-folder/./support-files/collations.txt</code> ; if not found there it would search for: <code>chmt-binary-folder/collations.txt</code>
<code>database</code>	optional	name of database for migrating	if specified, tables from no other databases would be migrated
<code>table</code>	optional	name of table for testing	if specified, database must be also specified; no other tables but specified will be migrated
<code>log-file</code>	optional	absolute path to output log file	if not specified, logs will be printed to the console

Log structure

The logs are similar to those from the Updater. There is information about every considered table founded in specified datadir. Each conversion finishes with [NOT NEEDED], [PASS] or [FAILED] status.

Collations-conversion-file structure

Each conversion directive is stored in one line of file:

```
collation_from_name;collation_from_id;collation_to_name;collation_to_id
```

For example:

```
big5_chinese_ci;1;binary;63
```

where both fields containing names are only informative (all conversions will be done using only ids).

Note: If CHMT will not recognize any of existing collations (means this collation wasn't in collation-conversion-file), it should stop migration process of related table.

Example collations-conversion-file (with conversion directives described above) can be obtained by running sql:

```
use information_schema;
select a.collation_name a_n, a.id a_id, b.collation_name b_n,
b.id b_id
from information_schema.collations a,
information_schema.collations b, character_sets c
where substr(a.collation_name, 1, locate('_',a.collation_name)-
1)=c.character_set_name and
substr(a.collation_name, 1, locate('_',a.collation_name)) =
substr(b.collation_name, 1, locate('_',b.collation_name))
and b.collation_name like '%bin' and c.maxlen=1
UNION
select a.collation_name a_n, a.id a_id, 'binary' b_n, 63 b_id
from information_schema.collations a, character_sets c
where ((substr(a.collation_name, 1, locate('_',a.collation_name)-
1)=c.character_set_name) or (locate('_',a.collation_name)=0)) and
(c.maxlen>1 or c.character_set_name ='binary')
order by a_id
into outfile '/some/path/my_collations.txt' fields terminated by
';';
```