Case Sensitivity in SQL™ Server Geodatabases

An ESRI ® Technical Paper • May 2007

Case Sensitivity in SQL Server Geodatabases

An ESRI Technical Paper

Contents

Case Sensitivity in SQL Server Geodatabases

Beginning with ArcGIS® 9.2, SQL Server geodatabase identifiers, such as feature class or table names, cannot be case sensitive. This means the contents of the ArcSDE® and geodatabase repository tables (SDE_ and GDB_ tables) cannot be created and stored using a case-sensitive collation. Your database collation must be case insensitive.

This does not mean you cannot perform case-sensitive searches against attribute data. Attribute data can be stored using a case-sensitive collation, and queries can be sensitive to the case of individual characters.

This document has been written to assist users who need to convert existing case-sensitive geodatabases to use case-insensitive collations and enable case-sensitive collations on individual character data columns.

This document consists of five sections. The first two sections, Upgrading to ArcSDE 9.2 and Creating a New Enterprise ArcSDE 9.2 Geodatabase, are part of the Quick-Start Guide. They will walk you through the required steps to either upgrade to or create a new ArcSDE 9.2 geodatabase that supports case-sensitive data. The subsequent three sections, Understanding SQL Server Collations, Understanding Case Sensitivity in SQL Server, and Understanding Case Sensitivity in the ArcSDE Geodatabase, provide a more detailed discussion of these topics.

Quick-Start Guide

Upgrading to ArcSDE 9.2

If you already have an understanding of SQL Server data storage, case sensitivity, and collations, you will not need to digest the contents of this document in detail. Perhaps you just want a limited set of steps that will guide you through the conversion of your geodatabase. Perform the following steps if you are upgrading to ArcSDE 9.2 from an 8.3, 9.0, or 9.1 geodatabase that uses a case-sensitive database collation.

Note: Geodatabase names cannot be case sensitive. This means all SQL Server geodatabases must reside on a SQL Server instance with a *case-insensitive* server collation. For information on setting or altering the SQL Server default server collation, see Setting and Changing the Server Collation in the SQL Server Books Online.

Begin by disconnecting all users and stopping the ArcSDE service. Perform a full backup of all databases to be upgraded and uninstall the ArcSDE software. Next, alter the SQL Server database collation of all the databases you are upgrading.

Note: If your geodatabase uses the *multiple spatial database* architecture (a database named 'sde' and one or more related databases) you must alter the collation of each database as described in Step 2 below. Steps 1 and 3, dropping and recreating the SDE_spatial_references table constraints, are only performed in the sde database.

Alter the SQL Server Database Collation

1. The SDE_spatial_references table contains two constraints that must be dropped before the database collation can be changed. The following Transact-SQL statements will drop the spatial_ref_xyunits and spatial_ref_zunits constraints. All code in this document references a database named sde. Modify the database name to match that of your geodatabase.

```
USE [sde]
GO
ALTER TABLE [sde].[SDE_spatial_references]
    DROP CONSTRAINT [spatial_ref_xyunits];
ALTER TABLE [sde].[SDE_spatial_references]
    DROP CONSTRAINT [spatial_ref_zunits];
```

2. After the constraints are dropped, change the database collation using the ALTER DATABASE statement.

```
ALTER DATABASE [sde] COLLATE
    SQL_Latin1_General_CP1_CI_AS;
```

The collation used in the above statement is the default for most English-language databases. The best way to choose a collation is to see what your SQL Server instance collation is set to (available on the Server Properties dialog box), making sure that it is case insensitive. All case-insensitive collations use the letters *CI* in their name. Execute the statement

```
SELECT * FROM fn_helpcollations
```

to get a list of all the supported SQL and Windows® collations.

For more information on collations in SQL Server, refer to the Understanding SQL Server Collations section.

3. After the database collation has been successfully altered, re-create the constraints on the SDE_spatial_references table.

```
USE [sde]
GO
ALTER TABLE [sde].[SDE_spatial_references] WITH CHECK
    ADD CONSTRAINT [spatial_ref_xyunits]
    CHECK (([xyunits]>=(1)))
GO
ALTER TABLE [sde].[SDE_spatial_references]
    CHECK CONSTRAINT [spatial_ref_xyunits]

ALTER TABLE [sde].[SDE_spatial_references]
    WITH CHECK ADD CONSTRAINT [spatial_ref_zunits]
    CHECK (([zunits]>=(1)))
GO
ALTER TABLE [sde].[SDE_spatial_references]
    CHECK CONSTRAINT [spatial_ref_zunits]
```

Install the 9.2 Case Sensitivity Patch

Ensure your ArcSDE component installation is version 9.2 service pack (SP) 2 before installing the Case Sensitivity Patch. This patch installs functionality to set a case-sensitive collation for all subsequently created character data while preserving the case insensitivity of geodatabase identifiers and metadata.

Note: It is extremely important that the upgrade of ArcSDE repository tables to 9.2 only be performed *after* the Case-Sensitivity Patch has been applied. Your ArcSDE repository tables must be upgraded directly from 8.3, 9.0 or 9.1 to 9.2 SP2 with the Case-Sensitivity Patch. Do not run the ArcSDE Post-Installation Wizard or the sdesetup command until the Case-Sensitivity patch has been installed and the collation of your database has been modified.

.

Upgrade the Geodatabase

The geodatabase can be upgraded using either the sdesetup –o upgrade command or the Setup Repository option using the Custom option of the ArcSDE Post Installation Wizard.

The upgrade will perform two actions. First, the collation of any SDE_ and GDB_ table columns containing character data will be changed to match the new case-insensitive database collation you just established. Next, a new DEFAULTS keyword parameter, COLLATION_NAME, will be added to the SDE_dbtune table.

After the upgrade is complete, you will be able to use your geodatabase. Any existing datasets that were created prior to the application of the patch and upgrade will use the old, case-sensitive database collation. By default, new datasets will use the new, case-insensitive database collation. If you want new datasets to use a case-sensitive collation, complete the following step.

Set the COLLATION_NAME Parameter

The COLLATION_NAME parameter can be used to set the collation of character data columns of all user-defined datasets. By default, all character data columns use the default database collation. By configuring this parameter, all new datasets will be created using the collation specified for the COLLATION_NAME parameter rather than the database collation.

Choose the case-sensitive version of your database collation. This usually means altering the *CI* in the collation name to *CS*. If you are unsure, check the SQL Server Books Online or execute the following query for a list of collation names:

```
SELECT * FROM ::fn_helpcollations()
```

To set the COLLATION_NAME parameter, use the sdedbtune command. For example, if your database collation is SQL_Latin1_General_CP1_CI_AS, execute the following command to set COLLATION_NAME to the case-sensitive version:

```
sdedbtune –o alter –k DEFAULTS –P COLLATION_NAME
  –v SQL_Latin1_General_CP1_CS_AS –i esri_sde –D sde
```

When a new dataset (table, feature class, etc.) is created, the collation specified in COLLATION_NAME is applied to every character data column. If COLLATION_NAME is blank, the database collation is used. If the UNICODE_STRING parameter in the SDE_dbtune table is set to FALSE, the COLLATION_NAME parameter is ignored.

For more information on Unicode data in the geodatabase, see the ArcGIS Desktop Help and search for "Unicode".

For more information on the SDE_dbtune table and sdedbtune command, see the ArcGIS Desktop Help and search for "dbtune".

Creating a New Enterprise ArcSDE 9.2 Geodatabase

The following steps will guide you through creating a new case-insensitive geodatabase and configuring it to support the creation of case-sensitive attribute data.

Install the 9.2 Case Sensitivity Patch

Ensure that your ArcSDE component installation is version 9.2 SP2 before installing the Case Sensitivity Patch. This patch installs functionality to set a case-sensitive collation for all subsequently created character data while preserving the case insensitivity of geodatabase identifiers and metadata.

Set the COLLATION_NAME Parameter

The COLLATION_NAME parameter is part of the SDE_dbtune table's DEFAULTS keyword. It is used to set the collation of character data columns of all new, user-defined datasets. By default, all character data columns use the default database collation. By configuring this parameter, all new datasets will be created using this collation rather than the database collation.

Choose the case-sensitive version of your database collation. This usually means altering the *CI* in the collation name to be *CS*. If you are unsure, check the SQL Server Books Online or execute this query for a list of collation names:

```
SELECT * FROM ::fn_helpcollations()
```

To set the COLLATION_NAME parameter, use the sdedbtune command. For example, if your database collation is SQL_Latin1_General_CP1_CI_AS, execute the following command to set COLLATION_NAME to the case-sensitive version:

```
sdedbtune -o alter -k DEFAULTS -P COLLATION_NAME
  –v SQL_Latin1_General_CP1_CS_AS –i esri_sde –D sde
```

When a new dataset (table, feature class, etc.) is created, the collation specified in COLLATION_NAME is applied to every character data column. If COLLATION_NAME is blank, the database collation is used. If the UNICODE_STRING parameter of the SDE_dbtune table is set to FALSE, the COLLATION_NAME parameter is ignored.

For more information on Unicode data in the geodatabase, see the ArcGIS Desktop Help and search for "Unicode".

For more information on the SDE_dbtune table and sdedbtune command, see the ArcGIS Desktop Help and search for "dbtune".

Understanding SQL Server Collations

The rules governing the storage, sorting, and comparison of character data are determined by a collation. These rules are usually chosen to support your language or locale.

Collations determine the code page that is used for single-byte character data (non-Unicode data) as well as its sort order, which determines the order in which data will be sorted and compared. The two basic types of sort order are dictionary and binary. ArcSDE supports dictionary sort orders, which are the most common. A component of a sort order is whether string comparisons are sensitive to things like the case or accenting of each character.

By default, the collation used for your instance of SQL Server, and usually your databases and the data contained within them, specifies a dictionary sort order and case-insensitive string comparisons. This means that an ORDER BY clause will return character data in alphabetical order, and the case of individual characters is disregarded when they are compared. For example, lowercase *a* is considered the same as uppercase *A* in a case-insensitive comparison.

SQL Server has two types of collations: Windows collations, which are designed to match Windows locales, and SQL collations, which match combinations of code page and sort order from earlier versions of SQL Server. If you do not accept the default collation, choose a Windows collation because SQL collations are provided only for backward compatibility.

Collations can be set at three different levels within SQL Server: at the instance or server level, the database level, and on individual columns. The server collation is chosen on installation of SQL Server. It controls the collation of server-level identifiers such as login names and database names. The server collation defaults to a Windows or SQL collation that matches the Windows locale. Unless otherwise specified on creation, databases default to the collation set for the server. The database collation controls the collation of database-level identifiers such as table and view names. Individual columns that store character data (char, varchar, nchar, nvarchar, text, and ntext data types) have a collation. Unless the collation is specified when columns are created, the default database collation is used. The column collation controls how data stored within a column is ordered and compared.

Understanding Case-Sensitivity in SQL Server

How SQL Server treats the case of letters when doing string comparisons is part of the sort order defined by the collation.

If data is stored using a case-sensitive collation, SQL Server will treat different cases of the same character as different characters. For example, if the collation of the owner_name column is case sensitive, the query SELECT * FROM parcels WHERE owner_name = 'SMITH' will not return values of 'Smith', 'smith', or 'SmItH' because the case of the characters does not match those in the WHERE clause.

Depending on the level of the collation, case sensitivity can also apply to identifiers such as the name of the table and column specified in the query. The database collation affects all the character data stored within the database system tables. This means system table columns storing the names of objects within the database, such as tables and views, will use case-sensitive collations. If the table is named parcels and your query specifies 'PARCELS' in the FROM clause, you will receive an error message indicating that the object 'PARCELS' does not exist.

A case-sensitive server collation will affect all server-level identifiers such as database names. A case-sensitive database level will affect all database-level identifiers. A case-sensitive column collation will affect only the data stored within that column.

Typically, the need for case sensitivity within SQL Server applies only to columns storing character data. The ability to perform case-sensitive string comparisons is a common requirement. The need for case-sensitive identifiers or object names is much rarer.

Understanding Case Sensitivity in the ArcSDE Geodatabase

ArcSDE geodatabases can support case-sensitive column collations on user-defined datasets. This means your character data can be stored using a case-sensitive collation and that all queries against that data will use case-sensitive string comparisons.

The geodatabase does not support case-sensitive identifiers such as object class or feature class names. This means both the server and database collation must be case insensitive as well as the column collations for all the SDE_ and GDB_ repository tables.

When you create a new dataset, the case of the table and column names is preserved within the geodatabase repository tables. For example, if you create mixed case column names, they will be preserved both within SQL Server and within the SDE_column_registry table where they are referenced and queried by applications such as ArcCatalog™. When you view the attribute data for a feature class, the column names will appear just as they were entered when the feature class was created. However, these same applications must have the ability to query for values, such as column and table names, using case-insensitive string comparison. Therefore, the geodatabase repository tables (SDE_ and GDB_ tables) must be stored using a case-insensitive collation, and the database that houses the geodatabase must also use a case-insensitive collation.

If you require case-sensitive string comparisons and searches of all your attribute data, you can specify that all the columns that store character data be created with a case-sensitive version of the current database collation. This is done by adding the name of the case-sensitive collation to the COLLATION_NAME parameter of the SDE_dbtune table's DEFAULTS keyword. By default, the COLLATION_NAME string is blank and all columns are created with the database collation. This is only supported with Unicode character data, so the UNICODE_STRING parameter must be set to TRUE. As long as the UNICODE_STRING parameter is TRUE (the default), all columns that store character data will be created with a Unicode data type (normally nvarchar). Once the COLLATION_NAME string has been set, that collation name will be applied to the creation of all new character data columns.

If you only require case-sensitive string comparisons for a handful of attribute columns, you can alter these manually within SQL Server. Using either Management Studio or Enterprise Manager, navigate to the table you want to modify, right-click it, and choose Design. On the Table Design dialog box, select the column you want to modify and type the collation name in the Collation field, or click the ellipsis (…) to choose from a list of all supported collations. This will alter the collation of a single column. If the table with the altered column is subsequently registered as versioned, the corresponding column on the Adds table will have the same collation. If the column you want to modify is on a table that has already been registered as versioned, you will need to manually change the collation on both the business table and the related Adds table.

> **Note:** The only collations that are supported for attribute columns within a geodatabase are the default case-insensitive database collation or the case-sensitive version of the default database collation. Do not set column collations to a different language from the database collation. If you need to store multiple languages or alphabets within a single geodatabase, use Unicode data types. Unicode is the default for all new data created with ArcGIS 9.2. For more information on Unicode data in ArcGIS, see the ArcGIS Desktop Help and search for "Unicode".