Chapter 3 — *MPRESS, the db-One Authoring System*

This chapter gives you an overview of MPRESS, the db-One Authoring System, and explains how to use MPRESS to create a database in db-One format. In addition, it includes the following topics:

- a detailed description of MPRESS Control Format (MCF) files and how they are used by MPRESS

- how to use MPRESS with an RDBMS

- how to use the features of MPRESS for Windows

- how to use the features of Command Line MPRESS

- how to update an existing db-One database

This chapter also includes two quick-start tutorials (one for MPRESS for Windows and one for Command Line MPRESS) to help you get up to speed quickly. The tutorials demonstrate MPRESS' features in the order in which you'll typically use them. In addition to the tutorials, procedures for using MPRESS are included in this chapter for your reference.

How to use this chapter

For an overview of MPRESS and for details about the files MPRESS uses and creates, continue reading this chapter, up to "What is MPRESS for Windows."

To get started using MPRESS right away, go to the MPRESS for Windows quick-start tutorial on page 32, or go to the Command Line MPRESS quick-start tutorial on page 55. You might read the introductory sections of this chapter later for details about MCF files and how to fine-tune properties to get the most out of MPRESS.

To skip the tutorials, go directly to "Using MPRESS for Windows" on page 41, or to "Using Command Line MPRESS" on page 59 for MPRESS procedures.

What is MPRESS?

The db-One Authoring System (MPRESS) is part of the db-One suite of products, designed to perform high-speed, accurate search and retrieval of information from very large text databases.

MPRESS compresses large text databases into **Metro One Data Format (MDF) files**, which are used by the other db-One products. You'll use MPRESS to author a database in MDF format, which is recognized by the db-One Search Engine. An MDF file consists of data and search indexes, and is tightly compressed so that search times are minimal. If you're using an RDBMS to manage your data, MPRESS can build an MDF file that contains indexes only.

Updating MDF files

Once you've created an MDF file, you can easily keep it up-to-date without creating it from scratch again. You can even search against the existing MDF file and the data to update it, *before* you've updated the MDF file.

To update an MDF file, you'll create compressed files called **Metro One Update Data (MUD) files** and **Number Mapper Files (NMF).** These files act like transaction files to update an MDF file (which you might think of as a master file).

The MUD and NMF files contain the records to be added and identify records to be deleted. They also provide information about how to merge the new records with the existing records.

After creating the MUD and NMF files, you'll merge the new records with the existing records and delete those records that are marked for deletion.

Two versions of MPRESS

There are two MPRESS applications: **MPRESS for Windows** (wmpress.exe) and **Command Line MPRESS** (mpress.exe). The Windows version has the advantage of letting you graphically configure, build, and test your database. However, with Command Line MPRESS you can set up a DOS batch (.bat) file to run a set of MPRESS commands.

You can use whichever version you prefer. Both versions use the same set of input files and given identical input, produce identical output files. If you want, you can switch between them, using the Windows version for some tasks, and the Command Line version for others.

This chapter covers procedures for using both versions of MPRESS.

MPRESS input and output files

MPRESS needs specific input files and creates specific output files.

The following table shows which input files are needed and which output files are produced as a result of running MPRESS.

Process	Input files	Output files
creating an MDF file	MTF file and an MCF file	compressed MDF file, mpress.log, error.log (created only if there are errors)
creating MUD and NMF files	Update Data file, an MCF file, and the MDF file to be updated	up to three compressed files–one MUD file (if any records are being added) and one to two NMF files, mpress.log, error.log (created only if there are errors)
merging the MUD and NMF files with the MDF file	all MUD and NMF files created by MPRESS for the MDF file, an MCF file, and the MDF file to be updated	updated MDF file, mpress.log, error.log (created only if there are errors)

> **Note:** When an MDF file is created, if it exceeds 2 GB (the operating system's file size limit), it is split into multiple parts.

What is an MTF file?

A **Metro One Text Format (MTF) file** is any character-delimited text file that contains the data to be indexed and compressed into an MDF file. Each line in the file is a record, ending with a carriage return. Each record contains one or more fields separated by a unique character. An MTF file can be created by any text editor, such as Microsoft Notepad.

> **Note:** Alternately, input data can be retrieved from an industry-standard database that allows access through an ODBC driver. (ODBC drivers are not supplied with MPRESS and must be purchased from a third-party vendor.)

What is an Update Data file?

An **Update Data file** is a text file that contains the records to be added to an existing MDF file and identifies the records to be deleted from an existing MDF file.

What is an MCF file?

An **MPRESS Control Format (MCF)** file is a text file that contains the input parameters, or properties, that dictate how the data is compressed as well as what functionality is available in the resulting MDF file. (MCF files are described in greater detail in the next section.)

By changing properties in the MCF file before running MPRESS, you can do all of the following tasks:

- Create a compressed MDF file.

- Create compressed MUD and NMF files.

- Merge the information in the MUD and NMF files with the existing MDF file.

Log files

The mpress.log file is a text file detailing information about the database and compression results. The error.log file lists any errors that occurred during compression. It is created only if there are errors.

The log files are written to the directory that contains the MDF file (the directory you specify for the OutputPath property in the MCF file). For more information about the OutputPath property, see "Mandatory properties" on page 23 of this chapter.

MPRESS Control Format files

An MPRESS Control Format (MCF) file contains all of the information needed by MPRESS to index and retrieve data; such as the data's format, how the data should be indexed, and which words within a field should be indexed.

If you will be authoring only with MPRESS for Windows, you don't need to know how to structure an MCF file. Therefore, you can to skip over the following section.

MCF file structure

MCF files are divided into **sections**. Each section contains parameters— called **properties**—that you'll assign values to in order to dictate how db-One works with the data. Most of the properties have defaults, but occasionally you might need to change them to optimize compression for your requirements.

> **Important:** MCF files are not case-sensitive. Section and property names can be upper or lower case or any combination of the two. For readability, this guide shows section and property names as a combination of upper and lower case.

An MCF file contains the following sections:

- Database Info

- Field1 through Field*n,* where *n* is the number of fields. (There may be up to 255 fields.)

- Compression Parameters

- Parsing Parameters

- Correction Parameters

- ODBC Parameters

- Update

The sections can be in any order within an MCF file and the properties within a section can also be in any order.

MCF files are formatted like Windows ".ini" files — with each property listed on its own line. Each section in the MCF file starts with a line containing only the section's name in square brackets as shown in the following example:

```
[Database Info]
property_name = value
property_name = value

[Field1]
property_name = value
property_name = value
```

Mandatory sections

The following table shows which sections are mandatory for each different process.

To …	You need these sections …
create an MDF file	Database Info, Fields, and ODBC Parameters (for databases accessed over ODBC Interfaces only).
create MUD and NMF files	Database Info, Fields, and Update. (Although other sections are not necessary, it's okay if they are present.)
merge the MUD and NMF files with an MDF file	Database Info and Update. (Although other sections are not necessary, it's okay if they are present.)

Mandatory properties

These properties are mandatory for creating a new database.

In the Database Info section:

- `MTFPathName` — the path to the MTF file containing the data (not needed for an ODBC-accessed input database).

- `OutputPath` — the path for all output files.

> **Note:** The OutputPath property is mandatory for Command Line MPRESS only. If you use MPRESS for Windows, MPRESS puts all output files in the directory that you specify for the MDFFileName property.

In each Field section:

- `FieldID` — a unique string that identifies the field.

In at least one Field section:

- `CompressField` — an indicator that the field should be replicated in the MDF file.