

Mass Loading XML Data

Tamino offers an XML Mass Load utility that allows the loading of large amounts of documents into Tamino in an efficient manner. This chapter covers the following topics:

- How Does Fast Mass Load Work?
 - Starting the Mass Loader
 - Error Messages
 - Input File Format
-

How Does Fast Mass Load Work?

Rather than loading XML documents online, the Mass Loader prepares the documents to be loaded offline and loads the data efficiently. The following considerations and restrictions should be taken into account when using the XML Mass Loader:

- The Mass Loader needs an active Tamino server with the appropriate collection and doctype defined for it to load documents into. The collection and doctype need to be defined *before* the mass loading can be started.
- The Mass Loader may run remote from a Tamino server, but it requires that its input file is local to the Mass Loader.
- All XML documents must be stored within one input file.
- There must be sufficient space on the Tamino server's temporary working space.
- On error, the XML doctype will be left in the same state as before the mass load was executed.
- The Mass Loader exclusively locks the complete collection in which the doctype resides. This means that queries and other Tamino operations that go against this collection will be refused.
- Mass loading is only possible in doctypes created with the current Tamino version.
- Mass loading into doctypes of the type "nonXML" is not supported.
- Mass loading into doctypes that have parts mapped to server extensions or X-Node is not supported. These sub-trees must be set to "Ignore update". Otherwise, the documents that use these features will be rejected.
- Mass loading bypasses the web server which means that web server security is also bypassed.

Normally, Tamino's mass load processing will tolerate the following documents:

- Documents that are not well formed.

- Documents in which closed content is turned on.
- Documents that are not part of the doctype.
- Documents in which parsing and schema level checks fail.

In these cases, the document is ignored by default, and processing is then continued. An error message is returned, stating that the document was rejected.

Starting the Mass Loader

The XML Mass Loader is called via the command line with the command `inoxmld`. The following options are available:

Parameter	Description	Default Value
Server=<machine name and port>	The machine name where the Tamino Server is running. The format for the machine name and port is: <machinename>:<port>	None
User=<UserID>	Tamino User Identification	(anonymous)
Collection=<collection>/<doctype>	Specification of collection name and doctype to mass load into	None
Input=<input filename>	Input file containing the mass load documents	Standard input
Norejects	Do not tolerate rejects	Tolerate having rejected documents
Log=<error log filename>	Filename for error message to be written to	Standard error
Rejects=<rejects filename>	Filename where each document that is rejected is written to	None

The parameters `Server` and `Collection` are mandatory, as well as an input filename. Please note that the `Server` parameter does not specify the Tamino server name, but that of the machine where the Tamino server is running on. If no input file is provided via the input parameter, then the input is assumed to come from Standard Input. The parameter `norejects` should only be used if you do not want to tolerate rejected documents while mass loading. If it is used, then when an error occurs, the complete mass load is stopped with no documents being loaded, and the doctype is left in the same state as before the mass load operation started.

Note:

Parameter names are case insensitive, but most of the parameter values themselves are case sensitive.

 **Here is an example for starting the Mass Loader via the command line:**

1. At the prompt, enter the command in the following format:

```
inoxmld server=mypc:3204 user=[userid] collection=hospital/patient
input=abc.xml norejects log=xyz.xml
```

The input document is processed. The data is loaded into Tamino.

Note:

The Server parameter does not specify the Tamino server name, but that of the machine where the Tamino server is running.

Error Messages

When an error occurs in the processing of a document, the line and column numbers given in the error report will be relative to the mass load document that was in error (line and column computation start at the ino:object tag).

Input File Format

The input file must have the following format:

```
<?xml version="1.0" encoding="iso-8855-1"?>
<!-- XML entities for all documents may be specified here -->
<ino:request>
<ino:object [ino:docname="<docname>"]>
<!-- document contents for 1 XML document including PI's and Comments -->
</ino:object >
<ino:object [ino:docname="<docname>"]>
<!-- document contents for 2 XML document including PI's and Comments -->
</ino:object >
.....
</ino:request>
```

The XML prolog may be specified only once at the start of the input file. The prolog and entities will be used by the mass loader, but will never be stored by Tamino. If there is no XML prolog, then the file must have a character encoding of UTF-8, UCS-2, or UCS-4. The docname attribute in line three and six is optional and may be moved to the root node.