# PPM Center Database DMS Solution

for HP Project and Portfolio Management Center version 9.13 and later

# About this Document

This paper provides HP PPM Center administrators some pointers on how to use the PPM Center Database DMS solution that was introduced in HP PPM Center version 9.13.

The following sections are covered:

- Technical Overview
- DMS Migration
- Database Sizing
- Oracle TEXT Indexes Creation and Maintenance (when using full-text search)

Note:
This document applies to PPM Center version 9.13 and later.

# Technical Overview

The new PPM Center Document Management System (DMS) solution — PPM Center Database DMS — allows you to store all the documents and associated metadata in your HP PPM Center database.

Contents of the documents (in whatever document format) are stored in BLOB fields. The PPM Center Database DMS solution leverages Oracle TEXT to provide full-text search, and the related indexes are CONTEXT indexes, both for metadata and document contents.

Note:

As of PPM Center version 9.14, it is mandatory to save the documents in the database schema hosting HP PPM Center business data. However, in the next release of PPM Center, it should be possible to save documents and full-text search indexes in a different Oracle DB Schema.

Note:

By default, document contents are stored in the default PPM tablespace USER_CLOB, while the metadata and the TEXT indexes (when created) are stored in the default PPM tablespace USER_DATA. DBAs can change the tablespaces used by PPM Center Database DMS. For example, to save DB disk space by compressing documents contents, or by tuning tablespace-level settings that would provide better performance for their specific usage. Such tablespace customizations are supported, but out of the scope of this document.

## Data Model

The documents information and contents are stored in the following four tables:

- **KNTA_DOCUMENTS**: This table is used by all three PPM Center DMS solutions (File System, Documentum Integration, and PPM Center Database DMS). It contains the documents information that is displayed in the PPM references overview, as well as the current "check out" status of the document.
- **KNTA_DOCUMENT_VERSIONS**: Contains the metadata specific to a document version:
  - Filename
  - File size
  - Version comment
  - Check in date and user
- **KNTA_DOCUMENT_TIP_CONTENTS**: Contains the binary contents of the latest version of a document, plus a copy of all the metadata of document and the latest version for full-text indexing.
- **KNTA_DOCUMENT_HISTORY_CONTENTS**: Contains the binary contents of all non-tip versions of a document, plus a copy of all non-tip versions metadata for full-text indexing.

# Differences in Comparision with Other PPM Center DMS Solutions

### Features

From a PPM Center user perspective, PPM Center Database DMS provides all the features of Documentum Integration, as well as a new feature to search only the latest versions of documents or to search all versions. The File System DMS solution only offers "attachment-like" behavior, and offers none of the features expected from a Document Management System.

Table 1. Supported DMS Features

| DMS Feature | Documentum | File System | PPM Center Database DMS |
|---|---|---|---|
| Versioning | Yes | — | Yes |
| Check in/Check out | Yes | — | Yes |
| Check out override | Yes | — | Yes |
| Full Text Search | Yes | — | Yes (after enabling full-text search) |
| Key Words | Yes | — | Yes |
| Tip only/History Search | — | — | Yes (after enabling full-text search) |

### UI Changes

There is almost no difference for a PPM User when using Documentum Integration or PPM Center Database DMS. The only difference is that when full-text search is enabled in PPM Center Database DMS, the search pages will display an option to "Search Historical Versions".

### Search Operators

The documents keywords search operators you can use are different between Documentum and PPM Center Database DMS. The search syntax is described on the Search Help page. You can click the question mark next to the document search field to open the Search Help page.
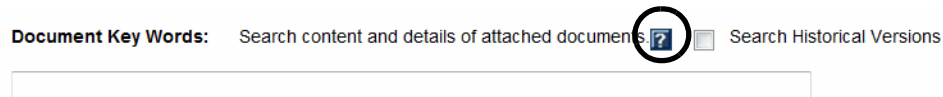


Table 2 summarizes the differences.

Table 2. Differences of Search Operators between Documentum and PPM Center Database DMS

| Operators | PPM Center Database DMS | Documentum |
|---|---|---|
| **"AND" queries** | You should insert the word and (case-insensitive) or & character between the search words. | By default, each query is an AND query. Since this is the default search behavior, there is no need to include the word and between search keywords. |
| **Phrase search** | By default, each query is a "Phrase" query. | You can also search for documents containing a specific phrase, or set of words in a specific order. Enclose the words in double quotes to enable this type of query. All words enclosed in double quotes must occur together and in the specified order for a document to be considered as a match. |
| **Exclusion of words** | To specifically exclude documents that contain a particular word, preface the keyword with a NOT (~) sign. | To specifically exclude documents that contain a particular word, preface the keyword by a minus (−) sign. In this case, all documents that contain the specified word are excluded from the results, even if they match other keywords in your query. |
| **"OR" queries** | You can insert the word or (case-insensitive) or \| character between the search words. | If you want to search for documents containing one OR another keyword, you can insert the word or between the search words. |
| **Combination search** | You may perform combination searches for documents by combining AND, OR, and/or ~ queries. If you want to search for documents containing both AND and OR queries, be sure to wrap sub-queries with parentheses " ( " and " ) ". | Any of these search formats can be used in combination. OR queries take precedence over AND queries. |
| **Search historical versions** | The Search Historical Versions option allows you to search content and/or version comments of historical versions of documents in addition to full search (including document content and document properties) of their current version. | Not supported. |
| **Synonyms** | Not supported. | To expand your search to include keywords that have a similar meaning as a word you have specified, preface your keyword with a tilde (~). In this case, a document is considered as a match if it contains the search keyword you specified, or additional words that have a similar meaning. |

# DMS Migration

This section provides step-by-step instructions on how to migrate to PPM Center Database DMS. Though this is practically as simple as clicking a button in the Administration Console, there are a few things of interest to know before proceeding.

---

---

## Before the Migration

- Tablespaces update: If your DBA wants to change the tablespace used to store the document contents, this should preferably be done before the migration, as there is no data in the tables at that point. If you already started a migration before but cancelled it, there are already some data in the PPM Center Database DMS DB tables; you can safely truncate these tables. But do NOT change anything in the KNTA_DOCUMENTS table, as it is used by all three DMS solutions to store document information and should never be modified manually.

- If you plan on using Full Text search, do NOT create the indexes on the tables before the migration. The reasons are:
  — Performance impact on migration, additional load on Database Server CPU during the migration due to indexes update.
  — Updating the index while new documents are bring added can result in more fragmented indexes, especially if the indexes are set to SYNC (ON COMMIT). You will get better results by creating the index after all documents are added.
  — Trying to remove a document (which can happen during the migration) while the TEXT index is currently being constructed can result in an Oracle error (ORA-29861: domain index is marked LOADING/FAILED/UNUSABLE).

- The default DMS Migration settings provides a good migration speed, but might result in a heavy load on PPM Service nodes and PPM Center database. The easy way to relieve the load during peak hours is to pause the migration (this can be easily done from the migration page in the Administration Console). However, if you want to alleviate the overall load on PPM Server during the migration process, you can:
  — Update the following parameters in the PPM Center's Administration Console before starting the migration (No PPM Server restart needed)
    – `DMS_MIGRATION_THREAD_COUNT` (Default: `3`): Number of threads used to migrate documents on a given PPM Service node.
    – `DMS_MIGRATION_DOCUMENTS_BATCH_SIZE` (Default: `1000`): Size of the queue of documents to migrate on a given PPM Service node. Every time the "DMS Migration" service runs on a Service node, the queue of documents to migrate is filled up.
    – `DMS_MIGRATION_DELAY_BETWEEN_DOCUMENT` (Default: `0`): Duration (in seconds) that a thread will wait between two documents to migrate.
  — Change the scheduling of the DMS Migration service on the Schedule Services page. Default setting is to run the service every 30 seconds. However, do NOT enable or disable that service manually from this page.

- If you are migrating from Documentum to PPM Center Database DMS, you might run out of Documentum sessions during the migration with the default configuration (including default Documentum Server configuration).

  The reason is that documents operations are asynchronous, and Documentum sessions can be busy for some time even when the document operation on PPM Server has completed. As the DMS Migration is a very demanding process, sometimes you might run out of Documentum sessions, which results in failure of some documents to be migrated.

  This is not a blocking issue as you can retry migrating failed documents until they are successfully migrated. However, it results in unneeded manual operations. You can reduce the probability of running into such an issue by resorting to the following options:

  — Use one migration thread only.
  — Add a delay between documents to migrate (of `1` second or more). This gives time for the asynchronous operations to complete. Considering the performance impact, you should only resort to this option after you have tried to migrate all documents at least once.
  — Reduce the documents migration batch size so there can be some idle time between two migration service triggering.
  — Increase the maximum number of Documentum sessions on Documentum server: edit the `$DOCUMENTUM/dba/config/<DOCBASE_NAME>/server.ini` file, and increase the value of `concurrent_sessions` (default: `100`).

## Starting the migration

To do so,

1  Log in to PPM Center and launch the Administration Console.

2  In the left panel of the Administration Console window, expand the **Administration Task** section, and then click **DMS Configuration**.

3  On the DMS Configuration page, click **Migrate**.

4  Select **PPM Center Database DMS** from the **Target DMS** drop-down list, then click **Next**.

5  Click **Start Migration**.

---

Warning:
There can be a delay of less than one minute between clicking **Start Migration** and the moment where the document starts to be actually migrated; this is due to the default scheduling of the DMS Migration service, which starts every 30 seconds.

---

## During the migration

You can pause the migration by clicking **Pause**, view detail of error (by clicking **View failed files logs**) if any document fails to be migrated, and retry to migrate any failed document.

---

Warning:
The progress bar is not updated automatically. You need to press **Refresh** to update the migration progress status.

---

## Migration Completed

Before moving to the next step (Transition Period), this is a good time to enable full-text search if you plan to use the feature.

If you move to Transition Period without enabling full-text search, the full-text search feature will not be available to PPM Center users and they cannot search documents by **Document Key Words**.

To enable full-text search in PPM Center Database DMS:

1   Create TEXT indexes, as described in Oracle TEXT Indexes Creation and Maintenance on page 12. Note that this step can take a long time (over 1 hour per 5 GB of documents to index).

2   In PPM Center Administration Console, go to **Application Configuration** and change the value of parameter `DMS_DB_ENABLE_FULLTEXT_SEARCH` to `true`, and then save the change.

## Transition Period

Once the migration is completed, you can move to "Transition Period", during which PPM Center Database DMS becomes active, but your old DMS solution (File System or Documentum) is still active in the background and save copies of newly created documents. This way, you can try PPM Center Database DMS for some time. If you find that the new DMS solution does not meet your requirements, or if you encounter any issue with the new DMS solution, you can always switch back to the previous DMS solution without causing loss of any documents to PPM Center users.

## After the Migration

Once you have properly tested PPM Center Database DMS in the Transition Period, you can simply click **Commit Migration** to stop synchronizing new documents to the old DMS. Once you click **Commit Migration**, you are officially moved to the PPM Center Database DMS, and the old DMS is retired.

# Database Sizing

The only sizing impact of using PPM Center Database DMS is in your Database Server, as everything happens in the database.

## Database Disk Space

The database disk space consumed by PPM Center Database DMS can be split into 3 different parts:

- Space taken up by Document contents (about equal to total documents size)
- Space taken up by Documents metadata (proportional to number of documents and propensity of PPM Users to input lengthy documents descriptions, key words, and version comments.
- Space taken up by TEXT Indexes
    - Index on document contents, proportional to total size and types of documents (index on pure text file such as a logs file might take more space than the file itself, while index on a picture file, even very large, does not take any space as no text content can be extracted from the picture).
    - Index on text metadata, proportional to the size of metadata (as all indexed metadata is text and can be indexed).

There is a tool available from HP Support — PPM DMS Statistics & PPM Database DMS Sizing tool — to automatically estimate the extra disk space needed in database to migrate an existing PPM Server environment to PPM Center Database DMS. The tool computes estimates based on quantity of documents, total size of documents, and types of documents. All this information are retrieved from the KNTA_DOCUMENTS table of PPM Center database.
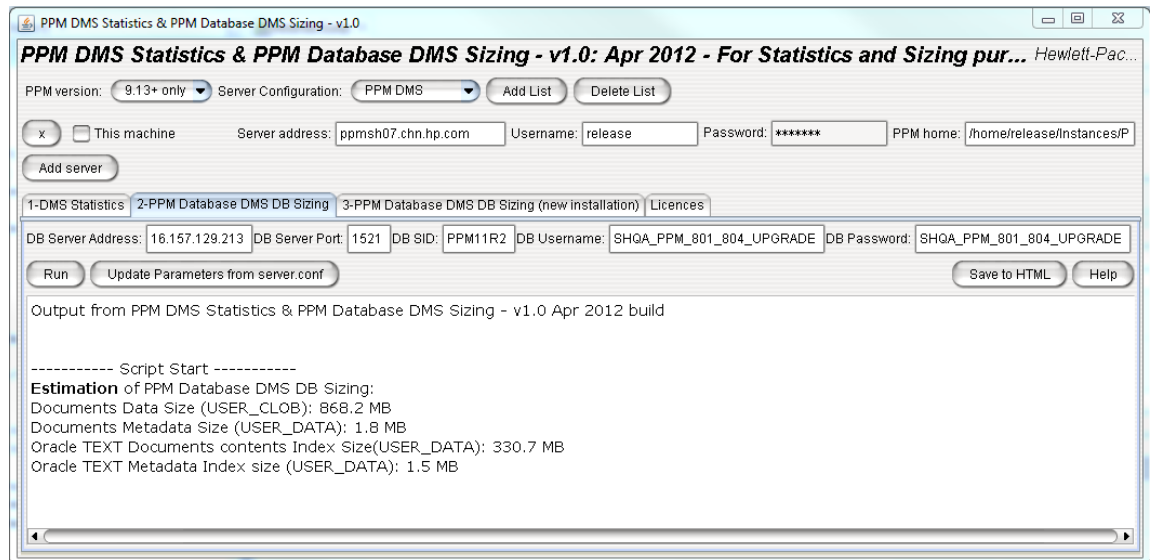


The tool can also provide estimates for new PPM Center installations, based on user's input on planned usage of documents in PPM Center (estimated total number and size of documents, text concentration of stored documents).
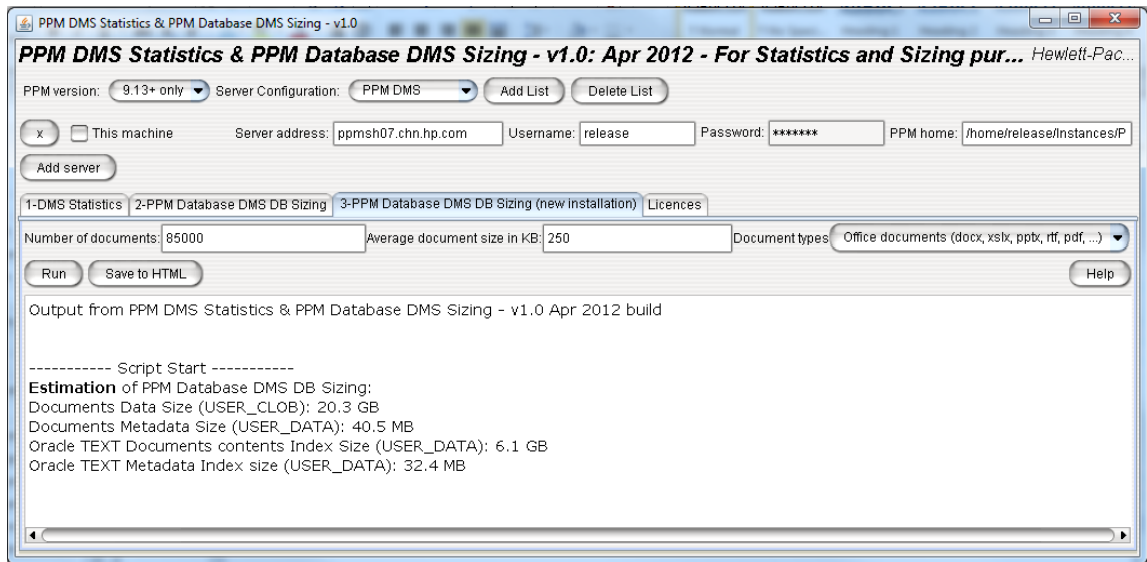
Customer Case

Here is an example of Database Space consumption from an existing PPM Center customer who used Documentum previously.

Customer documents statistics:

- Total number of documents: 128,853
- Average number of versions per document: 1.003
- Maximum number of versions for a document: 8
- Total number of documents versions: 129,221
- Cumulated size of all documents versions (estimation): 32,056 MB
- Document repartition by type: (see Table 3 below)

Table 3. Customer Case: Document repartition by type

| Document Type | Text Content | Files Count | Cumulated total size (MB) |
|---|---|---|---|
| RTF | Medium | 2,175 | 10,513 |
| MSG | Medium | 49,088 | 6,507 |
| DOC | Medium | 21,767 | 4,684 |
| TXT | High | 14,365 | 3,367 |
| XLS | Medium | 5,457 | 1,500 |
| PDF | Medium | 8,634 | 1,399 |
| ZIP | Unknown | 864 | 797 |
| XLSX | Medium | 1,248 | 650 |
| DOCX | Medium | 2,604 | 649 |
| BMP | None | 302 | 386 |
| none | Unknown | 1,922 | 342 |
| HTM | High | 16,783 | 319 |
| TIF | None | 667 | 242 |
| REP | Unknown | 22 | 141 |

Table 3. Customer Case: Document repartition by type

| Document Type | Text Content | Files Count | Cumulated total size (MB) |
|---|---|---|---|
| EXE | None | 21 | 76 |
| DOT | Medium | 126 | 57 |
| JPG | None | 349 | 54 |
| MDB | Medium | 2 | 48 |
| SQL | High | 782 | 27 |
| PPT | Medium | 13 | 21 |
| Other | Unknown | 2,030 | 278 |

Text-only documents contain high level text contents, Microsoft Office documents contain medium level text contents, and images as well as binary documents contain practically no text content. This is reflected in the document contents index size.

*Text Index creation time:* ~ 6 hours (on a 2-instance RAC server with 4 x Dual Core CPU Intel(R) Xeon(R) E5540 @ 2.53GHz per instance).

*Disk space consumed by PPM Center Database DMS:*

- Documents binary contents (BLOB Columns): 32,041 MB
- DMS Tables without BLOB columns (not including KNTA_DOCUMENTS): 30 MB
- Total Metadata Indexes: 21.7 MB
- Document Contents Indexes: 6,011 MB

As you can see, the size of document contents index (~6 GB) accounts for almost 20% of the documents contents size (~32 GB). This is a relatively ratio, which can be explained by the large amount of office documents in the customer's attachments.

It might be possible that the index size is larger than the document contents if all the attachments are pure text files (such as .txt, .sql, and .log files).

Warning:
The BLOB columns containing the documents are appearing slightly smaller than the estimated total size of documents (15 MB smaller). The reason is that the total size of the documents is an estimation, computed using "Versions count * file size of latest version" for each document. It appears that the earlier versions of documents are smaller in size in average, resulting in the observed discrepancy.

## Database Server CPU

If you have no plan to enable full-text search (or more precisely, if you do not need to create TEXT indexes), there is virtually no impact on database CPU, as reading and writing documents contents to database are mainly I/O intensive operations.

If you do create TEXT indexes, but schedule them to be updated only out of peak hours, there should be no need to consider an upgrade of DB Server CPU based on standard PPM Center sizings.

However, if you are using frequent index updates, or use SYNC (ON COMMIT) indexes for real time indexing, a CPU upgrade of your DB Server might be necessary, especially if your PPM Center users tend to store a large amount of text intensive files in PPM Center (more than one Gigabyte of new documents per week, with peak document activity concentrated on a few hours in the week).

# Oracle TEXT Indexes Creation and Maintenance

## TEXT Index Creation

Why is TEXT index not created automatically, like other PPM Indexes?

PPM Center Database DMS relies on creation of Oracle TEXT indexes on documents contents and metadata to provide full-text search on documents.

The index creation is not included as part of PPM Center 9.13 installation for multiple reasons:

- PPM Center Service Packs installation SQL Scripts can only be run with PPM DB User, not SYS; however, a PPM DB User needs to be granted CTXAPP and CREATE JOBS access grants before he can create TEXT indexes, and only SYS can provide these grants to PPM DB User.

- If for some reason you plan to use PPM Center Database DMS but have no plan to use full-text search, there is no need to waste DB resources with these indexes, as they are both space and CPU consuming.

- You need to decide or customize the parameters to use to create the TEXT index.

  HP does not offer the recommendation as the answers for the following questions vary with each customer:

  — Do you prefer real-time indexing at the cost of a performance impact and a fragmented index on the long run, or is it acceptable to refresh the index only every 24 hours?

  — Is the WORLD_LEXER HP uses by default acceptable for you? Or should you rather use a Japanese Lexer in case you store many Japanese documents?

  Your DBAs are likely make a better work at creating the TEXT indexes you need by providing default index creation scripts.

## Default Text Index Creation Script

The default TEXT index creation script can be found at the following path on an instance of PPM Center 9.13 or later:

`<PPM_Home>/utilities/database_dms/DatabaseDMS_FullTextSearch_Scripts.zip`

The zip package contains two files:

- The script to run as SYS user to give PPM DB User the required grants to create TEXT indexes.
- The default index creation script. You may want to customize this script to meet your requirements.

TEXT Index Creation Parameters

- The first parameter that you may want to update is the lexer used when indexing documents. By default, HP uses WORLD_LEXER as it supports a wide range of languages (including all languages supported by PPM Center). You can refer to Oracle documentation for more choices of lexer that might give you better results:

  **http://docs.oracle.com/cd/B28359_01/text.111/b28304/amultlng.htm#CEGBCDHJ**

- The other parameter that you may want to modify when creating indexes is the delay between TEXT indexes refresh. This is set in the "`SYNC (every sysdate+XXX)`" of the index creation SQL, where `XXX` is the average duration between 2 index refresh expressed in days. So, using "T/1440" allows you to easily express duration T in minutes.

  The default script creates indexes that will be refreshed every 10 minutes [`SYNC (every sysdate+1/144)`]. In other words, when users add a document to PPM Center, they may have to wait up to 10 minutes before the document can appear in the search results.

If you do not want to wait for a document to be indexed after adding it to PPM Center, you can use the option `SYNC (ON COMMIT)`. However, note that when doing so, there will be a noticeable performance impact as the document indexing is done as part of the transaction. Moreover, this results in a very fragmented index on the long run, requiring frequent optimizations, and possibly indexes full rebuilds.

## TEXT Indexes Maintenance

If your users start to witness slow document searches after some time while using PPM Center Database DMS, the possible root cause might be TEXT index fragmentation.

There are two solutions when this occurs:

- Drop and re-create the index(es). This can be time consuming if you have lots of documents, and during the index rebuilding time, not all documents might be searchable.
- Optimize the index(es). There are multiple level of index optimization depending on how complete you want the optimization to be. For example, to perform a fast optimization of the tip document contents index, you should run:

```
exec ctx_ddl.optimize_index('DMS_TIP_DOC_IDX','FAST');
```

Extensive documentation on Oracle TEXT index optimization can be found on Oracle web site, such as:

**http://docs.oracle.com/cd/B28359_01/text.111/b28303/ind.htm#i1007604**

**http://docs.oracle.com/cd/B28359_01/text.111/b28304/cddlpkg.htm#CCREF0638**

# Known Issues

There are a few known issues to PPM DMS Migration:

- Exception when starting migration if using a locale that does not use period as decimal separator, for example, French, or Russian. (This issue occurs on PPM Center version 9.13 only.)
- Migration process is slow if there are more than 50,000 documents to migrate.
- All documents will fail during migration, pointing to an error of insufficient privileges of PPM DB User on DBMS_RANDOM oracle package

All these issues are fixed in a hotfix available for PPM Center version 9.13 and 9.14. Contact HP Support to get the hotfix under the reference `HOTFIX_-_DMS_9.13_9.14_-_QCCR1L45846_-_QCCR1L45466.`