

[EN] Distributing Indexing and Searching Services

Indexing and Searching Services

Since the 5.0 version of the dLibra system, it has been possible to divide the search service (Search Server) known from version 4.0 into two separate services, for searching (Search Server) and indexing (Index Server), which are independent from each other. In the default configuration of the dLibra server, both services operate within the framework of the same Java virtual machine and share the same catalog with a search index. Such a configuration corresponds to the configuration from version 4.0 (regardless of whether the search service is located together with other services of the dLibra server or if it is run on a separate machine). Apart from the default configuration, the dLibra system makes available two other options, which differ with respect to the physical distribution of the services.

Search Server and Index Server on the Same Machine

In that solution, the Search Server and Index Server services are located on the same physical machine but operate within the framework of separate Java virtual machines. In order to obtain such a configuration, the user has to take the following steps:

- Copy the server directory (without the “logs” subdirectory) to the new location and, optionally, change the name of the directory so that it indicates that it refers to the Index Server service. Further in the text, we will – for clarity’s sake – call the original directory of the server *SE* and the copied catalog with the Index Server – *IS*.
- Delete the *SE/conf/is* subdirectory.
- Open the *SE/conf/server.xml* file in the <service-list> section; remove the part concerning the Index Server service.
- In subdirectory *IS/conf*, remove all subdirectories apart from *is*, *mx*, and *wrapper*.
- Open file *IS/conf/server.xml*; in section <service-list>, remove all entries apart from those which concern the Index Server service and the JMX interface. Also, change the value of parameter **serverPort** to a different one than its counterpart in file *SE/conf/server.xml* (for example, if parameter **serverPort** has value 10051 in file *SE/conf/server.xml*, then the value can be set to 11051 in file *IS/conf/server.xml*). If section <systemServicesUrl> is commented out, then the comment should be removed, and the URL should be entered in the System Services service (the address and port of the server correspond to the initial server).
- In the database, in table SYS_SERVICES, modify the entry pertaining to the Index Server service, by changing the value of column SER_PORT to the value set in file *IS/conf/server.xml* in the previous step.
- In the same table, add a new tuple corresponding to the JMX interface for the new Index Server. Here is an example for the Oracle database:

```
insert into SYS_SERVICES
(SER_ID, SER_TYPE, SER_DESCRIPTION,
 SER_VERSION, SER_CONNECTED, SER_PASSWORD,
 SER_HOST, SER_PORT)
values
(SYS_SERVICES_SER_ID_SEQ.NEXTVAL, 'mx', 'dLibra JMX Management Service',
 '5.0', 0, '@ME_PASSWD@',
 '@SERVER_HOSTNAME@', @SERVER_PORT@);
```

Parameters @ME_PASSWD@, @SERVER_HOSTNAME@, and @SERVER_PORT@ should be changed in accordance with the values in file *IS/conf/server.xml*.

- Copy the directory which contains search indexes (the path to that directory is saved in file *SE/conf/lucene.properties*, in key **indexDirectory**). Enter the path to the new index location in key **indexDirectory** of file *IS/conf/lucene.properties*. Do the same for the directory which contains the backup copies of search indexes, taking into account the fact that the path to that directory is saved at key **indexBackupDirectory**.
- Start up the dLibra server from the *SE* directory, and then from the *IS* directory. After the first startup of such a service configuration, file *services.dat* will be created in the directories of both servers. It should be used for generating new licenses.

Search Server and Index Server on Separate Machines

In that solution, the Search Server and Index Server services are located on separate physical machines and, consequently, operate within the framework of separate Java virtual machines. In order to obtain such a configuration, the user has to take the following steps:

- Copy the server directory (without the “logs” subdirectory) to the new location on a separate machine and, optionally, change the name of the directory so that it indicates that it refers to the Index Server service. Further in the text, we will – for clarity’s sake – call the original directory of the server *SE* and the copied catalog with the Index Server – *IS*.
- Delete the *SE/conf/is* subdirectory.
- Open the *SE/conf/server.xml* file in the <service-list> section; remove the part concerning the Index Server service.
- W podkatalogu *IS/conf* usuwamy wszystkie podkatalogi oprócz *is*, *mx*, *wrapper*.
- Open file *IS/conf/server.xml*; in section <service-list>, remove all entries apart from those which concern the Index Server service and the JMX interface. Also, change parameter **serverHost** to a value which corresponds to the IP address of the machine on which the Index Server has been

placed; the value of parameter **serverPort** can remain as is. If section **<systemServicesUrl>** is commented out, then the comment should be removed, and the URL should be entered in the System Services service (the address and port of the server correspond to the initial server).

- In the database, in table SYS_SERVICES, modify the entry pertaining to the Index Server service, by changing the values of columns SER_HOST and SER_PORT to the values set in file *IS/conf/server.xml* in the previous step.
- In the same table, add a new tuple corresponding to the JMX interface for the new Index Server. Here is an example for the Oracle database:

```
insert into SYS_SERVICES
(SER_ID, SER_TYPE, SER_DESCRIPTION,
 SER_VERSION, SER_CONNECTED, SER_PASSWORD,
 SER_HOST, SER_PORT)
values
(SYS_SERVICES_SER_ID_SEQ.NEXTVAL, 'mx', 'dLibra JMX Management Service',
 '5.0', 0, '@ME_PASSWD@',
 '@SERVER_HOSTNAME@', @SERVER_PORT@);
```

Parametry @ME_PASSWD@, @SERVER_HOSTNAME@ i @SERVER_PORT@ należy zamienić zgodnie z wartościami umieszczonymi w pliku *IS/conf/server.xml*.

- Copy the directory which contains search indexes (the path to that directory is saved in file *SE/conf/lucene.properties*, in key **indexDirectory**) to the machine with the Index Server service. Enter the path to the new index location in key **indexDirectory** of file *IS/conf/lucene.properties*. Do the same for the directory which contains the backup copies of search indexes, taking into account the fact that the path to that directory is saved at key **indexBackupDirectory**.
- Start up the dLibra server from the SE directory, and then from the IS directory. After the first startup of such a service configuration, file *services.dat* will be created in the directories of both servers. It should be used for generating new licenses.

Configuring the Synchronization of Search Indexes

The separation of the indexing and searching services makes them operate on independent search indexes. As a result, the index used for searching is not always current, and it has to be periodically refreshed. The dLibra server has a mechanism for synchronizing the indexes between the Index Server and Search Server services. The synchronization is run in accordance with the settings of the periodic task defined in file *SE/conf/se/jobs.xml*. In the default synchronization plan, the content index (as well as the `_spell` dictionary indexes) is synchronized separately from metadata indexes because of its size. The plan can be adjusted to individual needs, by modifying the CRON expressions for particular tasks. Additional periodic tasks related to particular index types can also be created.

Notes

- In both solutions, the Index Server service is separated from a configuration placed on one machine. For that reason, the Search Server is placed together with the remaining services, but in need, it can be separated and placed on another machine, just like the Index Server.
- Since index synchronization makes use of the system temporary directory, there should be enough free space on the drive. During a synchronization, a mechanism for creating backup copies which operates on ZIP archives is used, so the amount of free space should be sufficient for storing both the ZIP archive of all indexes and their unpacked versions, as well as the current version of search indexes.