

Rozmieszczenie usług indeksowania i wyszukiwania

Usługi indeksowania i wyszukiwania

Od wersji 5.0 dLibry istnieje możliwość rozdzielenia usługi wyszukiwania (Search Server) znanej z wersji 4.0 na dwie osobne usługi zajmujące się wyszukiwaniem (Search Server) i indeksowaniem (Index Server) niezależnie od siebie. Domyślna konfiguracja serwera dLibry zakłada, że obie usługi działają w ramach tej samej maszyny wirtualnej Java i współdzielą ten sam katalog z indeksem wyszukiwawczym. Konfiguracja taka odpowiada konfiguracji z wersji 4.0 (niezależnie od tego, czy usługa wyszukiwania umieszczona jest wraz z innymi usługami serwera dLibra czy uruchomiona jest na oddzielnej maszynie). Oprócz domyślnej konfiguracji dLibra udostępnia dwie inne możliwości różniące się fizycznym rozmieszczeniem usług.

Search Server i Index Server na tej samej maszynie

Rozwiązanie to zakłada, że usługi Search Server i Index Server będą umieszczone na jednej fizycznej maszynie, jednak będą działać w ramach oddzielnych maszyn wirtualnych Java. W celu uzyskania takiej konfiguracji należy wykonać następujące kroki:

- Kopiujemy katalog serwera (bez podkatalogu *logs*) do nowej lokalizacji i opcjonalnie zmieniamy nazwę katalogu w taki sposób, by wskazywała, że dotyczy usługi Index Server. Dalej na potrzebę przejrzystości opisu pierwotny katalog serwera nazywać będziemy *SE* natomiast skopiowany katalog z usługą Index Server nazywać będziemy *IS*.
- Usuwamy podkatalog *SE/conf/is*.
- Otwieramy plik *SE/conf/server.xml* i w sekcji **<service-list>** usuwamy wpis dotyczący usługi Index Server.
- W podkatalogu *IS/conf* usuwamy wszystkie podkatalogi oprócz *is*, *mx*, *wrapper*.
- Otwieramy plik *IS/conf/server.xml* i w sekcji **<service-list>** usuwamy wszystkie wpisy oprócz tych dotyczących usługi Index Server i interfejsu JMX. Oprócz tego zmieniamy parametr **serverPort** na wartość inną niż odpowiedni parametr w pliku *SE/conf/server.xml* (np. jeśli w *SE/conf/server.xml* parametr **serverPort** ma wartość 10051 to w pliku *IS/conf/server.xml* parametr ten można ustawić na wartość 11051). Jeśli sekcja **<systemServices Url>** jest zakomentowana, należy komentarz usunąć i wprowadzić URL do usługi System Services (adres i port serwera odpowiadają pierwotnemu serwerowi).
- W bazie danych modyfikujemy w tabeli *SYS_SERVICES* wpis dotyczący usługi Index Server zmieniając wartość kolumny *SER_PORT* na taką, jaka ustawiona została w pliku *IS/conf/server.xml* w poprzednim kroku.
- W tej samej tabeli dopisujemy nową krotkę odpowiadającą interfejsowi JMX dla nowego Index Servera. Przykład dla bazy danych Oracle:

```
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*.

- Kopiujemy katalog zawierający indeksy wyszukiwawcze (ścieżka do tego katalogu zapisana jest w pliku *SE/conf/lucene.properties* w kluczu **indexDirectory**). Ścieżkę do nowej lokalizacji indeksów wpisujemy do klucza **indexDirectory** pliku *IS/conf/lucene.properties*. Podobnie postępujemy z katalogiem zawierającym kopie zapasowe indeksów wyszukiwawczych, przy czym należy uwzględnić fakt, iż ścieżka do tego katalogu zapisana jest pod kluczem **indexBackupDirectory**.
- Uruchamiamy serwer dLibra z katalogu *SE*, następnie z katalogu *IS*. Po pierwszym uruchomieniu takiej konfiguracji usług w katalogach obu serwerów powstanie plik *services.dat*, który należy wykorzystać do wygenerowania nowych licencji.

Search Server i Index Server na osobnych maszynach

Rozwiązanie to zakłada, że usługi Search Server i Index Server będą umieszczone na osobnych fizycznych maszynach, więc naturalnie będą działać w ramach oddzielnych maszyn wirtualnych Java. W celu uzyskania takiej konfiguracji należy wykonać następujące kroki:

- Kopiujemy katalog serwera (bez podkatalogu *logs*) do nowej lokalizacji na osobnej maszynie i opcjonalnie zmieniamy nazwę katalogu w taki sposób, by wskazywała, że dotyczy usługi Index Server. Dalej na potrzebę przejrzystości opisu pierwotny katalog serwera nazywać będziemy *SE* natomiast skopiowany katalog z usługą Index Server nazywać będziemy *IS*.
- Usuwamy podkatalog *SE/conf/is*.
- Otwieramy plik *SE/conf/server.xml* i w sekcji **<service-list>** usuwamy wpis dotyczący usługi Index Server.
- W podkatalogu *IS/conf* usuwamy wszystkie podkatalogi oprócz *is*, *mx*, *wrapper*.
- Otwieramy plik *IS/conf/server.xml* i w sekcji **<service-list>** usuwamy wszystkie wpisy oprócz tych dotyczących usługi Index Server i interfejsu JMX. Oprócz tego zmieniamy parametr **serverHost** na wartość odpowiadającą adresowi IP maszyny, na której umieszczono Index Server, wartość

parametru **serverPort** może pozostać bez zmian. Jeśli sekcja **<systemServicesUrl>** jest zakomentowana, należy komentarz usunąć i wprowadzić URL do usługi System Services (adres i port serwera odpowiadają pierwotnemu serwerowi).

- W bazie danych modyfikujemy w tabeli *SYS_SERVICES* wpis dotyczący usługi Index Server zmieniając wartości kolumn *SER_HOST* i *SER_PORT* na takie, jakie ustawione zostały w pliku *IS/conf/server.xml* w poprzednim kroku.
- W tej samej tabeli dopisujemy nową krotkę odpowiadającą interfejsowi JMX dla nowego Index Servera. Przykład dla bazy danych Oracle:

```
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*.

- Kopiujemy katalog zawierający indeksy wyszukiwawcze (ścieżka do tego katalogu zapisana jest w pliku *SE/conf/lucene.properties* w kluczu **indexDirectory**) na maszynę z usługą Index Server. Ścieżkę do nowej lokalizacji indeksów wpisujemy do klucza **indexDirectory** pliku *IS/conf/lucene.properties*. Podobnie postępujemy z katalogiem zawierającym kopie zapasowe indeksów wyszukiwawczych, przy czym należy uwzględnić fakt, iż ścieżka do tego katalogu zapisana jest pod kluczem **indexBackupDirectory**.
- Uruchamiamy serwer dLibra z katalogu *SE*, następnie z katalogu *IS*. Po pierwszym uruchomieniu takiej konfiguracji usług w katalogach obu serwerów powstanie plik *services.dat*, który należy wykorzystać do wygenerowania nowych licencji.

Konfiguracja synchronizacji indeksów wyszukiwawczych

Rozdzielenie usług indeksującej i wyszukującej powoduje, że działają one na niezależnych od siebie indeksach wyszukiwawczych. Konsekwencją tego jest fakt, iż indeks wykorzystywany do wyszukiwania nie zawsze jest aktualny i wymaga okresowego odświeżania. Serwer dLibry wyposażony jest w mechanizm synchronizacji indeksów pomiędzy usługami Index Server i Search Server. Synchronizacja uruchamiana jest zgodnie z ustawieniami zadania okresowego zdefiniowanego w pliku *SE/conf/se/jobs.xml*. Domyślny plan synchronizacji zakłada, że indeks treści (a także indeksy słownikowe **_spell*) z racji swojego rozmiaru będzie synchronizowany oddzielnie od indeksów metadanych. Plan ten można dostosować do własnych potrzeb modyfikując wyrażenia CRON dla poszczególnych zadań. Możliwe jest również utworzenie dodatkowych zadań okresowych związanych z konkretnym typem indeksów.

Uwagi

- Oba rozwiązania zakładają, że wydzielenie usługi Index Server odbywa się z konfiguracji umieszczonej na 1 maszynie. W związku z tym usługa Search Server jest umieszczona razem z pozostałymi usługami, ale w razie potrzeby można ją wydzielić na osobną maszynę, podobnie jak Index Server.
- Z uwagi na to, iż synchronizacja indeksów wykorzystuje systemowy katalog tymczasowy należy zapewnić odpowiednią ilość wolnego miejsca na dysku. W czasie synchronizacji wykorzystywany jest mechanizm kopii zapasowych operujący na archiwach ZIP, należy więc tak dostosować ilość wolnego miejsca, aby było możliwe jednoczesne przechowywanie archiwum ZIP wszystkich indeksów oraz ich postaci niespakowanych równolegle z bieżącą wersją indeksów wyszukiwawczych.