



Copyright © 2014 NTT DATA INTRAMART CORPORATION

目次

- 1. 改訂情報
- 2. はじめに
 - 2.1. 本書の内容
 - 2.2. 対象読者
 - 2.3. 本書に記載されている外部サイトのURL
 - 2.4. 用語解説
 - 2.5. 本書の構成
- 3. Solrの概要
 - 3.1. Apache Solr とは
 - 3.2. RDBとの比較
- 4. Solrのセットアップ
 - 4.1. 前提条件
 - 4.2. セットアップ手順
- 5. Solrの設定
 - 5.1. コアの追加・削除
 - 5.2. コアごとの設定
- 6. Solrの操作
 - 6.1. インデックスの削除
 - 6.2. Solrの移行
 - 6.3. Solrのバックアップ

改訂情報

変更年月日	変更内容
2014-05-30	初版
2014-12-01	第2版 下記を追加・変更しました <ul style="list-style-type: none"> 「利用するJavaのバージョンについて」に「Java SE Development Kit 7u55」以降ではJVM起動オプションを指定する必要がない旨を追加 「Solr管理画面の確認」に説明を追加
2017-04-01	第3版 下記を追加・変更しました <ul style="list-style-type: none"> 「Solrのセットアップ」のセットアップツールのダウンロード方法の説明を修正
2018-04-01	第4版 下記を変更しました <ul style="list-style-type: none"> 「Solrのセットアップ」に最適化処理に関する注意事項を追加しました。
2018-08-01	第5版 下記を変更しました <ul style="list-style-type: none"> 「サーバの起動と停止」に Payara の説明を追加 「Solrの移行」に Payara の説明を追加
2020-08-01	第6版 下記を変更しました <ul style="list-style-type: none"> 「サーバの起動と停止」の「Jetty を利用する場合」に注意書きを追加 「Solrの移行」に Tomcat の説明を修正 「Solrの移行」に Jetty の説明を修正
2021-04-01	第7版 下記を変更しました <ul style="list-style-type: none"> Solrのバージョンアップに伴い、全体的に説明を修正 「利用するJavaのバージョンについて」の記述がサポート外のJDKの記述であるため削除

はじめに

本書の内容

本書では IM-ContentsSearch for Accel Platform で利用している全文検索エンジンサーバの Apache Solr について説明します。

IM-ContentsSearch モジュールの利用方法は「[intra-mart Accel Platform セットアップガイド](#)」 - 「[IM-ContentsSearch](#)」を参照してください。

テナント環境セットアップ時の設定については「[intra-mart Accel Platform セットアップガイド](#)」 - 「[Apache Solr接続情報](#)」を参照してください。

対象読者

以下の利用者を対象としています。

- IM-ContentsSearch をセットアップする方
- IM-ContentsSearch を運用・管理する方

本書に記載されている外部サイトのURL

本書内で記載されている外部URLは 2021年4月 現在のものです。

用語解説

本書中の表記について以下に記載します。

- IM-ContentsSearch for Accel Platform を省略して IM-ContentsSearch と表記します。
- Apache Solr を省略して Solr と表記します。
- リレーショナルデータベース を省略して RDB と表記します。
- Apache Solr がデプロイされたWeb Application Serverを Solrサーバ と表記します。
- セットアップ資材を展開したディレクトリを %SOLR_HOME% と表記します。

本書の構成

本書は以下の構成です。

- [Solrの概要](#)
Solr の概要・基本的な知識を紹介します。
- [Solrのセットアップ](#)

- [Solrの設定](#)

Solr の設定項目について説明します。

- [Solrの操作](#)

IM-ContentsSearch を利用するにあたって必要となる Solr の運用上の手順などについて説明します。

Solrの概要

ここでは Solr の概要や、基本的な知識について紹介します。

項目

- [Apache Solr とは](#)
- [RDBとの比較](#)

Apache Solr とは

Apache Solr は、高性能の全文検索エンジンライブラリ Lucene を使って構築されたJavaベースの全文検索エンジンサーバです。

Solr は全文検索を行う対象となる文書・データから索引語を抽出して、文書群への位置情報を単語ごとに保持した転置インデックスを保持します。

この転置インデックスをあらかじめ作成しておくことで、RDBでは実現が難しい高速な全文検索が実現可能です。



コラム

Solr についての詳しい情報は、Apache Software Foundation のWeb サイトでご確認ください。

<http://lucene.apache.org/solr/> (English)

また「Solr Wiki」では、Solr に関する様々な資料が掲載されていますのでご活用ください。

<http://wiki.apache.org/solr/> (English)

RDBとの比較

RDB と Solr における用語の対比表は以下の通りです。

用語比較

RDB	Solr	備考
スキーマ	→ コア	詳細については次項にて説明します。
テーブル	→ スキーマ	詳細については次項にて説明します。
レコード	→ ドキュメント	IM-ContentsSearch ではコンテンツ という形でラップしています。
カラム	→ フィールド	
プライマリキー	→ IDフィールド	

RDB	Solr	備考
ユーザ	→ なし	
コミット	→ コミット	詳細については次項にて説明します。
ロールバック	→ ロールバック	詳細については次項にて説明します。
なし	→ 最適化	詳細については次項にて説明します。

詳細説明

- コア

RDBのスキーマに相当する機能で、1つの Solr 上に複数のコアを保持することが可能です。Solr に対して複数コアの設定を行った場合、コアごとにアクセスするURLが異なります。尚、IM-ContentsSearch 用に配布している Solr では、標準で **default** という名前のコアを定義しています。

- スキーマ

RDBのテーブルに相当する機能です。Solr では1コアに1スキーマのみ定義可能です。スキーマ上のプライマリキーに相当するIDフィールドは、システムを通して一意となる値を設定する必要があります。

また、Solr のスキーマには以下の特徴があります。

1. 配列型の値をサポートしている
2. 動的フィールド（データ型のみ指定し後から自由に追加できるフィールド）がある

以上の特徴により、RDBよりも柔軟なデータ構造を持たせることができます。

- 更新処理

IM-ContentsSearch にはRDBのupdate文のようにレコード内の特定カラムを更新する機能がありません。

そのため、いずれかの値に変更があった場合には該当のドキュメントを再度作成する必要があります。

- トランザクション

Solr には分散トランザクションの概念がありません。

そのため複数のプロセスから同時に登録処理が行われた場合、先行するプロセスのコミット処理によって、後続するプロセスが登録したデータが途中でコミットされる可能性があります。

この問題を回避するために IM-ContentsSearch では、クローラジョブは1つのジョブネットにまとめ、更にそのジョブネットは並列実行を許可しない設定としています。

- コミット

Solr におけるコミットは、登録済みデータを検索結果に反映させる処理です。

RDBのようにトランザクション処理を確定するためのコミット処理ではありません。

- ロールバック

Solr におけるロールバックは、登録済みでコミット前のデータを破棄する処理です。

先述の通り Solr には分散トランザクションの概念がないため、複数のプロセスから同時に登録処理が行われていた場合、別プロセスが登録したデータも削除されます。

■ 最適化

Solr ではコミット処理を行うたびに索引情報を保持したバイナリファイルが作成されるため、コミットを繰り返すとファイル数が増え続けます。

ファイル数が増大することで、以下の問題が発生します。

1. 検索時にロードするバイナリファイルが増加しI/O負荷が高くなることで、検索速度が低下する。
2. ファイルディスクリプタ、またはファイルハンドルの消費量が多くなり、場合によっては不足する可能性がある。

この問題を回避するために、Solr には最適化 (optimize) 処理が用意されています。

最適化処理を実行することで、バイナリファイルを最小限にまとめ直し、上記の問題を回避することが可能です。

IM-ContentsSearch では、以下の方法で Solr に対して最適化処理を実行することが可能です。

1. クローラジョブネットの最後に最適化ジョブを実行する。 **(推奨)**
2. APIから最適化処理を実行する。 **(非推奨)**
3. ジョブの終了処理にて上記の最適化APIを実行する。 **(非推奨)**

注意

最適化処理はバイナリファイルのサイズ、すなわち登録されたデータの量によって処理時間が増加するため、実行を最小限にする必要があります。

すべてのコミット処理が終了するタイミングで最適化処理を実行するために、ジョブネットの実行ジョブの最後に最適化ジョブを設定することを推奨します。

Solrのセットアップ

IM-ContentsSearch を利用するための Solr のセットアップ方法について説明します。

項目

- [前提条件](#)
- [セットアップ手順](#)
 - [セットアップ資材の取得・展開](#)
 - [サーバの起動と停止](#)
 - [Solr管理画面の確認](#)

前提条件

- リリースノートに記載されているシステム要件を満たしている必要があります。
 - 詳細は「[リリースノート](#)」 - 「[Apache Solr システム要件](#)」を参照してください。



注意

Solr はディスクアクセスが大量に発生する場合があります。
 そのため、intra-mart Accel Platform がセットアップされたサーバと同一筐体に構築した場合、パフォーマンスに影響が出る可能性があります。
 運用環境では intra-mart Accel Platform と Solr を別サーバに構築することを強く推奨します。



注意

Solr における最適化処理では、一時的に登録されている全データを複製します。
 そのため、最適化処理時には、Solrサーバ上で、通常使用しているディスク容量の2倍以上のディスクの空き容量を必要としますので、注意してください。

セットアップ手順

IM-ContentsSearch では、Solr のセットアップ資材を提供しています。

- アプリケーションサーバ「Jetty」を利用した構築および起動



コラム

2021 Spring からアプリケーションサーバは同梱されている Jetty を使用します。
 Resin/Tomcat/Payara へデプロイするためのセットアップツールは提供しなくなりました。

セットアップ資材の取得・展開

1. セットアップ資材の取得

Solr のセットアップ資材を取得します。

セットアップツールは以下のURLよりダウンロードできます。ダウンロードには製品のライセンスキーが必要です。

[プロダクトファイルダウンロード](#)

2. セットアップツールの展開（解凍）

セットアップ資材のファイル <solr.zip> を取得後、任意のパスに展開します。

本書では、例として以下のディレクトリを指定します。

- Windowsの場合

```
C:\solr
```

- Linuxの場合

```
/usr/local/solr
```

セットアップ資材のファイルを展開したディレクトリを、以下 %SOLR_SETUP_HOME% と略します。

サーバの起動と停止

1. サーバの起動

- Windowsの場合

```
C:\solr\bin> solr.cmd start
```

- Linuxの場合

```
/usr/local/solr/bin/solr start
```

コラム

Solrの起動時、以下のようなエラーが発生する場合、ポート番号の設定を変更してください。

-p %PORT% と指定することで Solr が使用するポート番号を変更できます。

エラーログ出力例

```
ERROR: Process 4 is already listening on port 8983. If this is Solr, please stop it first before starting (or use restart). If this is not Solr, then please choose a different port using -p PORT
```

2. サーバの停止

Solr を停止する際には、停止する Solr ポート番号を指定します。

-p %PORT% の代わりに **-all** と指定することですべての Solr を停止することもできます。

- Windowsの場合

```
C:\solr\bin> solr.cmd stop -p %PORT%
```

- Linuxの場合

```
/usr/local/solr/bin/solr stop -p %PORT%
```

Solr管理画面の確認

Solr管理画面へ接続し、「search」ボタンが正常に動作することを確認します。

- 以下のURLをWebブラウザのアドレスバーに入力して「Solr管理」画面が表示できることを確認します。

```
http://<HOST>:<PORT>/solr/#/
```

The screenshot displays the Solr Administration interface. On the left is a navigation menu with options like Dashboard, Logging, Core Admin, Java Properties, Thread Dump, and Core Selector. The main content area is divided into several sections:

- Instance:** Shows the start time as 'less than a minute ago'.
- Versions:** Lists installed versions: solr-spec (8.7.0), solr-impl (8.7.0), lucene-spec (8.7.0), and lucene-impl (8.7.0).
- JVM:** Shows runtime details: Oracle Corporation Java HotSpot(TM) 64-Bit Server VM 1.8.0_221 25.221.
- Processors:** Shows 12 processors.
- Args:** Lists various JVM arguments such as `-DSTOP.KEY=solrrocks`, `-Djava.io.tmpdir=D:\solr\server\tmp`, and `-XX:+AlwaysPreTouch`.
- System:** Displays resource usage: Physical Memory (55.5% at 8.80 GB of 15.87 GB), Swap Space (67.8% at 14.49 GB of 21.37 GB), and JVM-Memory (54.8% at 280.34 MB of 512.00 MB).
- Security:** Lists Authentication Plugin, Authorization Plugin, Current Username, and User Roles.

At the bottom, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax.

図 Solr管理画面

- 「Core Selector」を選択し、コアの概要画面を表示します。

```
http://<HOST>:<PORT>/solr/#/default/query
```

図 コアの概要画面

- 「query」リンクをクリックします。

図 クエリー画面

- q (The Query String) のテキストエリアに「*:*」と入力されていることを確認し、「Execute Query」ボタンをクリックします。
- 検索に成功した場合、ブラウザに以下のようなXMLが表示されます。

```
<response>
  <lst name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">58</int>
    <lst name="params">
      <str name="hl.snippets">5</str>
      <str name="hl.useFastVectorHighlighter">true</str>
      <str name="echoParams">all</str>
      <str name="sort">record_date desc</str>
      <str name="start">0</str>
      <str name="q">*:*</str>
      <str name="rows">10</str>
      <str name="indent">on</str>
      <str name="version">2.2</str>
    </lst>
  </lst>
  <result name="response" numFound="0" start="0"/>
</response>
```

Solrの設定

Solr の設定項目について説明します。

項目

- コアの追加・削除
 - コアとは
 - intra-mart Accel Platformのマルチテナント環境における Solr の構成
 - コアの追加
 - コアの削除
- コアごとの設定
 - 基本設定 (solrconfig.xml)
 - スキーマの設定 (schema.xml)

コアの追加・削除

コアとは

コアはRDBのスキーマに相当し、コアごとにスキーマ定義やクエリの設定を持つことができます。

1つのSolrインスタンスの中に複数のコア（マルチコア）を設定することで、コアごとにインデックスが保持されます。

弊社から提供しているセットアップ資材のSolr環境は初期状態で **default** というコアが設定されています。

コアの設定ファイル（<schema.xml>,<solrconfig.xml>など）にはIM-ContentsSearch用の設定が最初から記載されています。

intra-mart Accel Platformのマルチテナント環境における Solr の構成

intra-mart Accel Platformでは、WARファイルによる複数テナントおよび、バーチャルテナントによる複数テナントで運用する場合には、テナントごとにコアを設定し、テナント単位でインデックスの管理を行います。

WARファイルによる複数テナントおよび、バーチャルテナントによる複数テナント環境での Solr の構成は以下の2つのパターンが考えられます。

1. テナントごとに Solr を構築する
2. 1つの Solr で、テナントごとにインデックスを管理する（コアを複数作成する）。

2の構成にする場合には、Solrコア構成の変更が必要です。

データの増加に応じてインデックスが増加することにより検索や最適化の処理速度が低下する場合がありますため、可能であれば、テナントごとにSolrのインスタンスを分けて（1の構成で）運用することを推奨します。

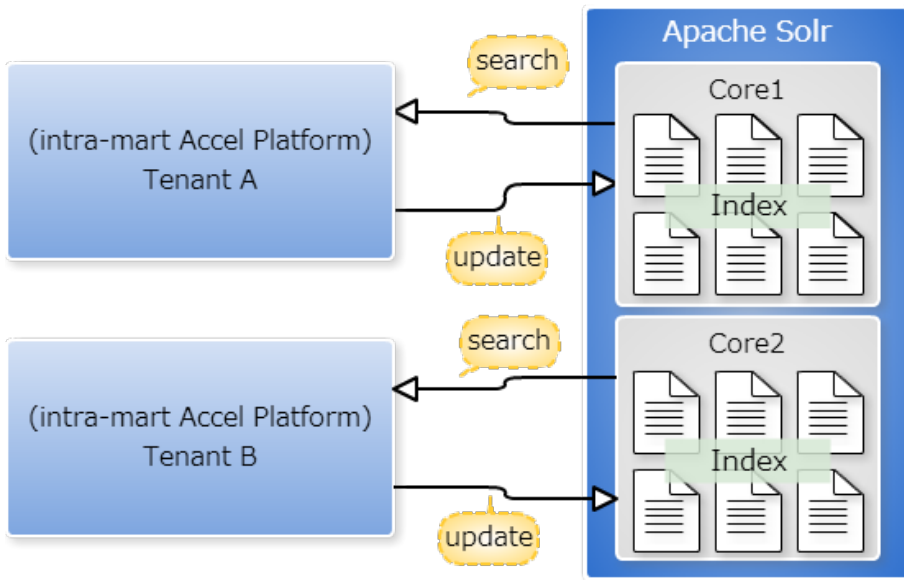


図 1つのSolrインスタンスに複数のコアを設定した場合のWARファイルによる複数テナントおよび、バーチャルテナントによる複数テナント環境の概念



注意

IM-ContentsSearchではコンテンツを識別するユニークキーに **ID** フィールドを指定しており、別々のテナントでコンテンツのIDが生成されるため、各テナントに同じSolrコアを設定することはできません。

コアの追加

コアを作成する例の手順を紹介します。
defaultのコアがあることを前提とします。

1. コマンドラインからコアを作成

コマンドラインから、以下のコマンドを実行します。

- Windowsの場合

```
C:\solr\bin> solr.cmd create -c %CORE_NAME% -d _iapdefault
```

- Linuxの場合

```
/usr/local/solr/bin/solr create -c %CORE_NAME% -d _iapdefault
```

2. 「Solr管理画面」の確認

ブラウザで「Solr管理画面」が表示できることを確認してください。

```
http://<HOST>:<PORT>/solr/#/%CORE_NAME%/core-overview
```

コアの削除

コアを削除する例の手順を紹介します。

1. コマンドラインからコアを作成

コマンドラインから、以下のコマンドを実行します。

- Windowsの場合

```
C:\solr\bin> solr.cmd delete -c %CORE_NAME%
```

- Linuxの場合

```
/usr/local/solr/bin/solr delete -c %CORE_NAME%
```

コアごとの設定

コアごとに下記の設定を行うことが可能です。

基本設定 (solrconfig.xml)

弊社から提供しているセットアップ資材に含まれる<solrconfig.xml>はIM-ContentsSearch用の設定が記述されています。

<solrconfig.xml>は %SOLR_HOME%/server/solr/%コア名%/conf ディレクトリ内に保存されています。

以下に記載する項目以外の内容は変更しないでください。

- インデックス保存ディレクトリ

インデックスが保存される場所です。初期状態では %SOLR_HOME%/server/solr/%コア名%/data ディレクトリに保存されます。

相対パスで指定した場合はアプリケーションサーバを起動したディレクトリによってカレントディレクトリが変わるため、絶対パスでの指定を推奨します。

```
<dataDir>${solr.data.dir:}</dataDir>
```

スキーマの設定 (schema.xml)

弊社から提供しているセットアップ資材に含まれる<schema.xml>はIM-ContentsSearch用のスキーマ定義が記述されています。

IM-ContentsSearchのためのフィールドやアナライザの定義が記載されているため、必要がない限り<schema.xml>の内容を変更しないでください。

<schema.xml>は %SOLR_HOME%/server/solr/%コア名%/conf ディレクトリ内に保存されています。

 コラム

IM-ContentsSearch が利用する Solr のフィールド（<schema.xml>で定義されているフィールド）は、IM-ContentsSearchの検索フィールド設定（solr-field-config）によって指定されています。

詳細は以下のドキュメントを参照してください。

- 「IM-ContentsSearch プログラミングガイド」 - 「フィールド」
- 「設定ファイルリファレンス」 - 「検索フィールド設定」

Solrの操作

IM-ContentsSearch を利用するにあたって必要となる Solr の運用上の手順などについて説明します。

項目

- [インデックスの削除](#)
- [Solrの移行](#)
- [Solrのバックアップ](#)

インデックスの削除

インデックスを削除する方法について説明します。

開発時や環境再構築時に一旦インデックスを削除したい場合などに実行してください。

管理画面から削除するクエリを指定して削除できます。いずれも、Solr が起動されている状態で実行する必要があります。

- コア名「default」内のインデックスをすべて削除する操作

1. 管理画面を開きます。

```
http://localhost:8983/solr/#/default/documents
```

2. 「Document Type」、「Document」を以下の通り入力します。

Document Type: Solr Command (raw XML or JSON)

Document: `{'delete': {'query': ':'}}`

The screenshot shows the Solr Admin interface. On the left is a navigation menu with options like Dashboard, Logging, Core Admin, Java Properties, Thread Dump, Overview, Analysis, Dataimport, Documents (selected), Files, Ping, Plugins / Stats, Query, Replication, Schema, and Segments info. The main content area shows the 'Documents' configuration for the 'default' core. It includes a 'Request-Handler (qt)' field with the value '/update', a 'Document Type' dropdown menu set to 'Solr Command (raw XML or JSON)', and a large text area for 'Document(s)' containing the JSON query: `{'delete': {'query': ':'}}`. Below this are input fields for 'Commit Within' (set to 1000) and 'Overwrite' (set to true), and a blue 'Submit Document' button. At the bottom of the page, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax.

3. 「Submit Document」リンクをクリックします。

i コラム

IM-ContentsSearch では、インデックスを削除するためのジョブを用意しています。
「ジョブ・ジョブネットリファレンス」 - 「削除クロール」 を参照してください。

Solrの移行

Solr に保存されている設定ファイルやインデックスの移行手順について説明します。
移行先のサーバでインデックスを再度作成する場合にはインデックスのコピーは必要ありません。
本書では intra-mart Accel Platform 2021 Spring(Bergamot) 以降の Solr を移行する手順を記載します。
過去のバージョンの移行手順はドキュメントアーカイブの管理者ガイドを参照してください。

i コラム

2020 Winter(Azalea) 以前の Solr からの移行はできません。
Solr を新規で構築し、インデックスを作成してください。

! 注意

Solr のデータの移行作業を行う前には、バックアップを行うことを推奨します。

- 作業は必ず Solr を停止させている状態で開始してください。
- 移行先の Solr は弊社提供のセットアップ資材で構築構築している環境とします。

%OLD_HOME%: 移行元の Solr インストールディレクトリ

%NEW_HOME%: 移行先の Solr インストールディレクトリ

1. solrディレクトリのコピー

<%OLD_HOME%/server/solr> ディレクトリを <%NEW_HOME%/server/solr> ディレクトリ配下にコピーします。

2. インデックス保存ディレクトリのコピー

インデックスが保存されているディレクトリをコピーしてください。

- <%OLD_HOME%/server/solr/%コア名%/conf/solrconfig.xml>の `dataDir` に指定しているディレクトリを変更していない場合

<%OLD_HOME%/server/solr/%コア名%> ディレクトリ配下に `data` ディレクトリがあるため、手順2でコピー済みです。

- <%OLD_HOME%/server/solr/%コア名%/conf/solrconfig.xml>の `dataDir` に指定しているディレクトリを変更している場合

複数のコアを設定している場合はそれぞれのディレクトリと <solrconfig.xml> の `dataDir`

3. その他環境設定

Solr に同梱されている Jetty の設定をデフォルト設定から変更している場合は、移行先の環境に合わせて変更してください。

4. Jettyの起動

Solr を起動し、正常に起動されることを確認してください。

5. Solr管理画面の確認

ブラウザでSolr管理画面が表示できることを確認してください。

```
http://<HOST>:<PORT>/solr/#/
```

Solrのバックアップ

1. solrディレクトリのバックアップ

%SOLR_HOME%/server/solr ディレクトリをバックアップしてください。

<solrconfig.xml> の `dataDir` の指定を変更していない場合は %SOLR_HOME% ディレクトリに設定ファイルとインデックスが保存されているため、下記の手順を行う必要はありません。

2. インデックスのバックアップ

インデックスが保存されているディレクトリをバックアップしてください。

複数のコアを設定している場合はそれぞれの <solrconfig.xml> の `dataDir` に指定されているディレクトリをそれぞれバックアップしてください。