doda Plus-CRM

Talendジョブ保守ガイド

第1.00版

2019年11月30日

株式会社　Cスタジオ

変更履歴

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版 | 変更日 | 変更箇所 | 変更内容 | 変更者 |
| 1.00 | 2019/11/30 |  | 初版リリース | Cスタジオ 竹内 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

目次

[はじめに 5](#_Toc25577477)

[Talendとは 5](#_Toc25577478)

[参考サイト 5](#_Toc25577479)

[開発環境の構築 6](#_Toc25577480)

[ハードウェア要件 6](#_Toc25577481)

[ソフトウェア要件 6](#_Toc25577482)

[JDKのインストール 6](#_Toc25577483)

[Talend Open Studioのインストール 6](#_Toc25577484)

[Talend Open Studioの起動 6](#_Toc25577485)

[プロジェクトのインポート 7](#_Toc25577486)

[プロジェクト概要 8](#_Toc25577487)

[プロジェクト構成 8](#_Toc25577488)

[ジョブ 8](#_Toc25577489)

[コンテキスト 8](#_Toc25577490)

[コード 8](#_Toc25577491)

[メタデータ 8](#_Toc25577492)

[ジョブ概要 9](#_Toc25577493)

[ジョブ共通 9](#_Toc25577494)

[初期処理 9](#_Toc25577495)

[後処理 9](#_Toc25577496)

[実行制御ジョブ 10](#_Toc25577497)

[スタートジョブ 10](#_Toc25577498)

[エンドジョブ 10](#_Toc25577499)

[クリアジョブ 10](#_Toc25577500)

[収集ジョブ 11](#_Toc25577501)

[ダウンロードジョブ 11](#_Toc25577502)

[収集親ジョブ 11](#_Toc25577503)

[収集サブジョブ 11](#_Toc25577504)

[連携ジョブ 12](#_Toc25577505)

[連携ジョブ 12](#_Toc25577506)

[ジョブの修正 13](#_Toc25577507)

[Talendジョブの修正 13](#_Toc25577508)

[Talendジョブのエクスポート 13](#_Toc25577509)

[ジョブのアップロード 13](#_Toc25577510)

[ジョブのデプロイ 14](#_Toc25577511)

[ジョブの実行 14](#_Toc25577512)

[ドキュメントの生成 14](#_Toc25577513)

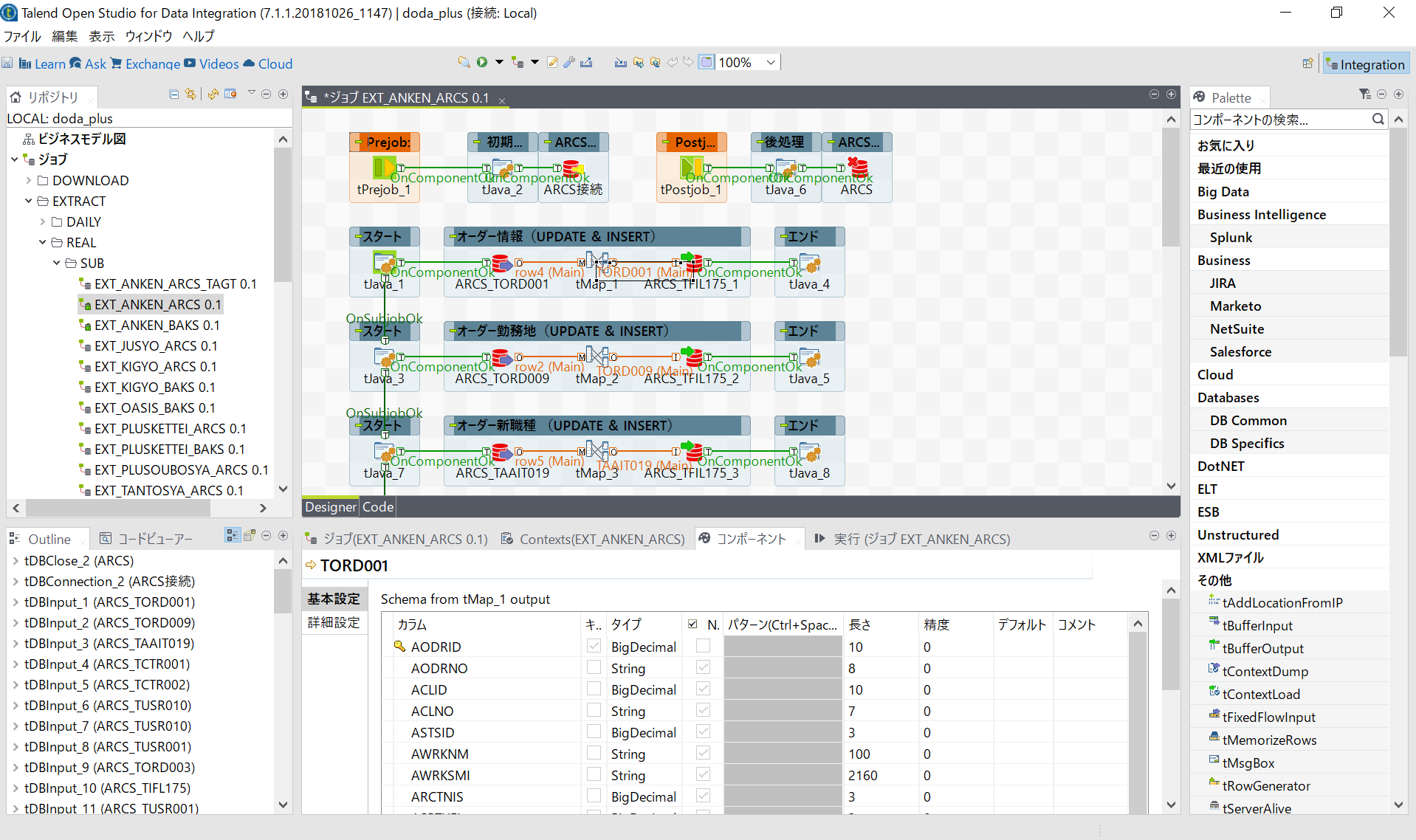
# はじめに

## Talendとは

**Talend**はフランスの[Talend社](https://jp.talend.com/)が提供するETLツールです。  
IDE（Eclipse）を拡張した「Studio」を使い、予め用意されているコンポーネントをデザイン画面上に配置し連結することで、GUIでETLジョブを作成することができます。  
「Studio」にはオープンソース版の「Talend Open Studio」と有償版の「Talend Studio」があり、対象となるデータタイプやボリュームによっていくつかの種類が用意されています。

本システムでは、オープンソース版の**Talend Open Studio Data Integration 7.1.1**を使用します。

「Studio」で作成したジョブはJavaのクラスが自動生成されJavaのアプリケーションとして実行されます。



Talend Open Studioの画面

## 参考サイト

Talend Open Studioの使用方法については、次のサイトを参考にしてください。

「Talend Data Integration Studioユーザーガイド」  
<https://help.talend.com/reader/RpIZs9~SFHdMENe49Qc4iw/beqbfgquVAOG5WHIO7bJJw>

# 開発環境の構築

## ハードウェア要件

Talend Open Studioに必要なハードウェア要件はこちらで確認してください。  
<https://help.talend.com/reader/otjx9J_SJHX4M48Yygs59g/g27_Jo7DkOXGnTc0TgiMzA>

## ソフトウェア要件

Talend Open Studioに必要なソフトウェア要件はこちらで確認してください。  
<https://help.talend.com/reader/otjx9J_SJHX4M48Yygs59g/h3tpFrNJEdIfMi5OsdGYow>

## JDKのインストール

Talendジョブを開発するには Java の開発環境 Oracle JDK ８（JREは不可）が必要です。  
事前にOracle JDK 1.8.xをインストールしてください。

## Talend Open Studioのインストール

Talend Open Studioのインストールファイル[**TOS\_DI-20181026\_1147-V7.1.1.zip**](https://jp.talend.com/products/data-integration-manuals-release-notes/)をダウンロードし、任意の場所に展開します。

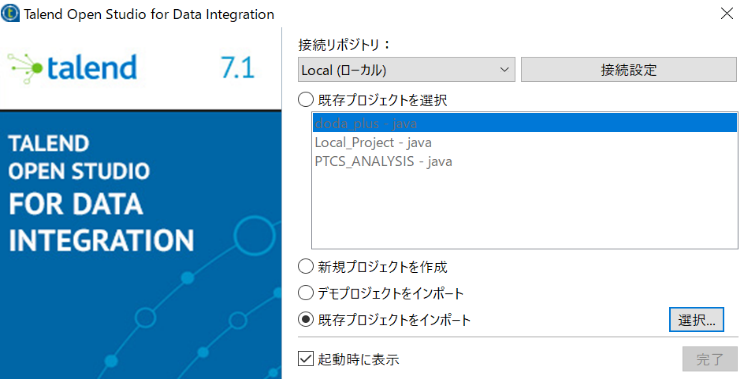
## Talend Open Studioの起動

Talend Open Studioを起動します。インストールしたフォルダ直下にある次のファイルを起動します。

|  |  |
| --- | --- |
| 環境 | 実行ファイル |
| Windows10 | TOS\_DI-win-x86\_64.exe |
| Linux | TOS\_DI-linux-gtk-x86.sh |
| Mac OS | TOS\_DI-macosx-cocoa.appをインストールします。 |

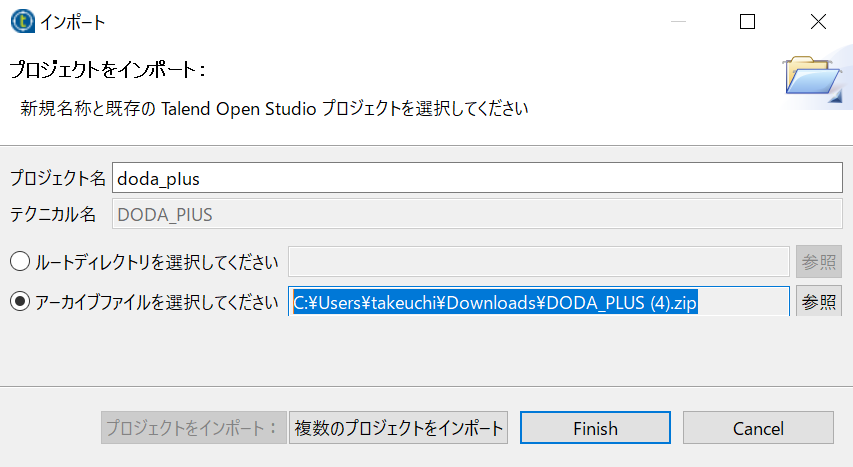
## プロジェクトのインポート

プロジェクトを選択するダイアログが表示されます。



「既存のプロジェクトをインポート」を選択して  
「完了」をクリックします。

「プロジェクト名」には**doda\_plus**を指定します。  
「アーカイブファイルを選択してください」を選択し  
　ZIP圧縮されたプロジェクトファイルを指定します。  
「Finish」をクリックします。



以上で開発環境が構築できました。次回からはプロジェクト選択ダイアログで「既存のプロジェクトを選択」から「**doda\_plus**」を選択して「完了」をクリックします。

環境により、追加のライブラリ（Jarファイル）を要求されることがあります。指示に従ってライブラリを追加してください。

# プロジェクト概要

## プロジェクト構成

doda\_plusプロジェクトは次のように構成されています。

### ジョブ

ジョブはシェルから実行される親ジョブと親ジョブから呼び出されるサブジョブに分かれます。  
doda\_plusプロジェクトでは次のようにフォルダで分けられています。

|  |  |
| --- | --- |
| フォルダ | 説明 |
| DOWNLOAD | ファイルサーバからファイルをダウンロードする親ジョブ群です |
| DOWNLOAD/SUB | スカウト配信ファイルをダウンロードするサブジョブ群です |
| EXTRACT | ジョブの初期処理、後処理を行う親ジョブ群です |
| EXTRACT/DAILY | 日次収集処理の親ジョブ群です |
| EXTRACT/DAILY/SUB | 日次収集処理をデータソース毎に行うサブジョブ群です |
| EXTRACT/REAL | ニアリアル収集処理の親ジョブ群です |
| EXTRACT/REAL/SUB | ニアリアル収集処理をデータソース毎に行うサブジョブ群です |
| UPLOAD | Salesforceに連携する親ジョブ群です |

### コンテキスト

ジョブの実行時に必要なパラメータを定義します。

コンテキストの内容は実行時にパラメータで上書きすることができます。

### コード

共通処理や関数をJavaで記述します。

### メタデータ

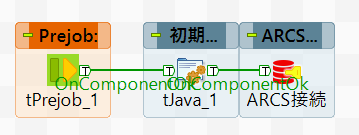
データのスキーマ構造を定義します。  
実際のデータベースに接続してテーブルのスキーマを取得することもできます。

# ジョブ概要

各ジョブの主な処理パターンについて説明します。

## ジョブ共通

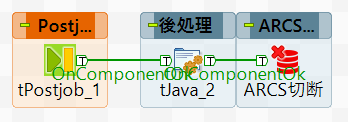
### 初期処理



各ジョブで初期処理を行います。

1. 設定ファイル（talend.config）を読み込みます。
2. データベースに接続します。

### 後処理

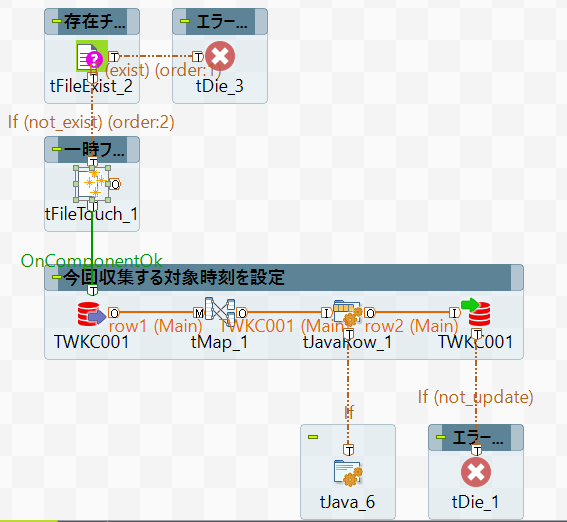


各ジョブで後処理を行います。

1. 処理時間をログに出力します。
2. データベースの接続を解除します。

## 実行制御ジョブ

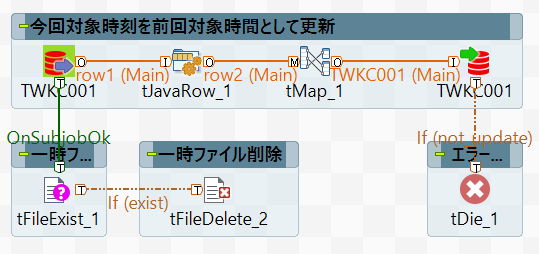
### スタートジョブ



日次やニアリアルのシェルの最初に実行されます。

1. 重複起動を避けるため管理ファイルの存在チェックをします。存在していた場合はエラー終了します。
2. 管理ファイルを作成します。
3. 管理テーブルの収集期間（FROM）、収集期間（TO）を設定します。

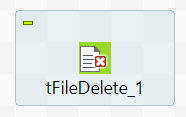
### エンドジョブ



日次やニアリアルのシェルで一連のジョブが正常終了した後で実行されます。

1. 管理テーブルの前回収集期間（FROM）、前回収集期間（TO）を設定します。
2. 管理ファイルを削除します。

### クリアジョブ

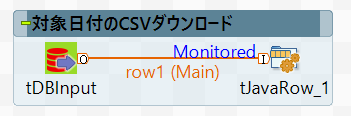


日次やニアリアルのシェルで一定の条件下で実行されます。

1. 管理ファイルを強制的に削除します。

## 収集ジョブ

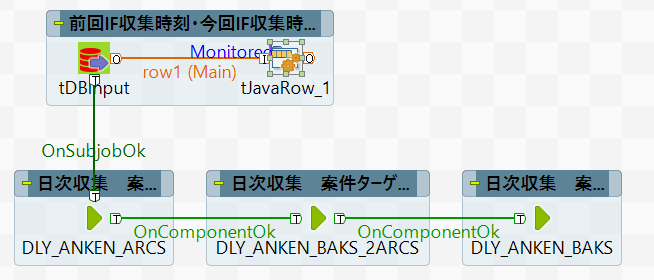
### ダウンロードジョブ



日次シェルから実行されます。

1. 管理テーブル（TWKC001）を照会してダウンロード対象のファイル名を生成します。
2. 共通関数getFSFileを実行してファイルをダウンロードします。

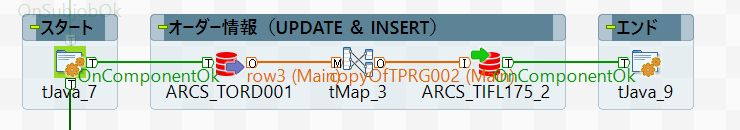
### 収集親ジョブ



日次やニアリアルのシェルから実行されます。

1. 管理テーブル（TWKC001）を照会して収集期間（FROM~TO）を取得します。
2. 収集サブジョブを実行します。

### 収集サブジョブ

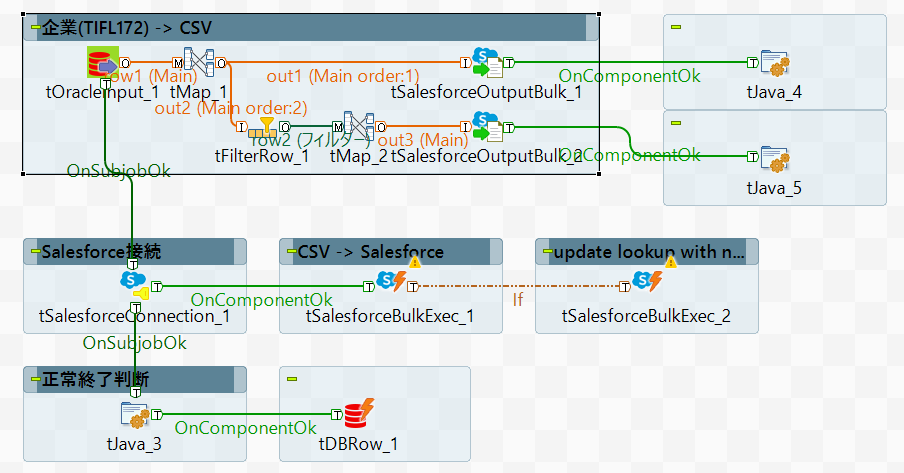


収集親ジョブから実行されます。

1. 収集対象のテーブル名をログに出力します。
2. 収集対象のテーブルからデータを照会します。
3. 照会したデータをIFテーブルのレコードにマッピングしIFテーブルに挿入または更新します。
4. 収集対象のテーブル名、レコード数をログに出力します。

## 連携ジョブ

### 連携ジョブ



日次やニアリアルのシェルから実行されます。

1. IFテーブルから連携対象のデータを照会します。
2. SalesforceのオブジェクトにマッピングしCSVファイルを作成します。
3. 連携対象のレコード数をログに出力します。
4. Salesforceのオブジェクトのルックアップ項目にNULL値が登録できないため、ルックアップ項目にNULLが含まれるレコードはリレーションシップのフィールド名ではなくオブジェクトのフィールド名で更新するためのCSVファイルを作成します。
5. NULLデータの連携対象のレコード数をログに出力します。
6. Salesforceに接続します。
7. ２で作成したCSVをSalesforceのバルクAPIでアップロードします。
8. ４で作成したCSVをSalesforceのバルクAPIでアップロードします。
9. アップロードの結果をログに出力します。
10. IFテーブルのレコードを連携済みに更新します。

# ジョブの修正

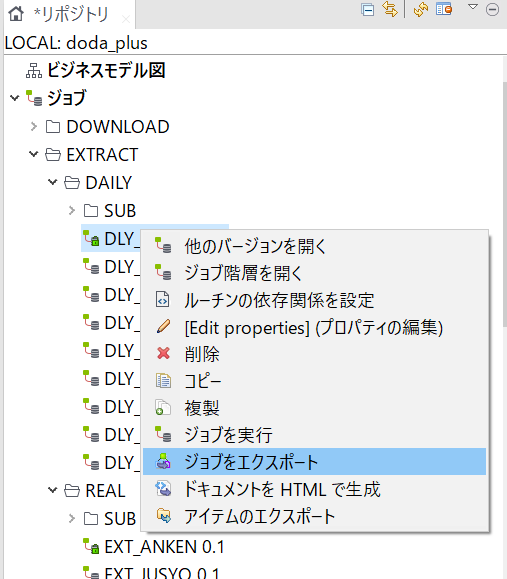
Talendジョブを修正する手順について説明します。

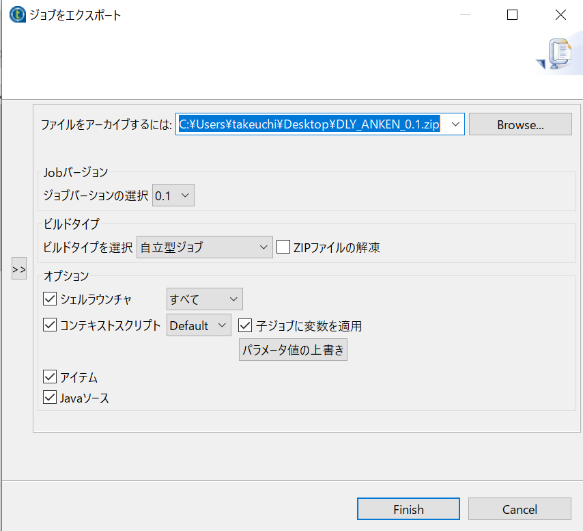
## Talendジョブの修正

Talend Open Studioを起動して対象のジョブを修正して保存します。

具体的なTalendの操作についてはTalend Studioユーザーガイドを参照して下さい。

## Talendジョブのエクスポート

リポジトリビューから対象の親ジョブを選択し、右クリックして「ジョブをエクスポート」を実行します。

出力先に任意の場所を指定します。

「子ジョブに変数を適用」にチェックします。

「Finish」をクリックしてジョブのアーカイブを作成します。

## ジョブのアップロード

作成したジョブのアーカイブをTalendサーバの次のディレクトリにアップロードします。

/opt/talend/DODA\_PLUS/*環境名*/archive

## ジョブのデプロイ

Talendサーバで次のコマンドを実行してジョブをデプロイします。

/opt/talend/DODA\_PLUS/*環境名*/tool/deploy.sh

ジョブが **/opt/talend/DODA\_PLUS/*環境名*/job/ジョブ名** ディレクトリに展開されます。

また、archiveディレクトリにはアーカイブファイルに日時を付けたファイル名で保存されます。

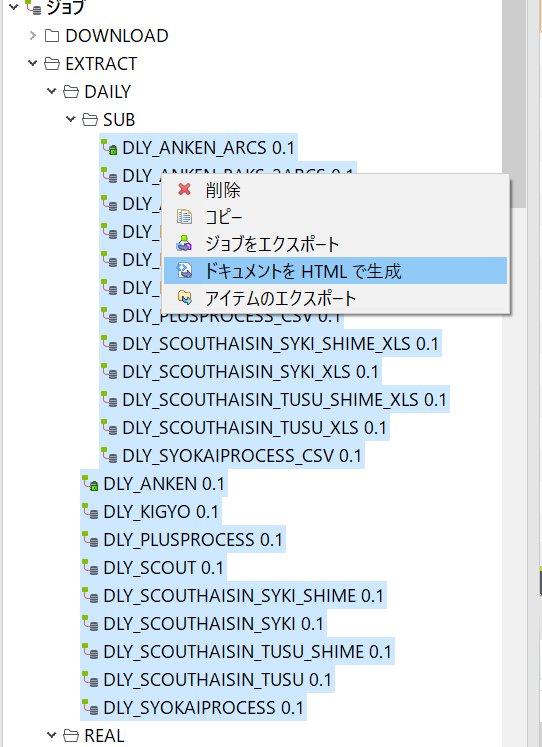
* 新しいジョブを追加した場合はdeploy.shの修正が必要になります。

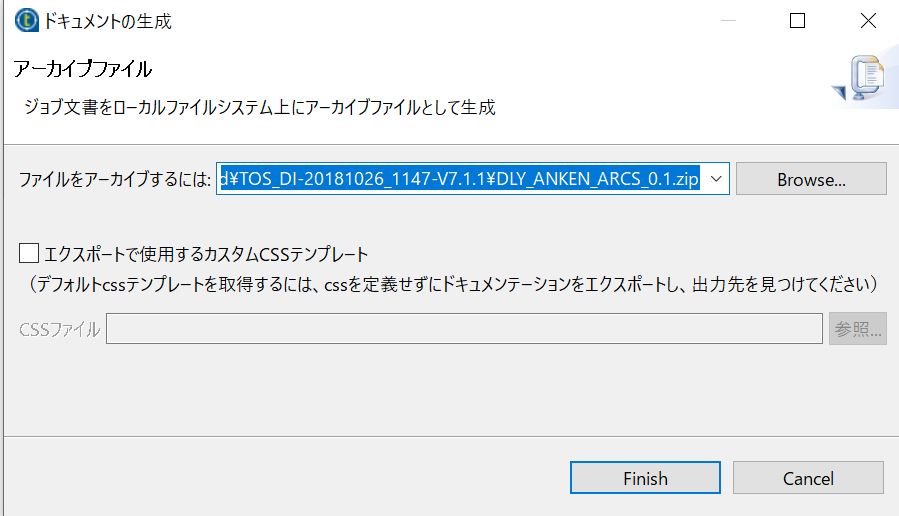
## ジョブの実行

単体でジョブを実行するにはTalendサーバで次のコマンドを実行します。

/opt/talend/DODA\_PLUS/*環境名*/job/ジョブ名\_run.sh

## ドキュメントの生成

リポジトリビューから対象のジョブを選択し、右クリックして「ドキュメントHTMLの生成」を実行します。

出力先に任意の場所を指定します。

「Finish」をクリックしてドキュメントのアーカイブを作成します。