

PDI ファーストステップガイド

2025 年 2 月発行

はじめに

このガイドは、PDI（Pentaho Data Integration）を初めて操作するときのガイドです。

Pentaho 評価版のインストール方法、データ変換の作成および実行方法、画面上の部品の配置方法など PDI の基本的な使い方を説明します。

■ 注意事項

ご利用にあたっては、OS、Java についての前提製品の必要知識を有する方を対象としています。

本書内の画面などは、Pentaho ソフトウェア Ver10.2 を使用しています。他バージョンでは、本書の内容と異なる挙動を示すおそれがあることにご注意ください。

製品の改良により予告なく記載されている仕様が変更になることがあります。

■ 輸出時の注意

本製品を輸出される場合には、外国為替及び外国貿易法の規制並びに米国輸出管理規則など外国の輸出関連法規をご確認の上、必要な手続きをお取りください。

なお、不明な場合は、弊社担当営業にお問い合わせください。

■ 商標類

HITACHI は、株式会社 日立製作所の商標または登録商標です。

Excel、Microsoft および Windows は、マイクロソフト 企業グループの商標です。

Oracle®、Java、MySQL 及び NetSuite は、Oracle、その子会社及び関連会社の米国及びその他の国における登録商標です。

Pentaho は、Hitachi Vantara LLC の商標または登録商標です。

その他記載の会社名、製品名などは、それぞれの会社の商標もしくは登録商標です。

■ 対象読者

PDI を初めて操作する方を対象にしています。

■ 用語の定義

本書で使用している用語を次の表に示します。

用語	説明
PDI	Pentaho Data Integration（PDI）。 各種データソースからデータウェアハウス、データマートを作成する際の、データ抽出・変換・集約・ロード処理といった ETL 処理（Extract, Transformation & Loading）のプログラムを開発する構築支援ツール。
ステップ	データベースやファイルなどの入出力、突き合わせや集計処理などのデータ加工を実行するアイコン。    

用語	説明
ホップ	ステップ間を結ぶ線。 ホップは、データの流れを表し、処理を視覚的に理解することができます。 また、ホップはステップからステップヘフィールド定義情報などを受け渡すことを表します。 

■ KB（キロバイト）などの単位表記について

1KB（キロバイト）、1MB（メガバイト）、1GB（ギガバイト）、1TB（テラバイト）はそれぞれ $1,024$ バイト、 $1,024^2$ バイト、 $1,024^3$ バイト、 $1,024^4$ バイトです。

■ 著作権

All Rights Reserved. Copyright (C) 2025, Hitachi, Ltd.

目次

はじめに	1
1. Pentaho 評価版インストール	4
1.1 事前準備	4
1.1.1 実行ユーザの登録	4
1.2 評価版インストール	5
1.3 インストール後の設定	19
1.3.1 Java のパスの設定	19
1.3.2 開発用コンポーネントの設定	19
2. PDI を使用する	21
2.1 PDI の紹介	21
2.2 PDI (Spoon) を起動する	21
2.2.1 PDI の画面構成	22
2.3 単純なデータ変換を作成する	22
2.3.1 新しいデータ変換定義画面を開く	22
2.3.2 CSV ファイルを読み込む設定を追加する	22
2.3.3 Excel ファイルを出力する設定を追加する	25
2.3.4 作成したデータ変換を保存する	29
2.4 データ変換を実行する	29
3. 実行後のデータを加工する	32
3.1 実行結果のデータの中身を確認する	32
3.1.1 各ステップの実行結果を確認する	32
3.1.2 データをグラフ化する	33
3.2 不要なデータを除外する	37
3.3 データ変換を再実行する	40
4. 付録	41
4.1 Pentaho Server の起動・停止方法	41
4.1.1 起動	41
4.1.2 停止	42

1. Pentaho 評価版インストール

この章では、Pentaho ソフトウェア Ver10.2 のサーバ・ワークステーション機能を Microsoft Windows にインストールする手順/設定手順について示します。

注意事項

- この章で示す Pentaho ソフトウェア Ver10.2 評価版のインストール手順は、インストールをスムーズに実施することを目的としています。記載内容は日立での動作確認結果などの情報を基にした独自内容になっています。
- この章で示すインストール手順は、あくまでも動作させるための最低限の手順であり、これらの手順によって、Pentaho ソフトウェアの動作を保証するものではありません。機能の必要性に応じて、設定や構成を変更する必要があります。
- 記載内容に関するご質問、ご相談などは、受け付けておりません。

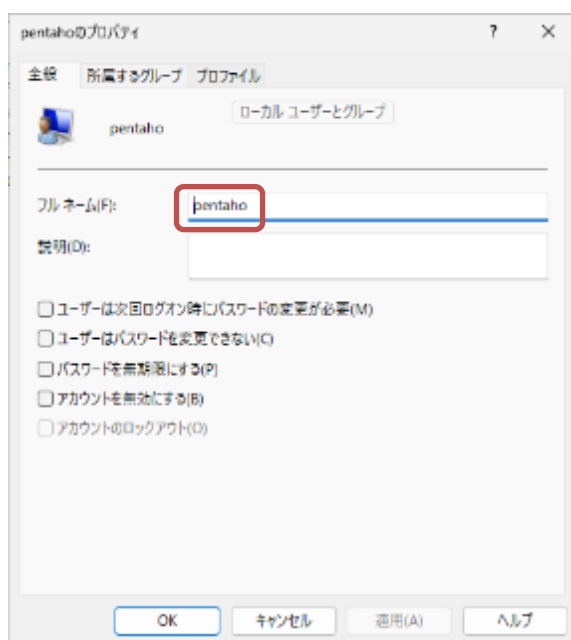
1.1 事前準備

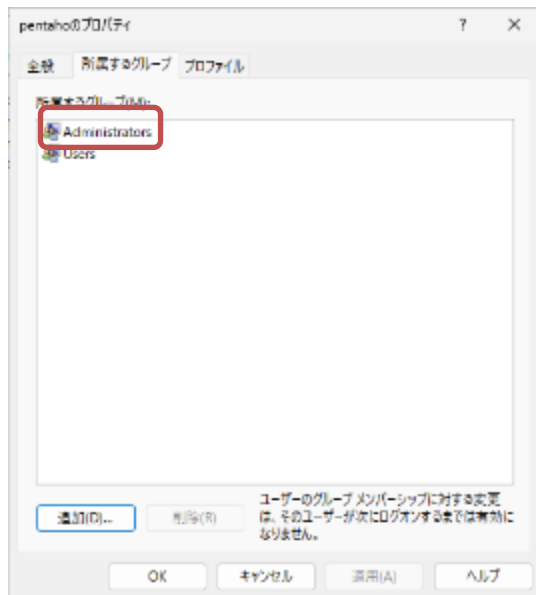
1.1.1 実行ユーザの登録

Pentaho ソフトウェアのインストールの実行ユーザ（ユーザ名例：pentaho）を登録します。

ユーザの追加は、スタートメニューの「すべて」－「Windows ツール」－「コンピューターの管理」で表示される「コンピューターの管理」画面の「ローカルユーザーとグループ」から実施できます。また、同ユーザに管理者権限を付与してください。

ユーザ名「pentaho」が追加された状態のプロパティ画面は次のようになります。





1.2 評価版インストール

Pentaho ソフトウェアのインストールを行います。実行は「1.1.1 実行ユーザの登録」で作成したユーザが行います。

Pentaho ソフトウェアには開発用コンポーネントとサーバ用コンポーネントがあり、評価版ではこれらを同じ環境にインストールすることも、別々の環境にインストールすることもできます。

なお、商用版では開発用コンポーネントとサーバ用コンポーネントとでサポートする OS が異なるため、開発用コンポーネントとサーバ用コンポーネントは同一マシンにインストールできません。Pentaho ソフトウェアを評価する目的でのみインストールしてください。

開発用コンポーネントとサーバ用コンポーネントについて次に示します。

- 開発用コンポーネント
データ変換定義やレポートの開発・デバッグツールのことで、クライアントマシンにインストールして使用します。
データ変換をお試しで使用する場合であれば、開発用コンポーネントのみのインストールで問題ありません。
- サーバ用コンポーネント
サーバマシン上でデータ変換の実行やレポートの作成を行うためのコンポーネントです。スケジュール実行など Pentaho Server の機能を使用したい場合にインストールします。

(1) インストーラの配備

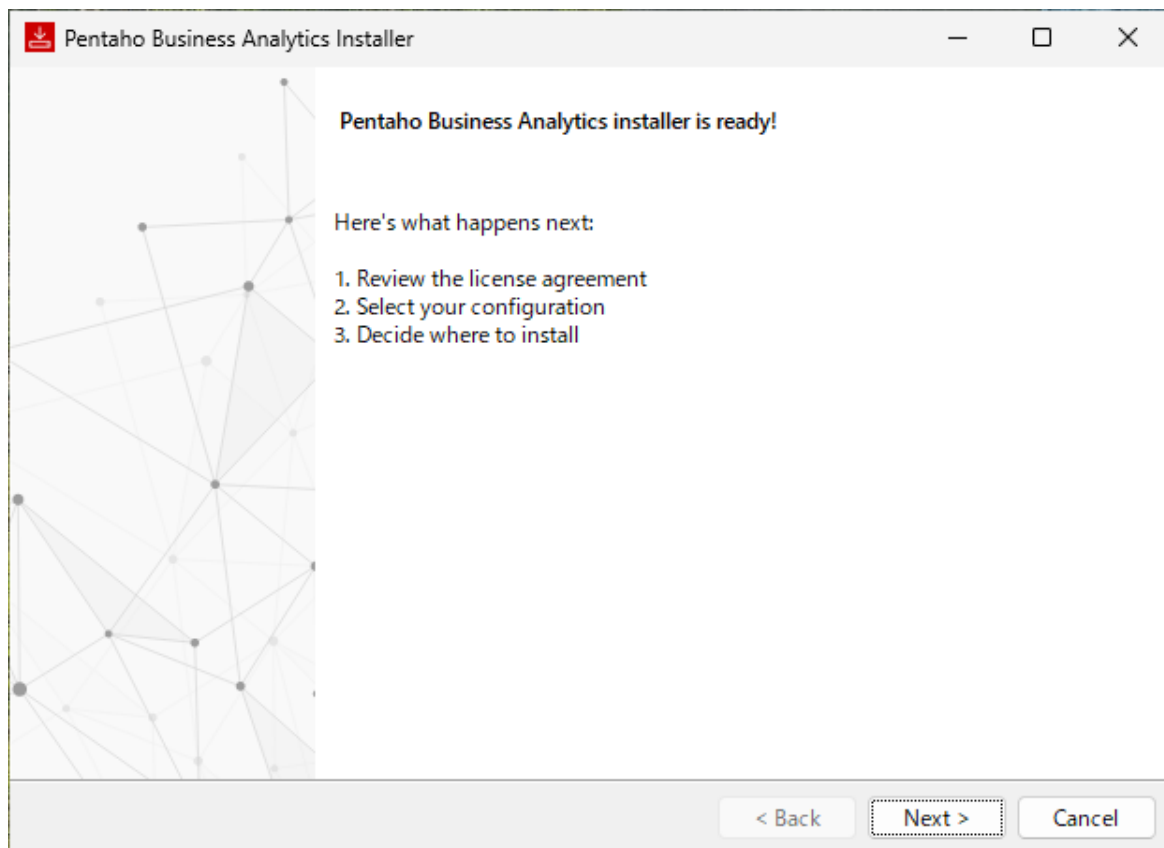
Pentaho ソフトウェアのインストーラをダウンロードし、Pentaho ソフトウェアを実行する環境に保存します。

PC > ローカルディスク (C:) > media

名前	更新日時
 pentaho-business-analytics-10.2.0.0-222-...	2025/02/05 9:56

(2) インストーラの起動

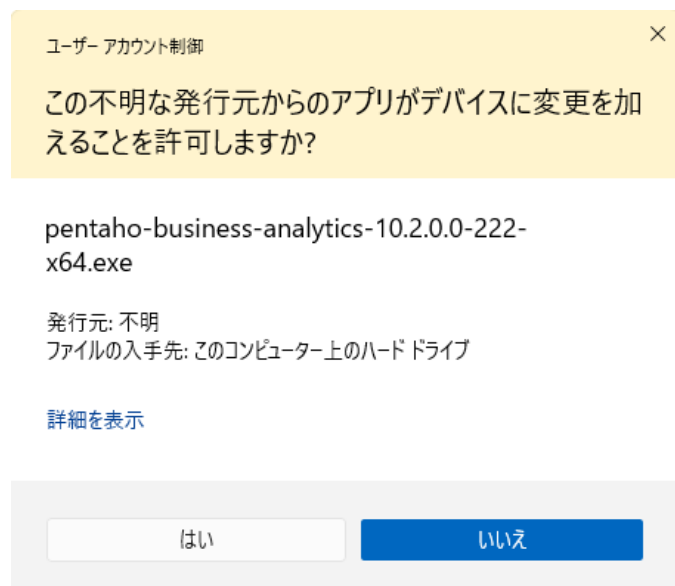
Pentaho ソフトウェアのインストーラをダブルクリックすると、Pentaho Business Analytics Installer が起動します。



[Next >] をクリックしてインストールを続行します。

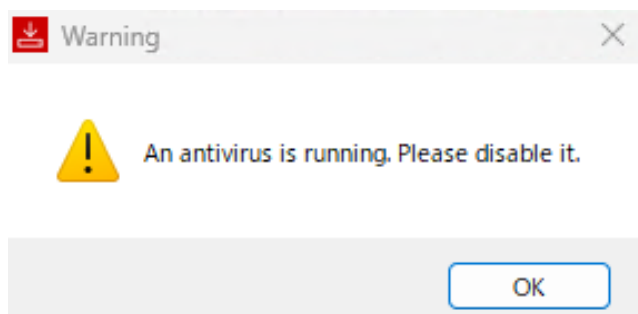
[注意]

次のようなユーザーアカウント制御のダイアログが表示されることがあります。[はい] をクリックしてインストールを続行してください。



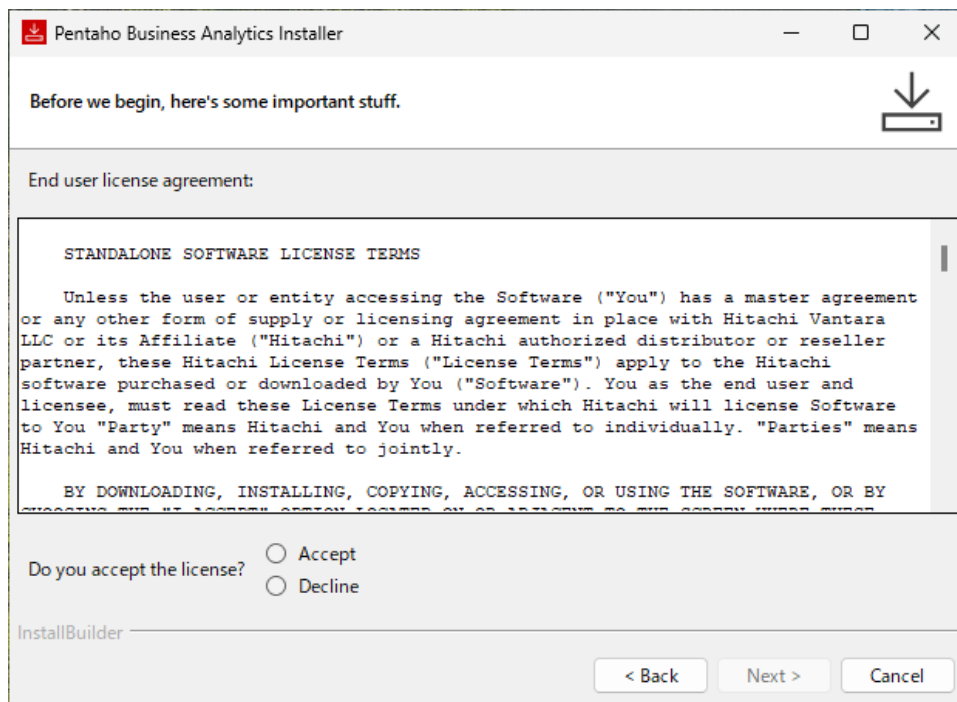
[注意]

次のような Antivirus を無効化するダイアログが表示されることがあります。Antivirus を無効化してからインストールを続行してください。



(3) 使用許諾の確認

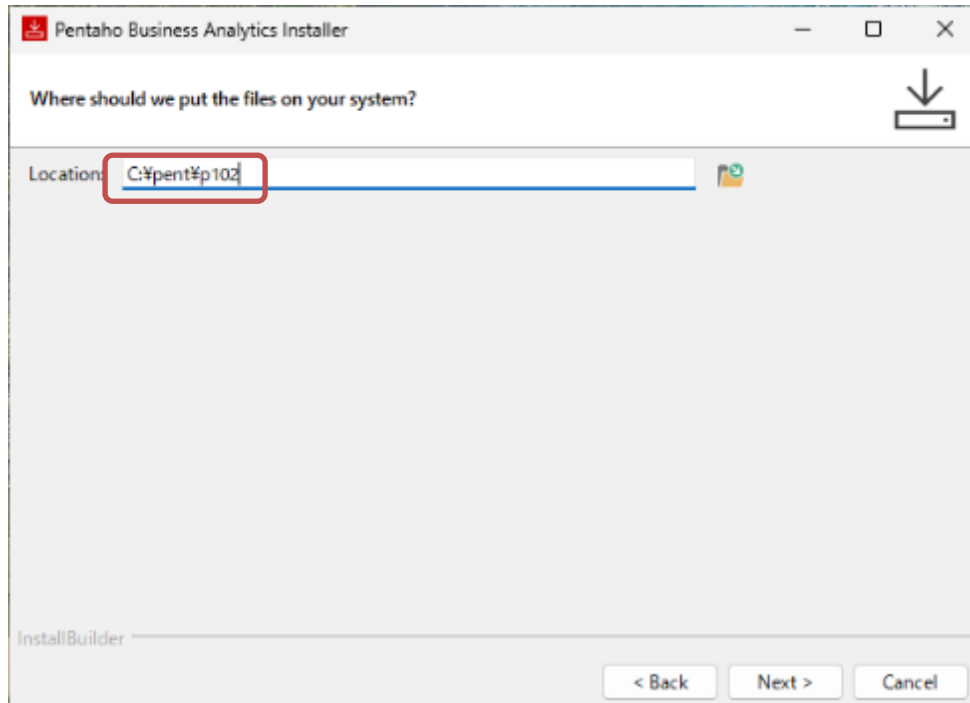
使用許諾を確認してください。使用許諾に合意する場合は [Accept] を選択し、[Next >] をクリックしてインストールを続行してください。



(4) インストールディレクトリの入力

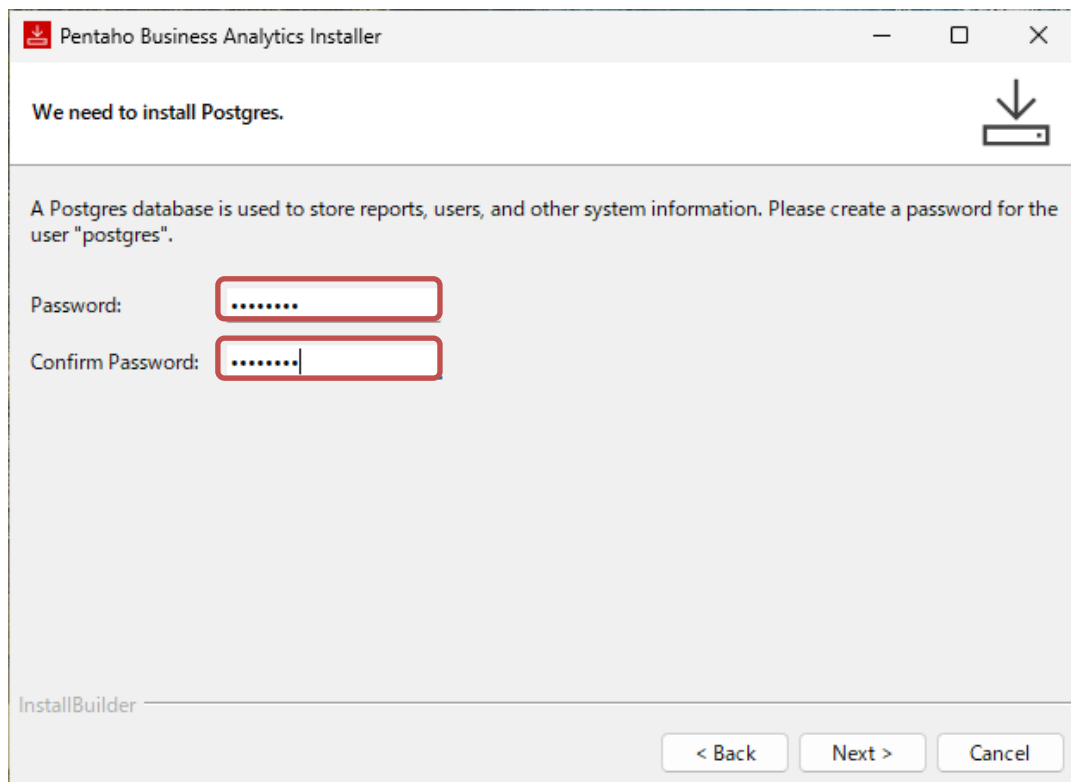
インストールディレクトリを入力します。

次の例では「C:\pent%p102」と入力しています。パス長はこれより長くならないよう指定してください。入力後、[Next >] をクリックしインストールを続行します。



(5) Repository 用 PostgreSQL の管理者パスワードの入力(サーバ用コンポーネントをインストールする場合のみ)

「Password:」と「Confirm Password:」にパスワードを入力し、「Next >」をクリックしインストールを続行します。



The screenshot shows the 'Pentaho Business Analytics Installer' window. The title bar includes the Pentaho logo and the text 'Pentaho Business Analytics Installer'. The main content area has a header 'We need to install Postgres.' with a download icon. Below this, a message states: 'A Postgres database is used to store reports, users, and other system information. Please create a password for the user "postgres".' There are two password input fields: 'Password:' and 'Confirm Password:'. Both fields contain masked characters (dots) and are highlighted with red rectangular boxes. At the bottom of the window, there are three buttons: '< Back', 'Next >', and 'Cancel'. The 'InstallBuilder' logo is visible in the bottom left corner.

(6) インストールオプションの選択

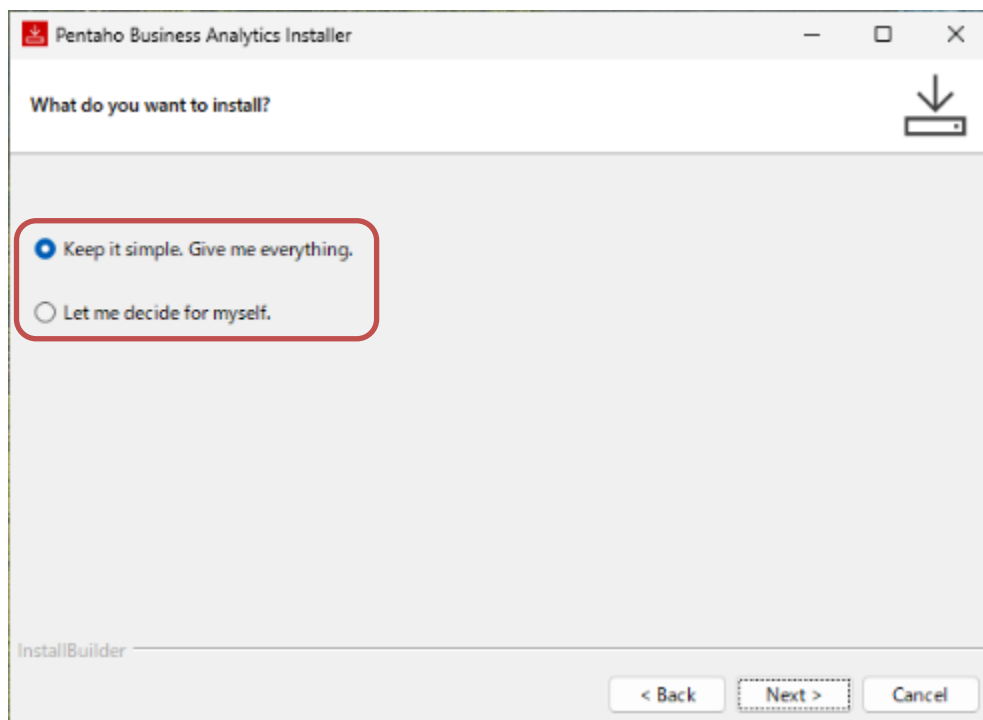
開発用コンポーネントとサーバ用コンポーネントとを同じ環境にインストールするか、別の環境にインストールするかを選択します。

⚠ 注意

商用版では開発用コンポーネントとサーバ用コンポーネントとでサポートする OS が異なります。したがって、開発用コンポーネントとサーバ用コンポーネントを同一マシンにインストールできません。Pentaho ソフトウェアを評価する目的でのみインストールしてください。

画面ではインストールオプションとして次のどちらかを選択してください。

- 開発用コンポーネントとサーバ用コンポーネントの両方を一つの環境にインストールしたい場合
「Keep it simple, Give me everything.」を選択してください。
- 開発用コンポーネントとサーバ用コンポーネントをそれぞれ別の環境にインストールしたい場合
「Let me decide for myself.」を選択してください。

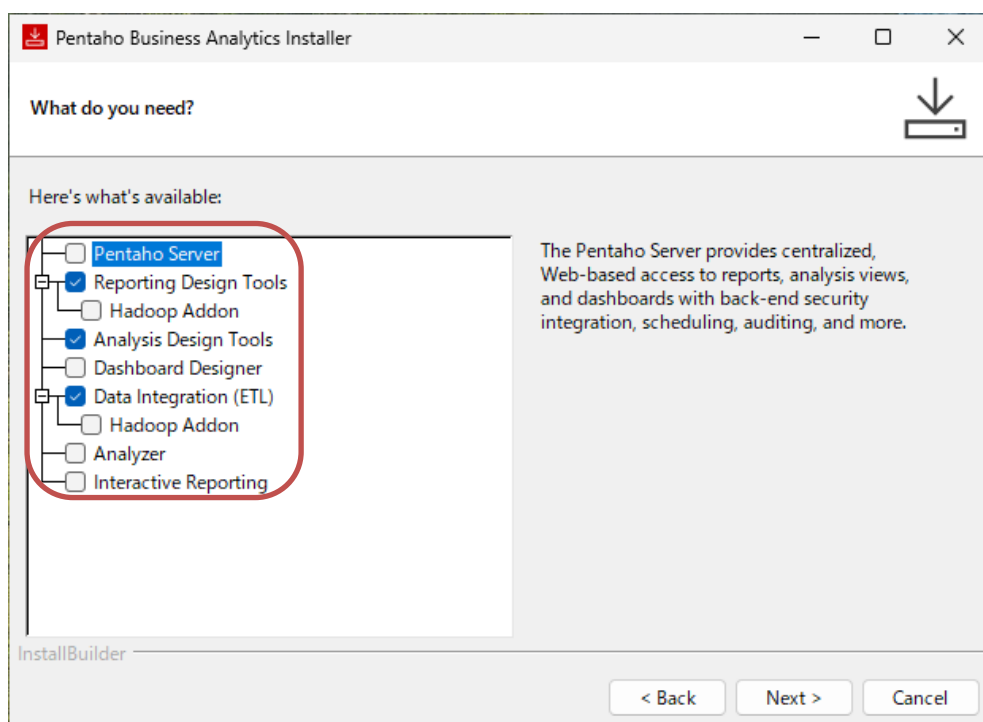


選択後、[Next >] をクリックしてインストールを続行します。

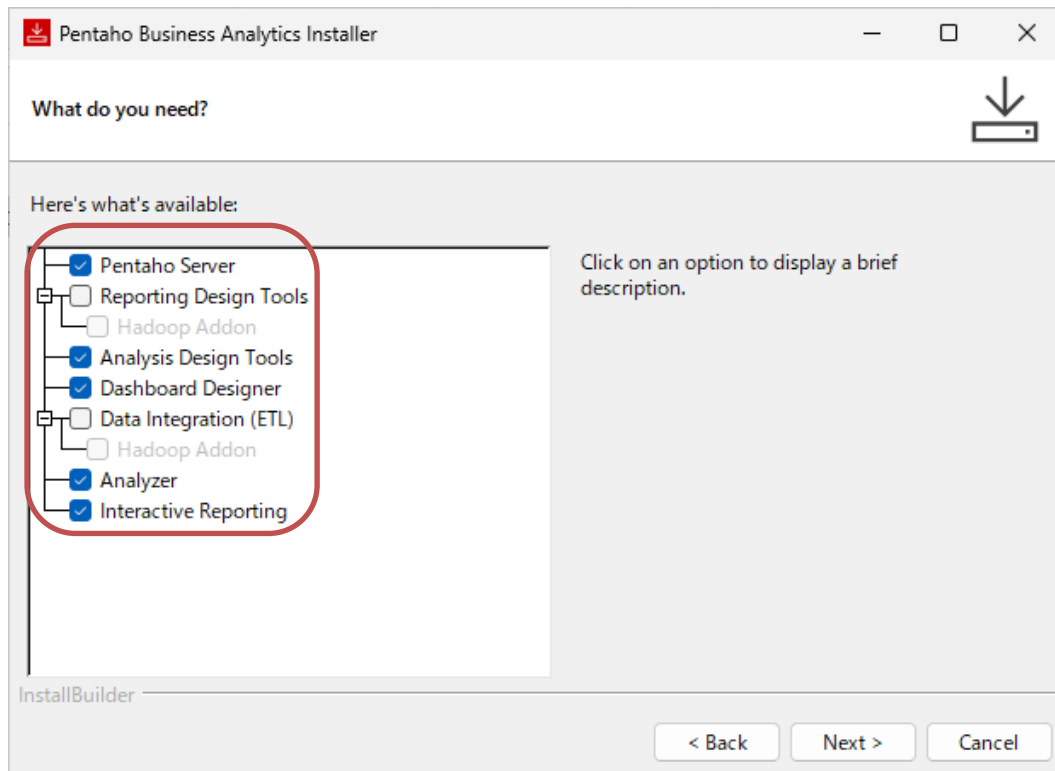
(7) インストールするコンポーネントの選択 (「1.2(6) インストールオプションの選択」で「Let me decide for myself.」を選択した場合)

開発用コンポーネントとサーバ用コンポーネントを別の環境にインストールする場合に、環境に応じて次のように選択してください。

- 開発用コンポーネントのみインストールする場合



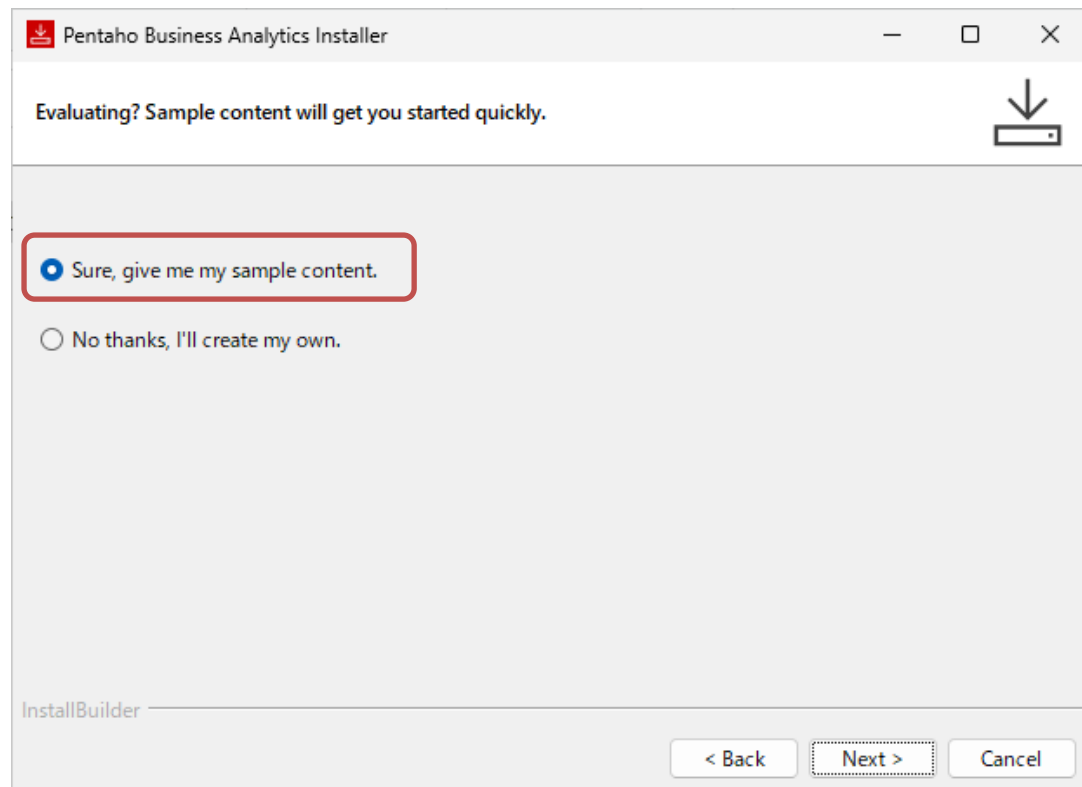
- サーバ用コンポーネントのみインストールする場合
次に示す画面のとおりに設定してください。



(8) サンプルコンテンツのインストール (「1.2(6) インストールオプションの選択」で「Let me decide for myself.」を選択しサーバ用コンポーネントをインストールする場合のみ)

サンプルコンテンツをインストールするかどうかを指定します。

このドキュメントではサンプルコンテンツを使用するため、サンプルコンテンツをインストールすることを示す [Sure, give me my sample content.] を選択し、[Next >] をクリックしてインストールを続行します。

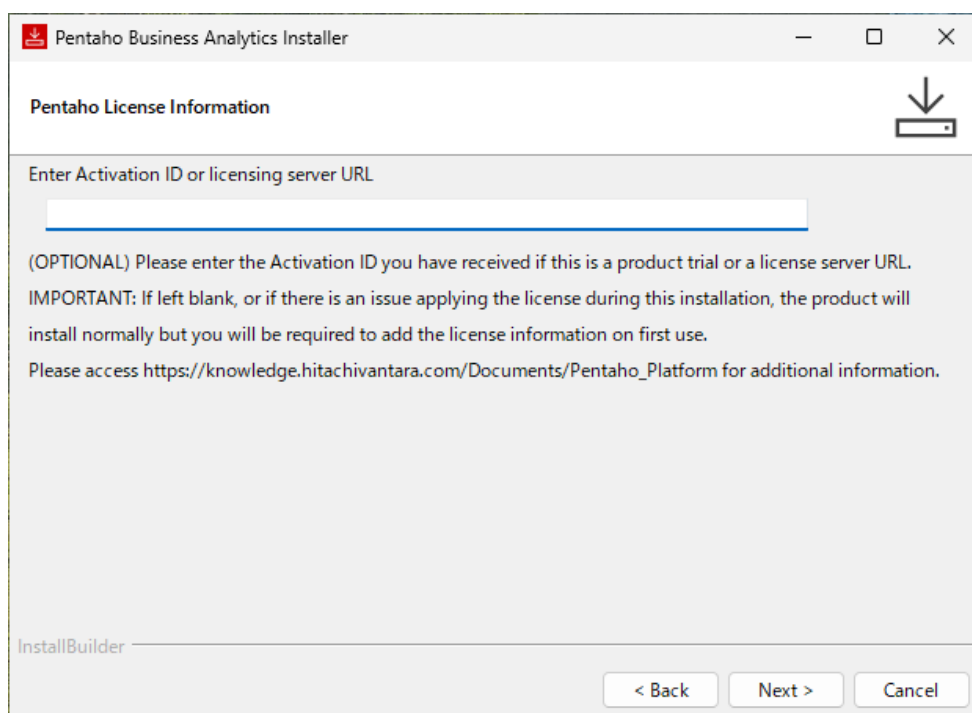


(9) 評価用ライセンスの有効化

評価用ライセンスの有効化の設定を行います。

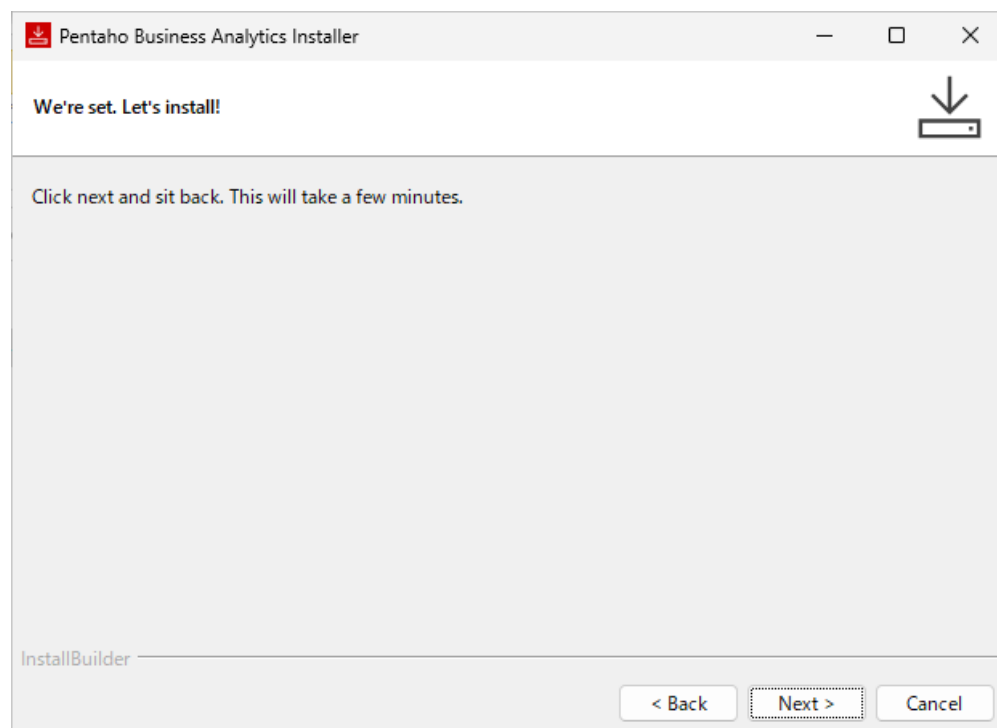
Web アクセスにプロキシが必要な環境かどうかに応じて以下のようにします。

- Web アクセスにプロキシを必要としない環境の場合
開発元から提供されたアクティベーション ID を入力し、[Next >] をクリックしてインストールを続行します。（この場合、ライセンスはインストール処理中に自動で有効化されます）
- Web アクセスにプロキシを必要とする環境の場合
ここでは何も入力せずに [Next >] をクリックしてインストールを続行します。（この場合、ライセンスは「1.2(12) 評価用ライセンスの手動有効化（「1.2(9) 評価用ライセンスの有効化」で「Web アクセスにプロキシを必要とする環境の場合」を選択しプロキシ設定をする場合のみ）」にて手動で有効化します）

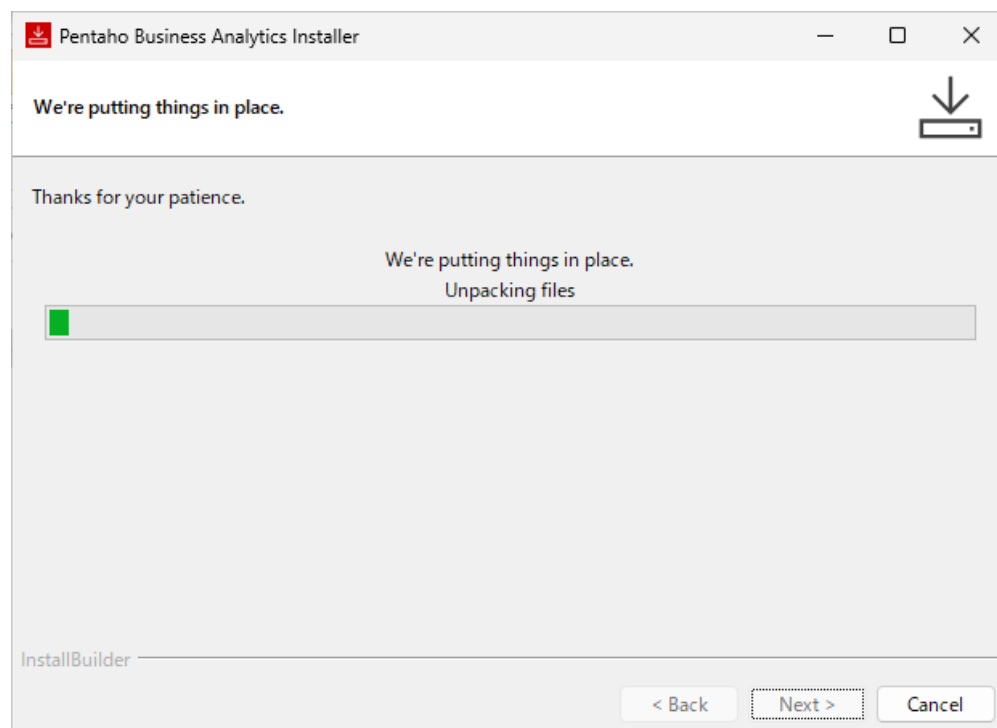


(10) インストール実行

インストールの実行準備ができましたので、[Next >] をクリックして、インストールを実行します。



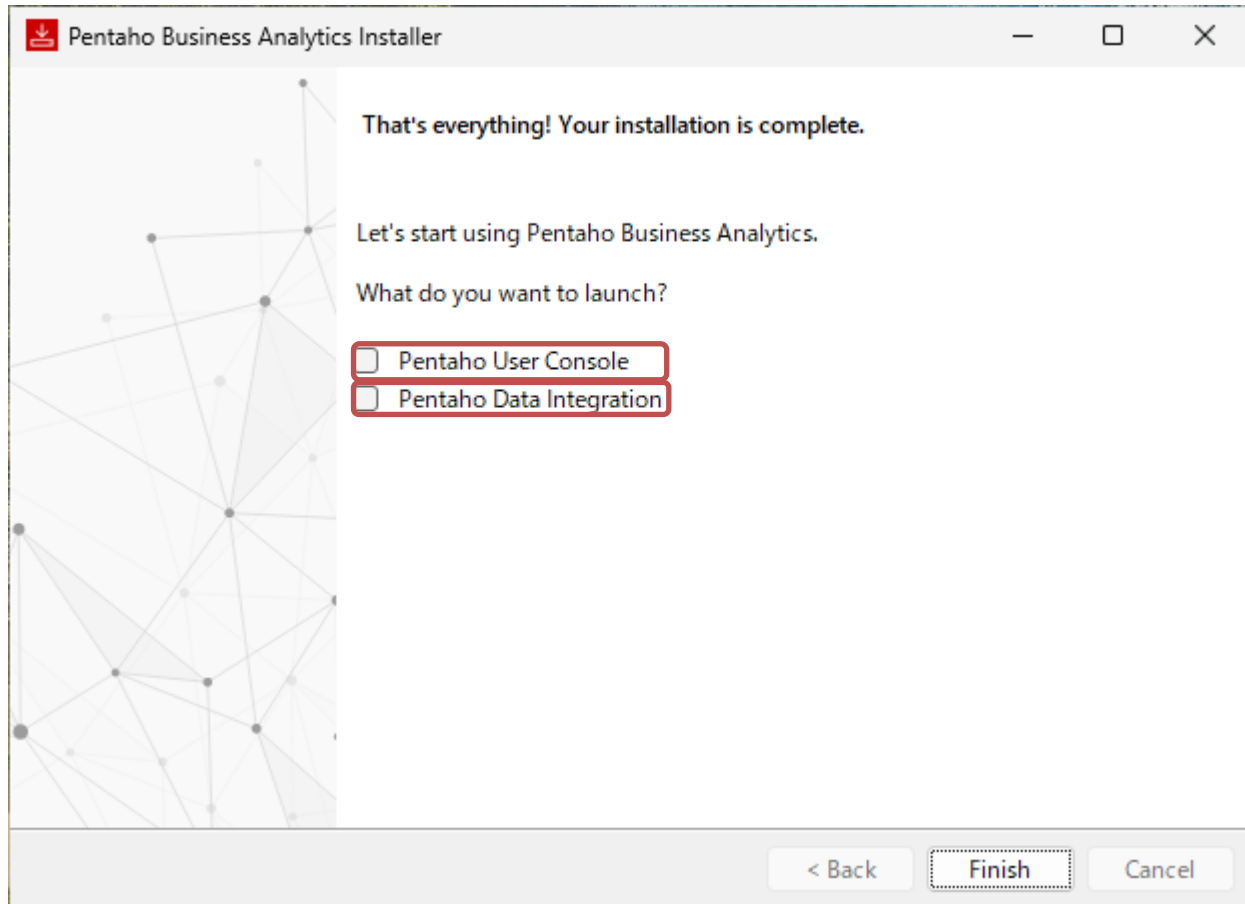
インストールの実行中は次に示す画面が表示されます。



(11)インストール完了

ここではインストール後に起動したいプログラムを選択しますが、インストール後にも日本語環境向けの設定が必要ですので、それぞれのチェックボックスをオフにします。

- [Pentaho User Console] : オフ（サーバ用コンポーネントをインストールした場合）
 - [Pentaho Data Integration] : オフ（開発用コンポーネントをインストールした場合）
- [Finish] をクリックし、インストールを完了します。



(12)評価用ライセンスの手動有効化（「1.2(9) 評価用ライセンスの有効化」で「Web アクセスにプロキシを必要とする環境の場合」を選択しプロキシ設定をする場合のみ）

評価用ライセンスを手動で有効化します。

開発用コンポーネントとサーバ用コンポーネントのそれぞれの設定手順がありますので、両方実施してください。

1. 開発用コンポーネントの評価用ライセンスの有効化
 - ① 以下のバッチファイルに対して、赤字で示した設定を追記します。
ファイル : C:\pentap102\license-installer\install-license.bat

箇所：69 行目付近

【認証なしプロキシの場合】

```
call "%_PENTAHO_JAVA%" %LICENSE_PARAMETER% -DCONSOLE_HOME=. -  
Dlog4j.configurationFile=log4j2.xml -Dhttps.proxyHost=ホスト名/IP アドレス -Dhttps.proxyPort=  
ポート番号 -cp %CLASSPATH% com.pentaho.dsc.elm.ELMUserInstaller %CMD_LINE_ARGS%
```

【認証ありプロキシの場合】

```
call "%_PENTAHO_JAVA%" %LICENSE_PARAMETER% -DCONSOLE_HOME=. -  
Dlog4j.configurationFile=log4j2.xml -Dhttps.proxyHost=ホスト名/IP アドレス -Dhttps.proxyPort=  
ポート番号 -Dhttps.proxyUser=ユーザ名 -Dhttps.proxyPassword=パスワード -cp %CLASSPATH%  
com.pentaho.dsc.elm.ELMUserInstaller %CMD_LINE_ARGS%
```

※上記のホスト名/IP アドレス、ポート番号、ユーザ名、パスワードは以下に置き換えます。

ホスト名/IP アドレス：プロキシサーバのホスト名または IP アドレス

ポート番号：プロキシサーバのポート番号

ユーザ名：プロキシサーバの認証ユーザ名

パスワード：プロキシサーバの認証パスワード

- ② ユーザー権限でコマンドプロンプトを起動します。
- ③ カレントディレクトリを変更してライセンスインストーラを実行します。 ※xxxxxxx の部分は開発元から提供されたアクティベーション ID に置き換えます。

```
C:\Users\pentaho>cd C:\pent\p102\license-installer  
C:\pent\p102\license-installer>install_license.bat xxxxxxxxx
```

- ④ 登録されたライセンス情報が表示されることを確認します。

Status:	Product Name:	Version:	Begin Date:	Exp. Date:
Valid	TRIAL	10.2	Tue Feb 04 00:00:00 JST 2025	Fri Mar 07 23:59:59 JST 2025

2. サーバ用コンポーネントの評価用ライセンスの有効化

- ① 「4.1.1 停止」に従って Pentaho Server を停止します。
- ② 以下のファイルを開きます。

C:\¥pent¥p102¥server¥pentaho-server¥tomcat¥bin¥pentahoserverw.exe

名前	作成日時	ファイルの種類	サイズ
makebase.sh	2024/02/14 17:15	SH ファイル	4 KB
pentahoserverw	2024/02/14 17:15	アプリケーション	126 KB
service	2025/02/05 16:31	Windows バッチファ...	11 KB
setclasspath	2024/02/14 17:15	Windows バッチファ...	4 KB

[注意]

次のようなユーザーアカウント制御のダイアログが表示されることがあります。[はい] をクリックしてインストールを続行してください。



- ③ 出現したダイアログの「Java」タブを開き、「Java Options:」欄に以下の赤字の設定を追加して、[OK] をクリックします。

-Dhttps.proxyHost=ホスト名/IP アドレス

-Dhttps.proxyPort=ポート番号

-Dhttps.proxyUser=ユーザ名

-Dhttps.proxyPassword=パスワード

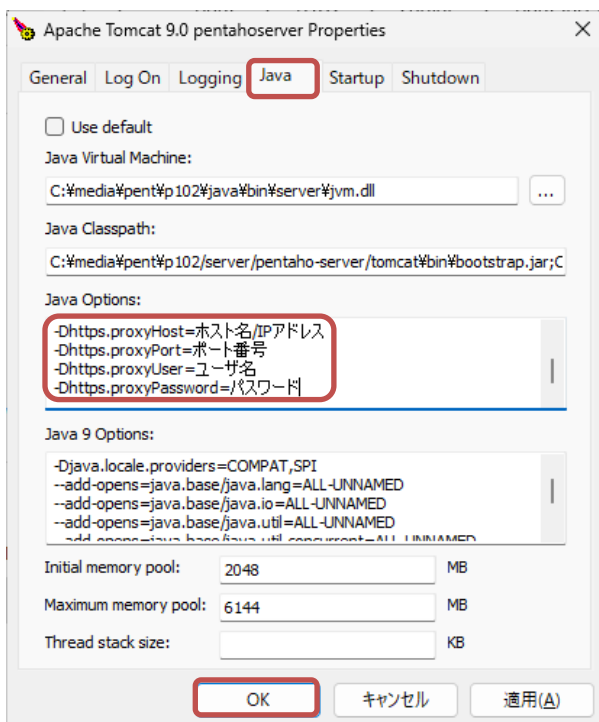
※上記のホスト名/IP アドレス、ポート番号、ユーザ名、パスワードは以下に置き換えます。

ホスト名/IP アドレス：プロキシサーバのホスト名または IP アドレス

ポート番号：プロキシサーバのポート番号

ユーザ名：プロキシサーバの認証ユーザ名

パスワード：プロキシサーバの認証パスワード

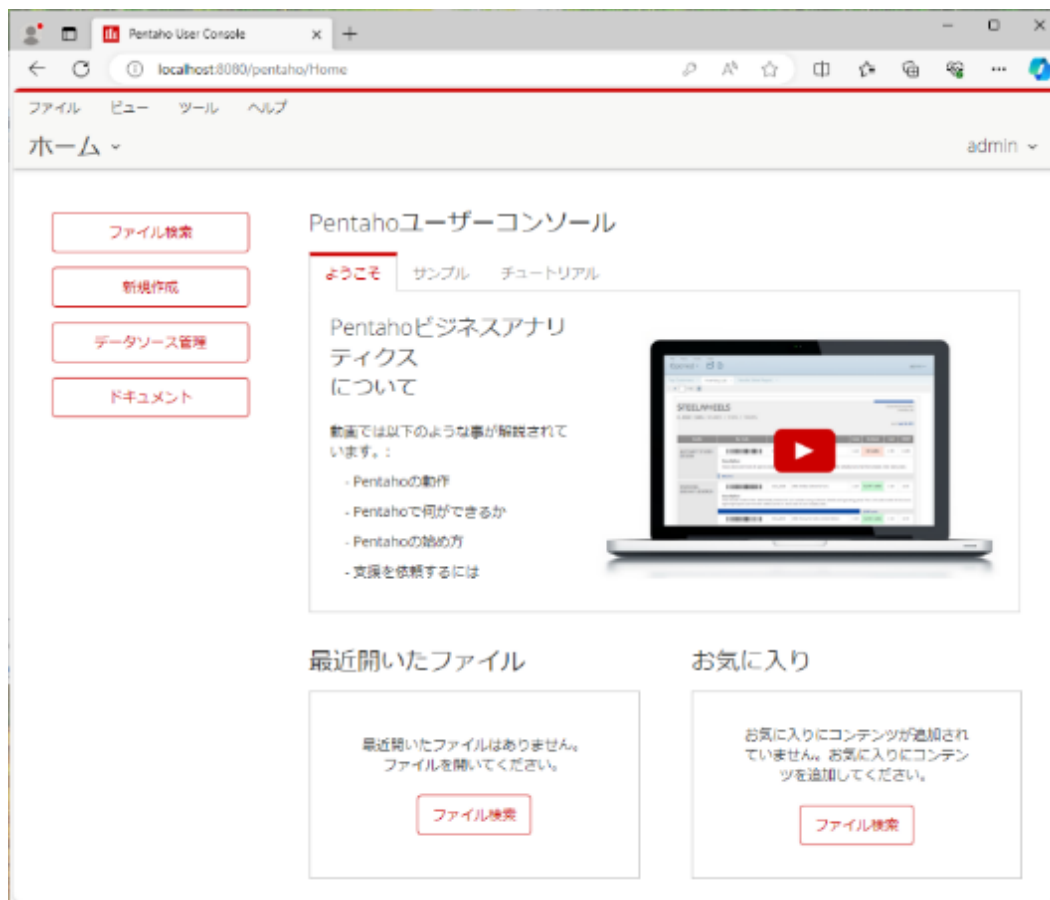


- ④ 「4.1.2 起動」に従って Pentaho Server を起動します。
- ⑤ Pentaho User Console に admin ユーザでログインします。

- Pentaho User Console の URL
<http://localhost:8080/pentaho>
- admin ユーザのユーザ名
admin
- admin ユーザのパスワード
password

⑥ ホーム画面が表示されることを確認します。

※正しく表示されない場合は、1. 開発用コンポーネントの評価用ライセンスの有効化を正しく実施できているか、プロキシ設定が正しいかなどをご確認ください。



1.3 インストール後の設定

この節では、Windows 環境にインストールした Pentaho ソフトウェアの起動に必要な設定の手順について示します。

[注意]

編集には、改行コードの LF（ラインフィード）が編集できるテキストエディタが必要です。ご準備をお願いします。

1.3.1 Java のパスの設定

システム環境変数 PENTAHO_JAVA_HOME に Pentaho ソフトウェア同梱の Java のパスを設定します。

- ① コマンドプロンプトを「管理者として実行」で起動します。
- ② 以下のコマンドを実行します。（※インストールディレクトリが「C:\pentaho」の場合）
>SETX /M PENTAHO_JAVA_HOME C:\pentaho\java

※この手順は Workstation に既存の Java が存在しない場合を前提とします。Workstation に既存の Java がある場合は、上記の設定に加え、システム環境変数 JAVA_HOME とシステム環境変数 PATH に設定済みの既存の Java のパスを、Pentaho ソフトウェア同梱の Java のパスに変更します。

1.3.2 開発用コンポーネントの設定

それぞれのバッチファイルに対して、赤字および下線で示した設定を追記します。修正後のバッチファイルの内容は、PC を再起動しなくても有効となります。

(1) Aggregation Designer

対象ファイル：C:\pentaho\design-tools\aggregation-designer\startaggregationdesigner.bat

49 行目付近

```
"%PENTAHO_JAVA%" -Dfile.encoding=UTF-8 %JAVA_ADD_OPENS% -jar "%dp0lib\pentaho-application-launcher-10.2.0.0-222.jar"
```

(2) Data Integration (Spoon)

対象ファイル：C:\pentaho\design-tools\data-integration\Spoon.bat

158 行目付近

```
if "%PENTAHO_DI_JAVA_OPTIONS%"==" set PENTAHO_DI_JAVA_OPTIONS="-Xms1024m" "-Xmx2048m" -Dfile.encoding=UTF-8"
```

(3) Metadata Editor

対象ファイル：C:\pentaho\design-tools\metadata-editor\metadata-editor.bat

100 行目付近

```
if "%PENTAHO_JAVA_OPTIONS%"==" set PENTAHO_JAVA_OPTIONS="-Xms1024m" "-Xmx2048m" -Dfile.encoding=UTF-8"
```

(4) Report Designer

対象ファイル：C:\pent\p102\design-tools\report-designer\report-designer.bat

64 行目付近

```
set OPT="-Dswing.useSystemFontSettings=false" "-Xms1024M" "-Xmx2048M" "-Dfile.encoding=UTF-8"
```

(5) Schema Workbench

対象ファイル：C:\pent\p102\design-tools\schema-workbench\workbench.bat

85 行目付近

```
"%_PENTAHO_JAVA%" -Xms1024m -Xmx2048m %JAVA_ADD_OPENS% %JAVA_LOCALE_COMPAT% -Dfile.encoding=UTF-8 -cp "%CP%" -Dlog4j.configurationFile=file:///%%ROOT%%\schemaWorkbench\log4j2.xml mondrian.gui.Workbench
```

2. PDI を使用する

PDI の起動方法や画面構成、簡単な使用方法を説明します。

2.1 PDI の紹介

Pentaho Data Integration (PDI) は、各種データソースからデータウェアハウス、データマートを作成する際の、データ抽出・変換・集約・ロード処理といった ETL 処理 (Extract, Transformation & Loading) のプログラムを開発する構築支援ツールです。PDI は、GUI によるプログラムレスでのビジュアル開発によって、高生産性、高品質なデータウェアハウスの構築を可能とします。

2.2 PDI (Spoon) を起動する

以降で説明するデータ変換や加工などを実施するには、PDI が提供する GUI (Spoon) を最初に起動する必要があります。

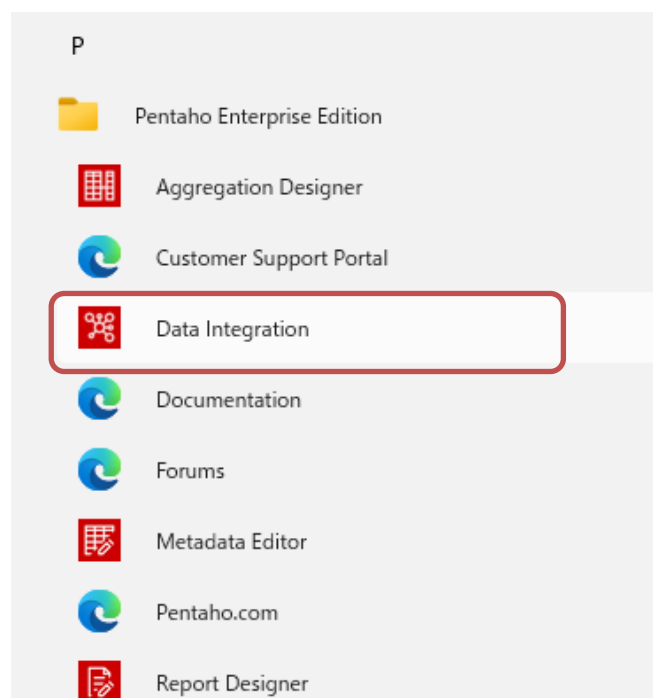
PDI を起動するユーザに管理者権限があるかどうかで起動方法が異なります。

なお、初回起動時は初期化处理のため 10 分程度掛かることがあります。

- 管理者権限のあるユーザの場合

起動するユーザに管理者権限がある場合はスタートメニューから起動します。

[スタートメニュー] — [すべて] — [Pentaho Enterprise Edition] — [Data Integration] で起動します。



- 一般ユーザの場合

C:\pentyp102\design-tools\data-integration\Spoon.bat をダブルクリックして、PDI を起動します。

2.2.1 PDI の画面構成

PDI の画面は次のとおりです。



2.3 単純なデータ変換を作成する

ここでは例として、サンプルコンテンツの CSV ファイルを使った、単純なデータ変換の作成手順を示します。CSV ファイルを読み込んでそのまま Excel ファイルに出力するデータ変換を作成します。

2.3.1 新しいデータ変換定義画面を開く

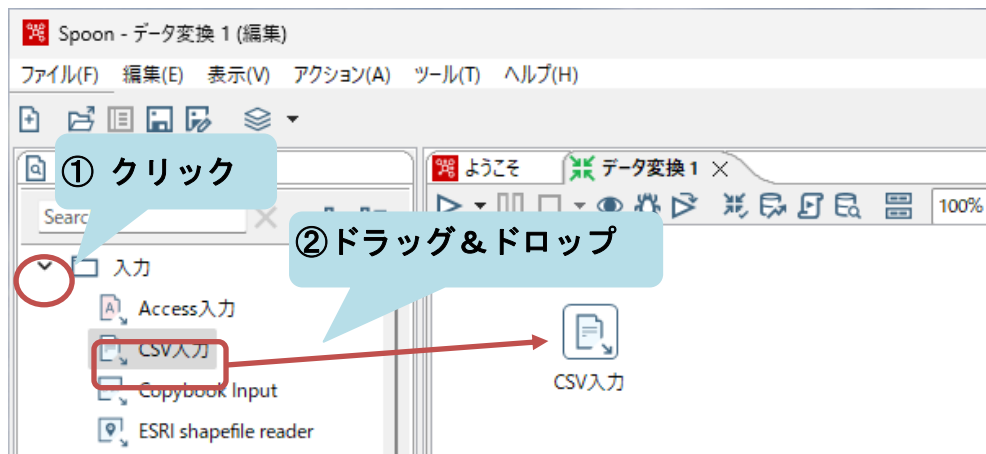
[ファイル] メニューー [新規] - [データ変換] を選択して、新しいデータ変換定義画面を開きます。



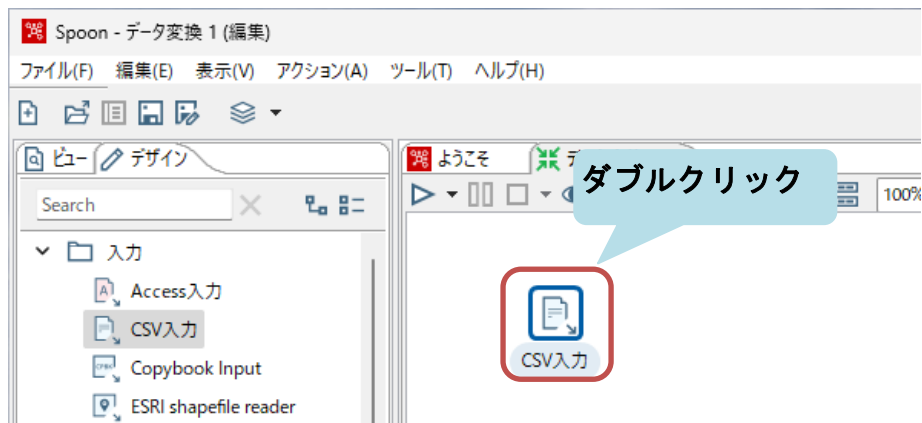
2.3.2 CSV ファイルを読み込む設定を追加する

PDI に含まれるサンプルの CSV ファイルを読み込むよう次の手順で設定します。

1. 部品一覧の [入力] - [CSV 入力] をキャンバスにドラッグ&ドロップします。



2. 「CSV 入力」 ステップをダブルクリックし設定画面を開きます。



3. 設定画面に次の内容を設定します。

#	設定項目	設定内容	備考
1	ファイル名	C:\pent¥p102¥design-tools¥data-integration¥samples¥transformations¥files¥sales_data.csv (上記のパスは、「1.2(4)インストールディレクトリの入力」でインストールディレクトリを「C:\pent¥p102」と指定した場合)	サンプルの CSV ファイルの格納場所をフルパスで指定します。 [参照] ボタンをクリックするとファイルダイアログから選択できます。
2	データ型を自動変換	オフ	データ型の変換を省略して性能向上を図る機能です。後続処理でエラーになることがあるので、オフにすることをお勧めします。
3	文字コード	ISO-8859-1	左記のとおり指定してください。
4	フィールド	([フィールドを取得] ボタンをクリックし、次に表示されるダイアログで [OK] ボタンをクリックしてください。)	読み込むファイルの 1 行目をフィールド名として利用し、2 行目以降のデータで型を推測します。

設定後の内容は次のとおりです。

#	フィールド名	データタイプ	書式	長さ	精度	通貨記号	桁区切り文字	数値読み文字	空白除去
1	ORDERNUMBER	Integer	#	15	0	¥	,	,	なし
2	QUANTITYORDERED	Integer	#	15	0	¥	,	,	なし
3	PRICEEACH	BigNumber	#,##	5	2	¥	,	,	なし
4	ORDERLINENUMBER	Integer	#	15	0	¥	,	,	なし
5	SALES	BigNumber	#,##	7	2	¥	,	,	なし
6	ORDERDATE	String		15		¥	,	,	なし
7	STATUS	String		10		¥	,	,	なし
8	QTR_ID	Integer	#	15	0	¥	,	,	なし
9	MONTH_ID	Integer	#	15	0	¥	,	,	なし
10	YEAR_ID	Integer	#	15	0	¥	,	,	なし
11	PRODUCTLINE	String		12		¥	,	,	なし
12	MSRP	Integer	#	15	0	¥	,	,	なし
13	PRODUCTCODE	String		8		¥	,	,	なし
14	CUSTOMERNAME	String		30		¥	,	,	なし

正しく読み込めるかどうかはプレビュー画面で確認できます。確認手順を次に示します。

1. 「プレビュー」 ボタンをクリックします。

#	フィールド名	データタイプ	書式	長さ	精度	通貨記号	桁区切り文字	数値読み文字	空白除去
1	ORDERNUMBER	Integer	#	15	0	¥	,	,	なし
2	QUANTITYORDERED	Integer	#	15	0	¥	,	,	なし
3	PRICEEACH	BigNumber	#,##	5	2	¥	,	,	なし
4	ORDERLINENUMBER	Integer	#	15	0	¥	,	,	なし
5	SALES	BigNumber	#,##	7	2	¥	,	,	なし
6	ORDERDATE	String		15		¥	,	,	なし
7	STATUS	String		10		¥	,	,	なし
8	QTR_ID	Integer	#	15	0	¥	,	,	なし
9	MONTH_ID	Integer	#	15	0	¥	,	,	なし
10	YEAR_ID	Integer	#	15	0	¥	,	,	なし
11	PRODUCTLINE	String		12		¥	,	,	なし
12	MSRP	Integer	#	15	0	¥	,	,	なし
13	PRODUCTCODE	String		8		¥	,	,	なし
14	CUSTOMERNAME	String		30		¥	,	,	なし

2. プレビューする行数をダイアログに入力して [OK] ボタンをクリックすると、次のようなプレビュー画面が表示されます。

プレビューデータ画面

ステップ6(行) CSV入力 (1000 行)

#	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME
1	10107	30	95.7	2	2871	2/24/2003 0:00	Shipped	1	2	2003	Motorcycles	95	S10_1678	Land of Toys I
2	10121	54	81.25	5	2785.9	5/17/2003 0:00	Shipped	2	5	2003	Motorcycles	95	S10_1678	Parkin Collecto
3	10134	41	94.74	2	3884.54	7/1/2003 0:00	Shipped	3	7	2003	Motorcycles	95	S10_1678	Lyon Souveni
4	10145	65	83.26	6	3746.7	8/25/2003 0:00	Shipped	3	8	2003	Motorcycles	95	S10_1678	Toys(Ground)
5	10159	49	100	14	5205.47	10/18/2003 0:00	Shipped	4	10	2003	Motorcycles	95	S10_1678	Corporate Gift
6	10168	36	95.66	1	3475.76	10/28/2003 0:00	Shipped	4	10	2003	Motorcycles	95	S10_1678	Technics Store
7	10160	24	86.17	9	2482.77	11/11/2003 0:00	Shipped	4	11	2003	Motorcycles	95	S10_1678	Executive Desi
8	10168	48	100	1	5512.32	11/18/2003 0:00	Shipped	4	11	2003	Motorcycles	95	S10_1678	Herku Gifts
9	10201	22	98.57	2	2188.54	12/1/2003 0:00	Shipped	4	12	2003	Motorcycles	95	S10_1678	Mini Wheels C
10	10211	41	100	14	4708.44	1/15/2004 0:00	Shipped	1	1	2004	Motorcycles	95	S10_1678	Auto Lancia-4
11	10223	37	100	1	3665.66	2/20/2004 0:00	Shipped	1	2	2004	Motorcycles	95	S10_1678	Australian Col
12	10257	25	100	7	2525.12	4/5/2004 0:00	Shipped	2	4	2004	Motorcycles	95	S10_1678	Witchamatic
13	10251	28	100	2	3185.64	5/18/2004 0:00	Shipped	2	5	2004	Motorcycles	95	S10_1678	Yakin Collecto
14	10263	34	100	2	3435.76	6/29/2004 0:00	Shipped	2	6	2004	Motorcycles	95	S10_1678	Gift Depot Inc
15	10275	45	92.85	1	4177.25	7/25/2004 0:00	Shipped	3	7	2004	Motorcycles	95	S10_1678	La Rochelle G
16	10285	36	100	6	4089.68	8/27/2004 0:00	Shipped	3	8	2004	Motorcycles	95	S10_1678	Marta's Replic
17	10299	27	100	9	2587.79	9/30/2004 0:00	Shipped	3	9	2004	Motorcycles	95	S10_1678	Toys of Landon
18	10309	41	100	5	4894.38	10/15/2004 0:00	Shipped	4	10	2004	Motorcycles	95	S10_1678	Boone Mini Tr
19	10318	46	94.74	1	4358.04	11/2/2004 0:00	Shipped	4	11	2004	Motorcycles	95	S10_1678	Discus Clean

閉じる OK ログアウト

3. 表示されたプレビュー画面で問題がないか確認します。確認する項目の例を次に示します。

- ・ 列名が表示されている
- ・ データが複数行表示されている
- ・ 文字化けしていない

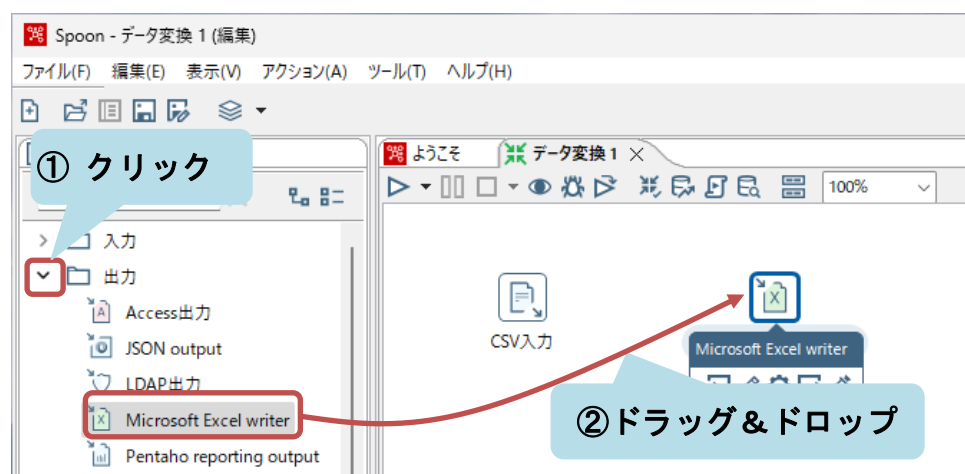
問題なければプレビュー画面の [閉じる] ボタンをクリックしてプレビュー画面を閉じます。

4. 設定完了後、[OK] ボタンをクリックし設定画面を閉じます。

2.3.3 Excel ファイルを出力する設定を追加する

読み込んだ CSV ファイルを Excel ファイルに出力するよう次の手順で設定します。

1. 部品一覧の [出力] - [Microsoft Excel writer] をキャンバスにドラッグ&ドロップします。



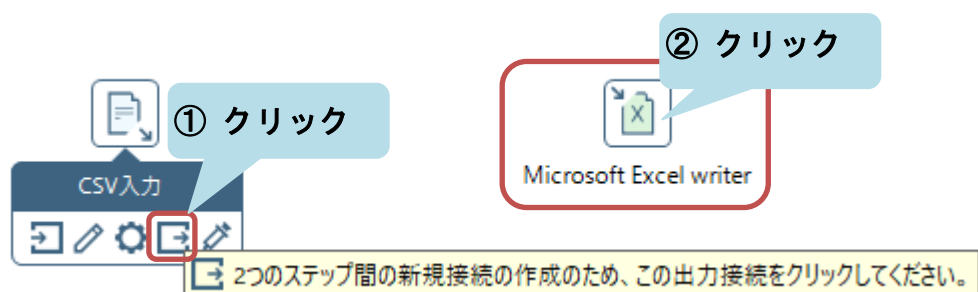
2. キャンバス上の「CSV 入力」ステップと「Microsoft Excel writer」ステップをつなげます（ホップを設定）。

次の手順で実施してください。

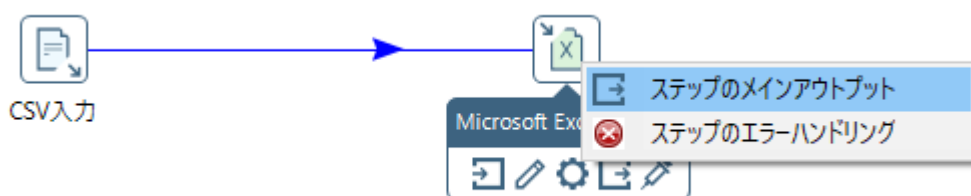
- a. 「CSV 入力」ステップの上にマウスポインタを持って行くと次のような吹き出しが表示されます。



- b.  をクリックし、マウスポインタを「Microsoft Excel writer」ステップまで移動させクリックします。



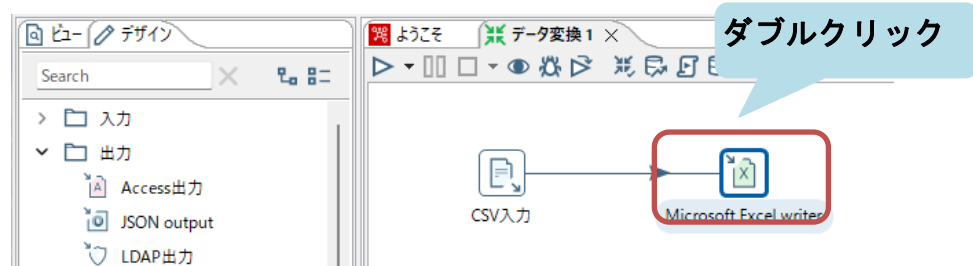
- c. 「ステップのメインアウトプット」を選択し、接続（ホップ）を完了します。



ホップは次の方法でも設定できます。

- 「[Shift] キーを押した状態で、[CSV 入力] ステップをクリックしたまま、マウスポインタを「Microsoft Excel writer」ステップまで移動させクリックを放します。
- ホイールがついたマウスであれば、[CSV 入力] ステップでスクロールをクリックしたまま「Microsoft Excel writer」ステップまで移動させクリックを放します。

3. [Microsoft Excel writer] ステップをダブルクリックし設定画面を開きます。



設定画面には次の内容を設定します。

#	設定項目	設定内容	備考
1	ファイル名	(任意の場所を指定してください。拡張子 (「.xls」など) は不要です。)	変換後の Excel ファイルの格納場所をフルパスで指定します。ファイル名は CSV ファイル名と異なる名称も指定できます。 [参照] ボタンをクリックするとファイルダイアログから選択できます。

設定後の設定画面は次のとおりです。

Microsoft Excel writer

ステップ名 Microsoft Excel writer

ファイル & シート コンテンツ

ファイル

ファイル名 C:\Excel\Excel.xls 参照(B)...

拡張子 xls [Excel 97 以降]

Create parent folder ☐

ストリームXSLXデータ ☐

データを区切る行数 0

ファイル名にステップ番号を含める ☐

ファイル名に日付を含める ☐

ファイル名に時刻を含める ☐

日付のフォーマットを指定する ☐

日付フォーマット

ファイル名を表示...

出力ファイルが既に存在する場合は 新たな出力ファイルと置き換える

最初の行を取得時にファイルを作成する ☐

結果にファイル名を追加する ☒

シート

シート名 Sheet1

アクティブシートに設定 ☒

出力ファイルにシートが存在する場合は 新たなシートと置き換える

シートを保護する (XLSフォーマットのみのみ) ☐

ユーザー名

パスワード

テンプレート

新規ファイル作成時にテンプレートを使用する ☐

テンプレートファイル template.xls 参照(B)...

新規シート作成時にテンプレートを使用する ☐

テンプレートシート

Hide Template Sheet ☐

ヘルプ OK(O) キャンセル(C)

設定完了後、[OK] ボタンをクリックして設定画面を閉じます。

2.3.4 作成したデータ変換を保存する

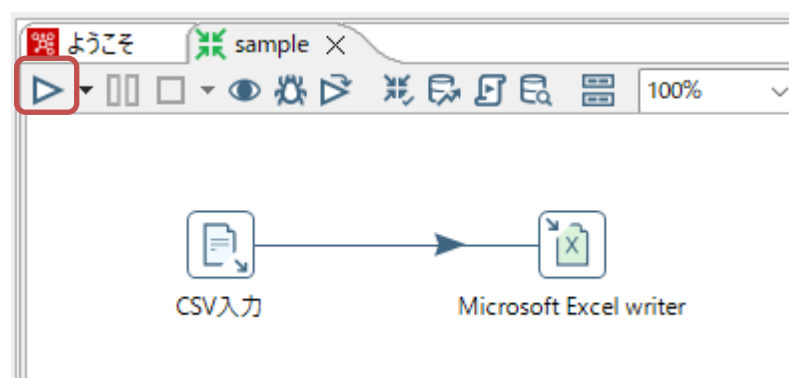
［ファイル］メニューー［保存］を選択して、データ変換定義を保存します。新規に保存する場合はファイルダイアログが表示されますので、任意の場所に保存します。



2.4 データ変換を実行する

作成したデータ変換を実行する方法について説明します。

1. キャンバスの左上にある  (実行) ボタンをクリックします。



2. 「データ変換の実行」ダイアログが表示されますので、「実行」ボタンをクリックします。

データ変換の実行

Run configuration:
Pentaho local

詳細

☒ 実行前にログをクリアする ログレベル: 基本のログ

☐ セーフモードを有効にする

☒ パフォーマンスメトリクスを収集

パラメーター 変数

パラメーター	デフォルト値	値	Description

引数

☒ Always show dialog on run

ヘルプ 実行(L) キャンセル(Q)

3. 実行すると右下に実行結果のペインが表示されます。

Spoon - sample

実行結果

2025-02-26 11:02:30.370 - General - Logging plugin type found with ID: CheckpointingSaver
2025-02-26 11:02:30.888 - Carte - Installing timer to purge stale objects after 1440 minutes.
2025-02-26 11:02:32.414 - Spoon - Save file as...
2025-02-26 11:02:33.589 - Spoon - データ変換をインストールしました。
2025-02-26 11:02:35.604 - Spoon - データ変換をインストールしました。
2025-02-26 11:02:35.609 - Spoon - データ変換を実行しました。
2025-02-26 11:02:36.204 - sample - License validated.
2025-02-26 11:02:36.211 - sample - Dispatching started for transformation [sample]
2025-02-26 11:02:36.212 - CSV入力 - Header row skipped in file 'C:\pencil\1326design\tools\integrations\samples\transformations\Files\Roles_data.csv'
2025-02-26 11:02:36.262 - CSV入力 - 処理を完了 (J=2024, Q=0, R=0, W=2625, U=0, E=0)
2025-02-26 11:02:36.262 - Microsoft Excel writer - 処理を完了 (J=0, Q=2024, R=2625, W=2625, U=0, E=0)
2025-02-26 11:02:36.289 - Spoon - データ変換が完了しました。

正常に実行したステップの右上に緑色のチェックマークがつきます。



実行に失敗したステップは赤枠で囲まれ右上に赤色の停止マークがつきます。



- データ変換に成功すると、指定した場所に Excel ファイルが出力されます。

Excel ファイルを開くと、読み込んだ CSV ファイルの中身が出力されていることを確認できます。

	A	B	C	D	E	F	G	H	I	J	K
1	ORDERNO	QUANTITY	PRICEEA	(ORDERLIN	SALES	ORDERDA	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCT
2	10107	30	95.7	2	2871	2/24/2003	Shipped	1	2	2003	Motorcycle
3	10121	34	81.35	5	2765.9	5/7/2003	0 Shipped	2	5	2003	Motorcycle
4	10134	41	94.74	2	3884.34	7/1/2003	0 Shipped	3	7	2003	Motorcycle
5	10145	45	83.26	6	3746.7	8/25/2003	Shipped	3	8	2003	Motorcycle
6	10159	49	100	14	5205.27	10/10/2003	Shipped	4	10	2003	Motorcycle
7	10168	36	96.66	1	3479.76	10/28/2003	Shipped	4	10	2003	Motorcycle
8	10180	29	86.13	9	2497.77	11/11/2003	Shipped	4	11	2003	Motorcycle
9	10188	48	100	1	5512.32	11/18/2003	Shipped	4	11	2003	Motorcycle
10	10201	22	98.57	2	2168.54	12/1/2003	Shipped	4	12	2003	Motorcycle
11	10211	41	100	14	4708.44	1/15/2004	Shipped	1	1	2004	Motorcycle
12	10223	37	100	1	3965.66	2/20/2004	Shipped	1	2	2004	Motorcycle
13	10237	22	100	7	2222.12	4/5/2004	0 Shipped	2	4	2004	Motorcycle

- 確認できたら Excel ファイルを閉じます。

3. 実行後のデータを加工する

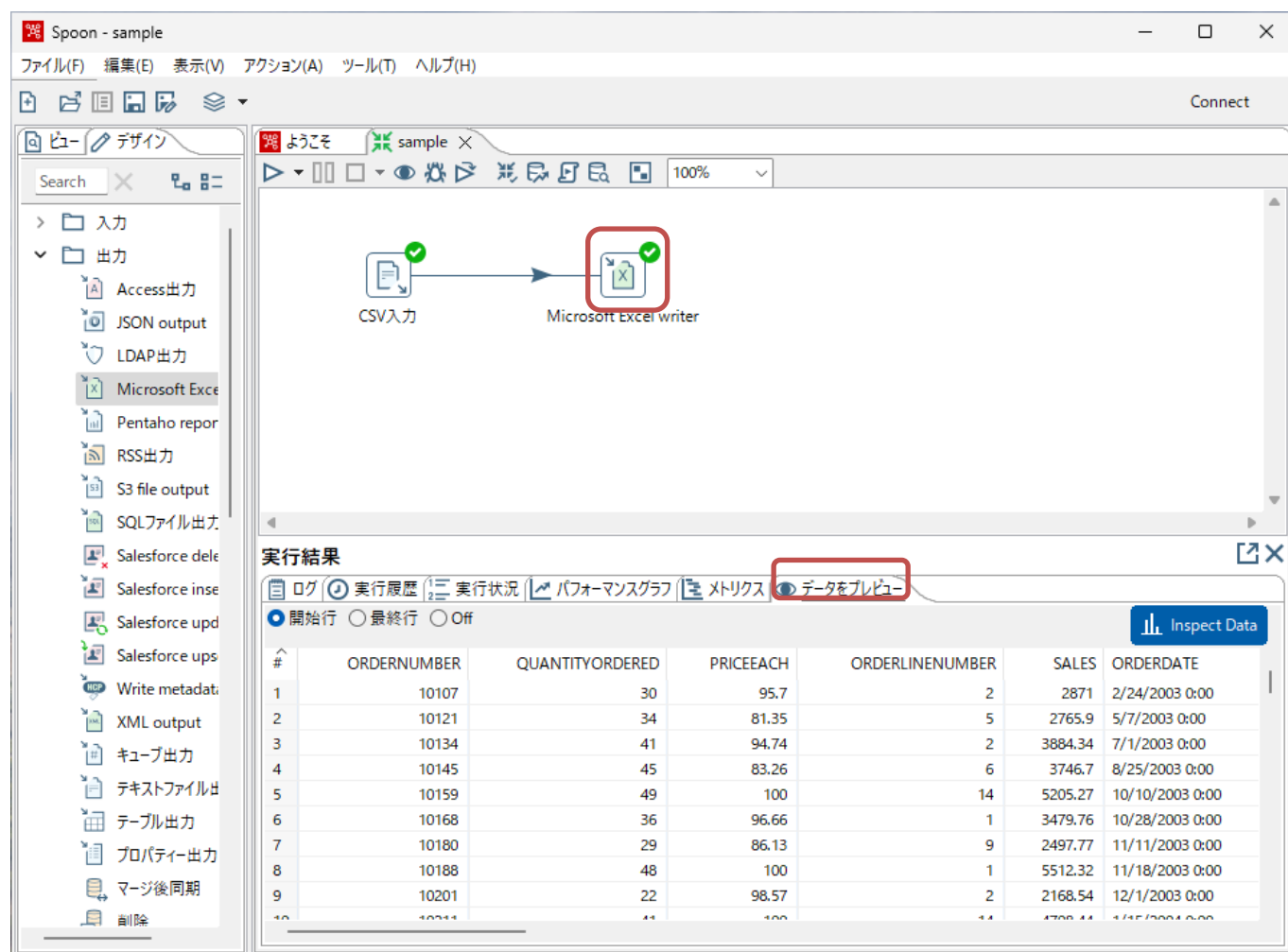
実行後のデータの確認とデータの加工の流れを説明します。

3.1 実行結果のデータの中身を確認する

PDI ではステップごとに実行結果を確認できます。

3.1.1 各ステップの実行結果を確認する

実行すると右下に実行結果ペインが表示されます。[データをプレビュー] タブをクリックするとキャンバスで選択しているステップの実行結果を確認できます。



The screenshot shows the Spoon - sample application interface. The canvas displays a data flow from 'CSV入力' to 'Microsoft Excel writer'. The '実行結果' (Execution Results) pane at the bottom shows the 'データをプレビュー' (Preview Data) tab selected, displaying a table of order data.

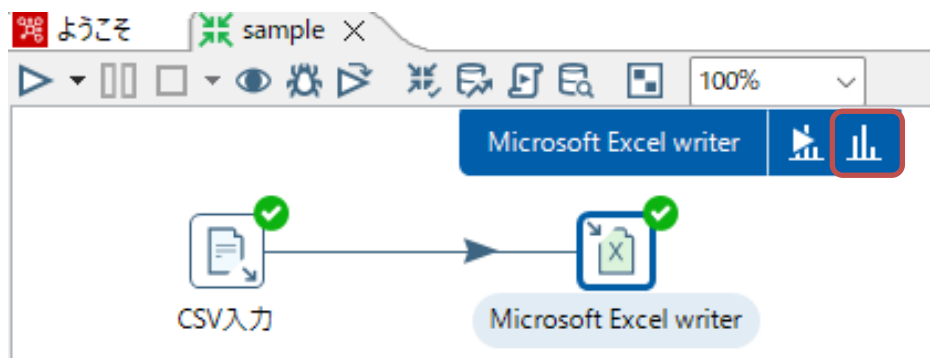
#	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE
1	10107	30	95.7	2	2871	2/24/2003 0:00
2	10121	34	81.35	5	2765.9	5/7/2003 0:00
3	10134	41	94.74	2	3884.34	7/1/2003 0:00
4	10145	45	83.26	6	3746.7	8/25/2003 0:00
5	10159	49	100	14	5205.27	10/10/2003 0:00
6	10168	36	96.66	1	3479.76	10/28/2003 0:00
7	10180	29	86.13	9	2497.77	11/11/2003 0:00
8	10188	48	100	1	5512.32	11/18/2003 0:00
9	10201	22	98.57	2	2168.54	12/1/2003 0:00

3.1.2 データをグラフ化する

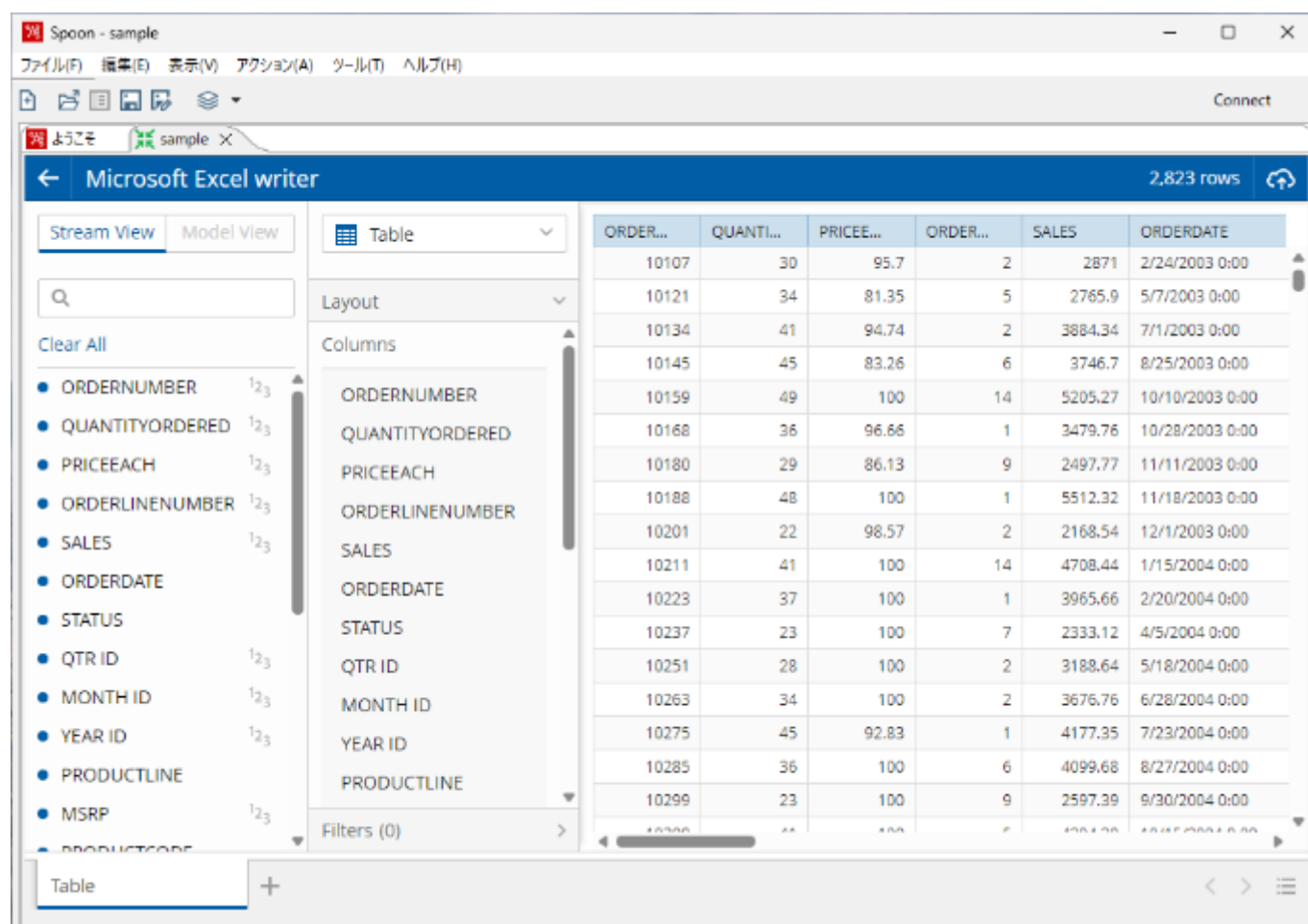
PDI 上でデータをグラフ化して視覚的にデータを把握できます。

ここでは [Microsoft Excel writer] ステップで [POSTALCODE] 列の欠損があるかどうかを確認します。

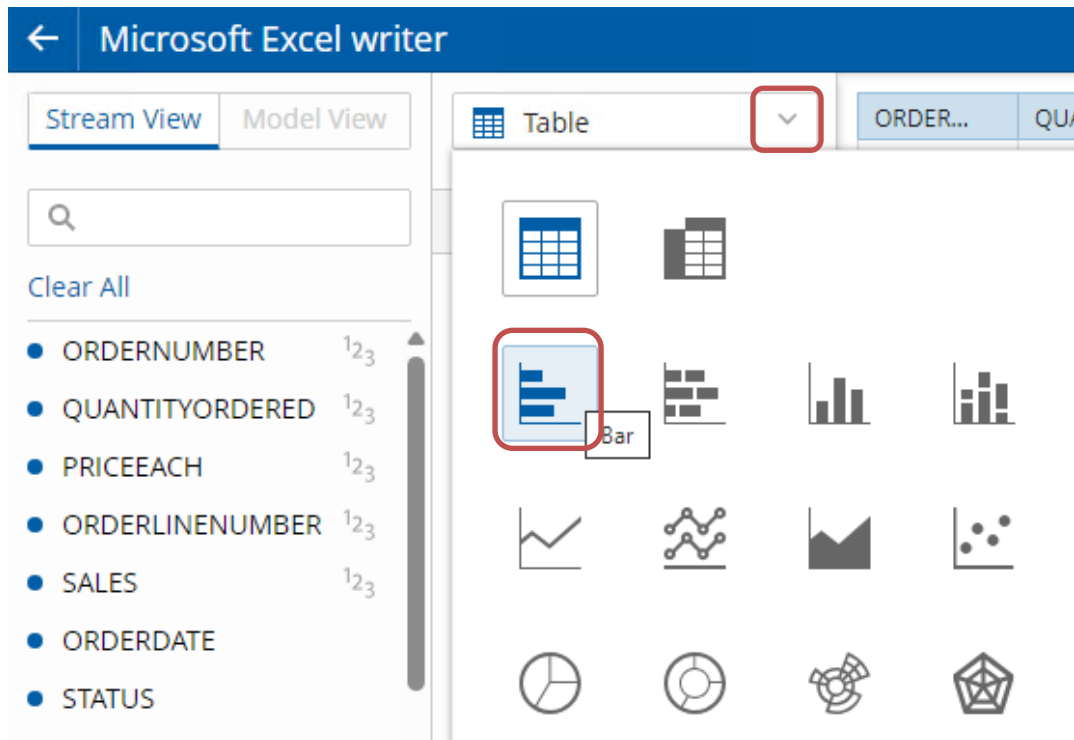
1. [Microsoft Excel writer] ステップをクリックし、右上の  をクリックします。



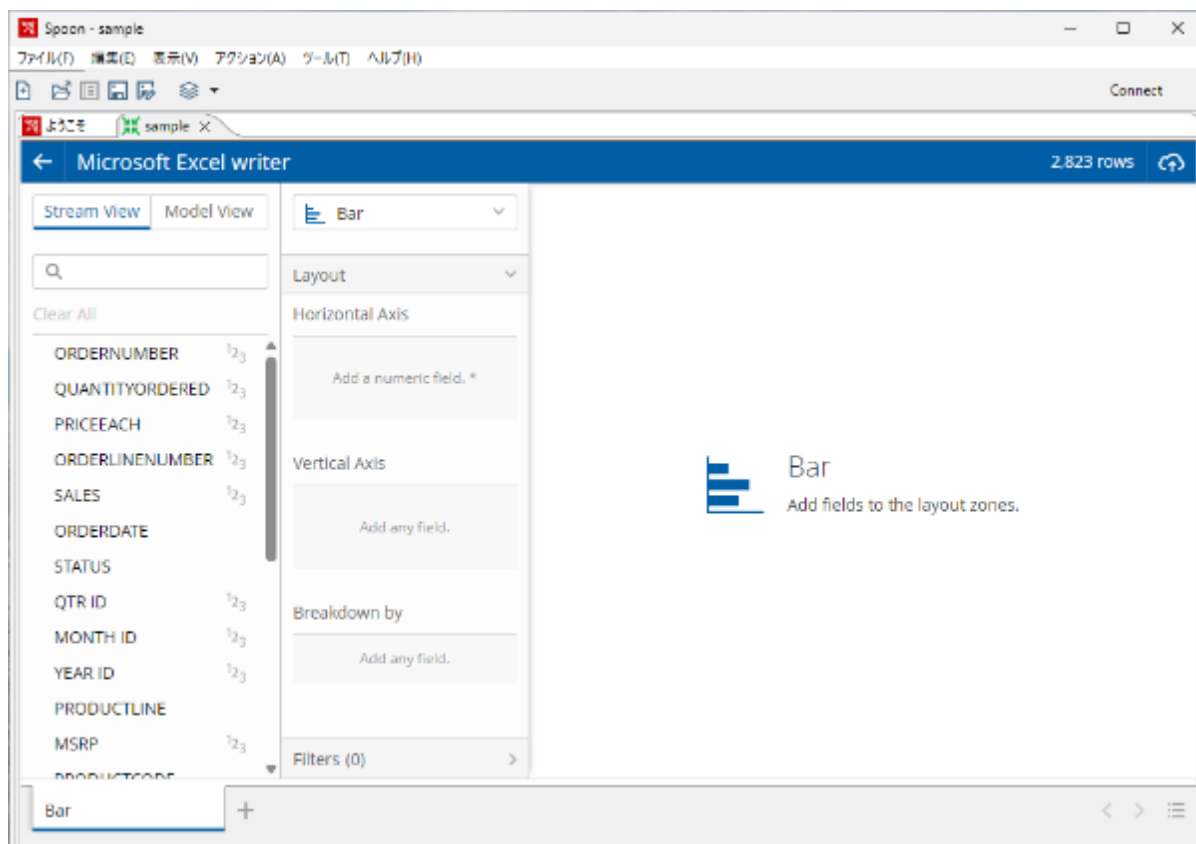
次のような画面が表示されます。



2. [Table] のプルダウンメニューから横棒グラフを選択します。



次のような画面が表示されます。

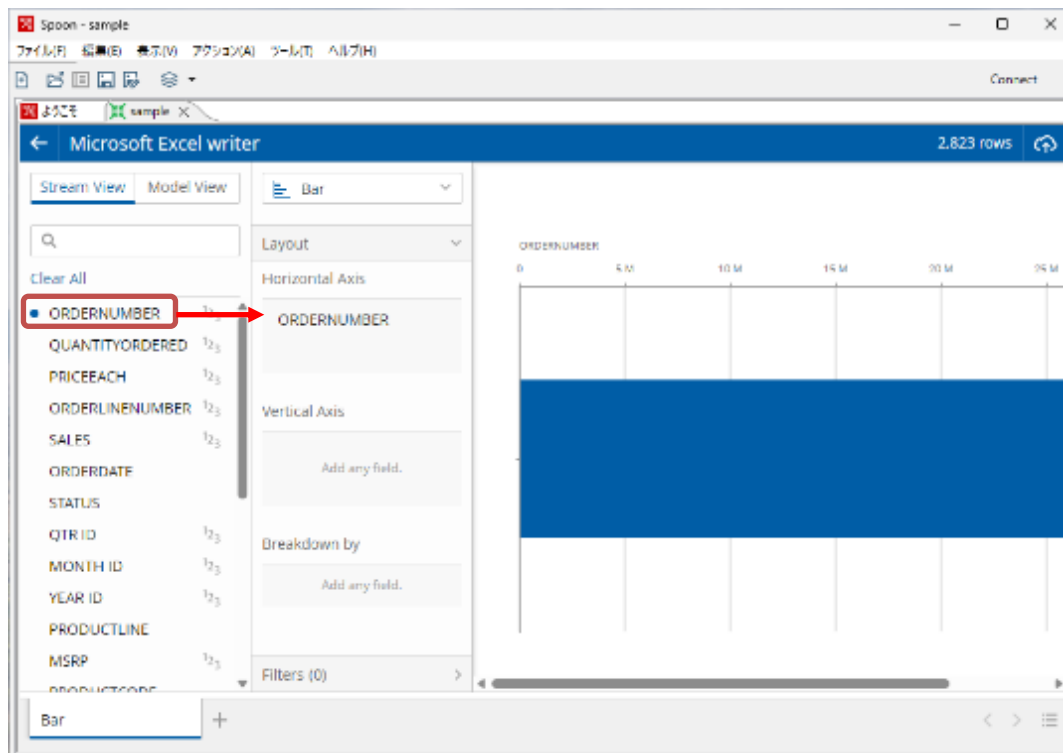


3. [Layout] にフィールドを設定します。

ここでは、横棒グラフの横軸に [ORDERNUMBER] 列、縦軸に [POSTALCODE] 列を設定します。

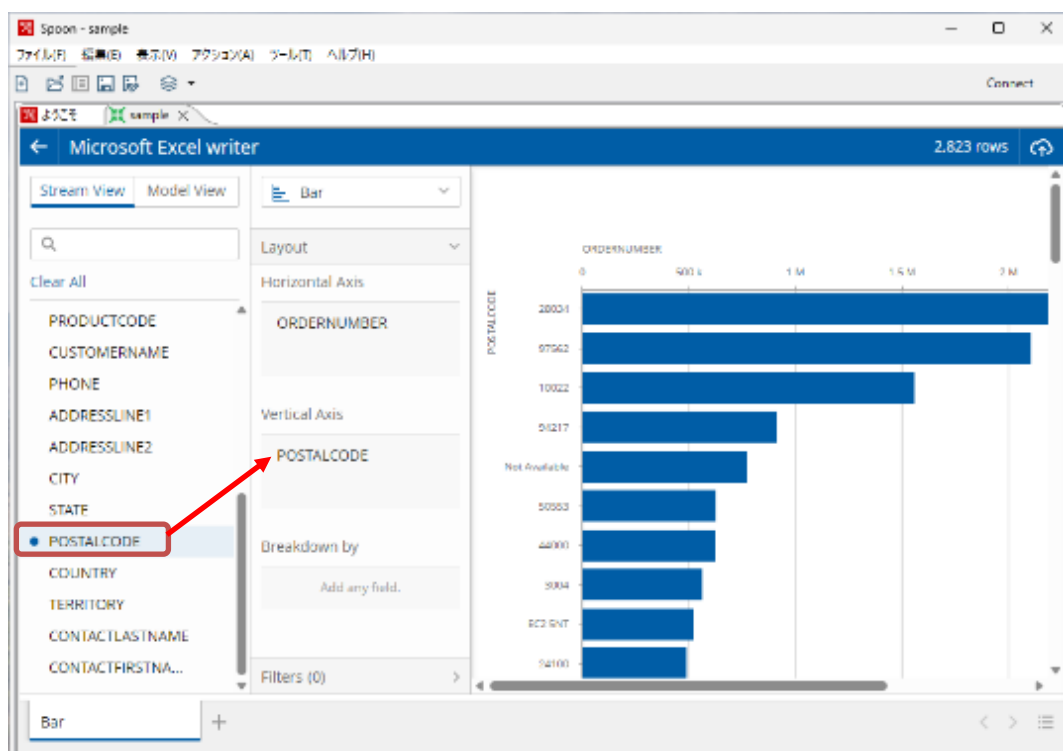
a. [ORDERNUMBER] 列の配置

[Horizontal Axis] に [ORDERNUMBER] 列をドラッグ&ドロップします。



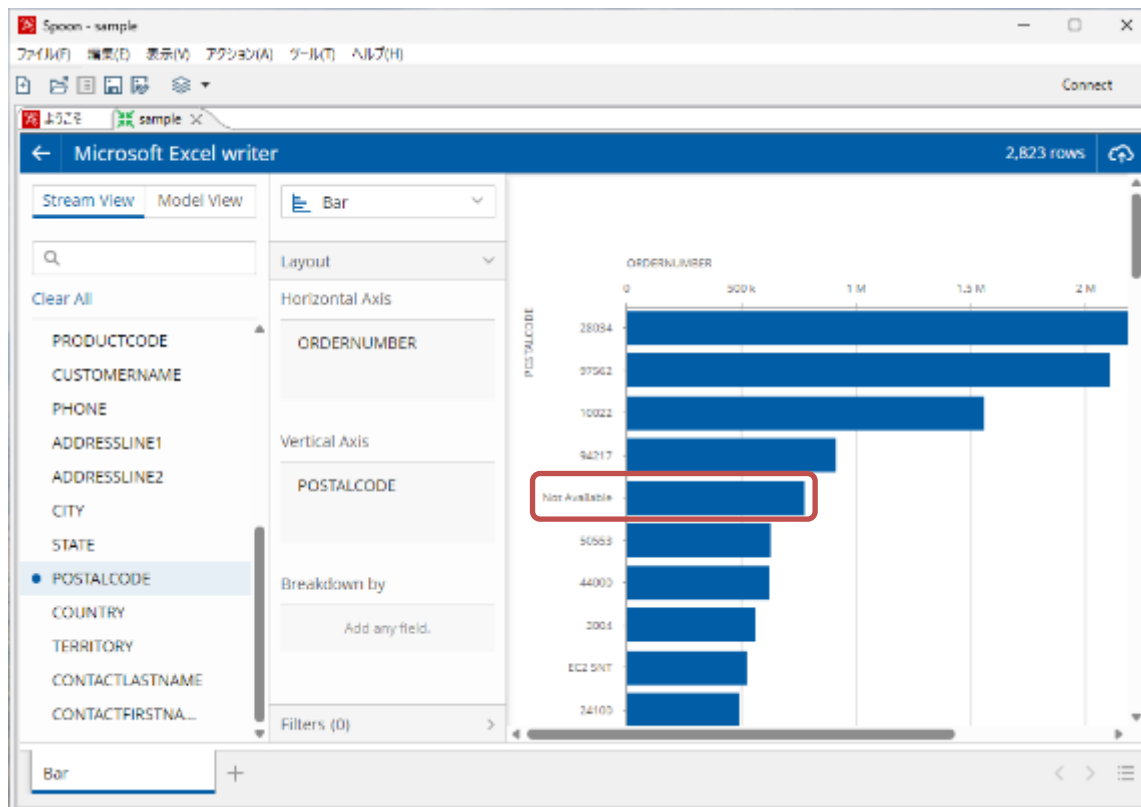
b. [POSTALCODE] 列の配置

[Vertical Axis] に [POSTALCODE] 列をドラッグ&ドロップします。

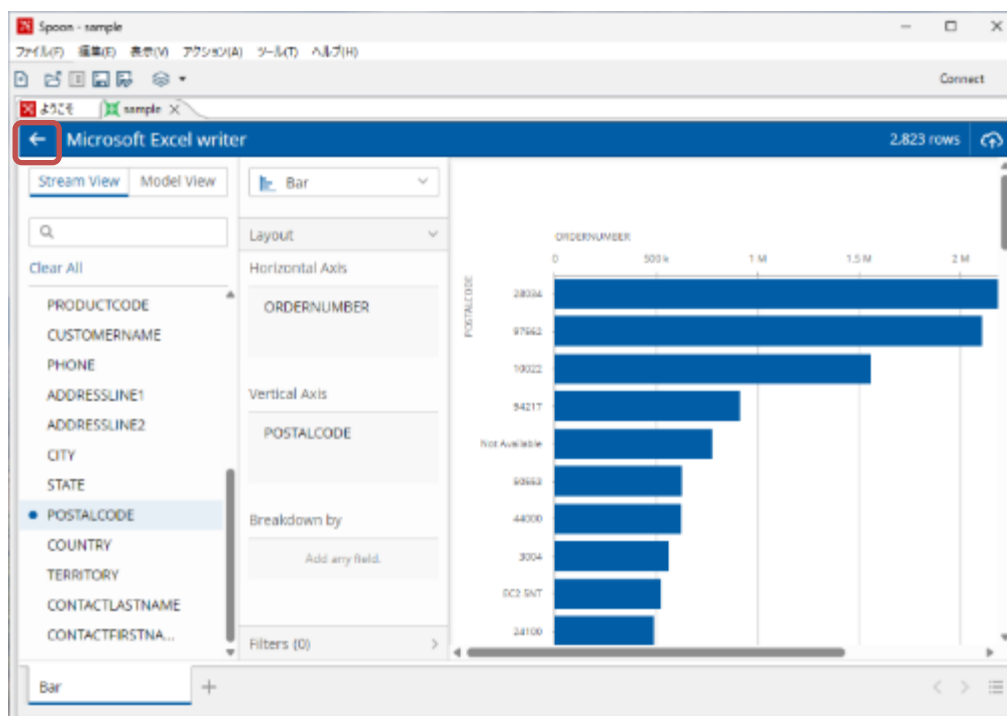


4. 「Not Available」があることを確認します。

「Not Available」（欠損）が表示されている「POSTALCODE」列に、欠損があることを示します。
(横棒の大きさは「ORDERNUMBER」の合計値を表しており、欠損の個数ではありません。)



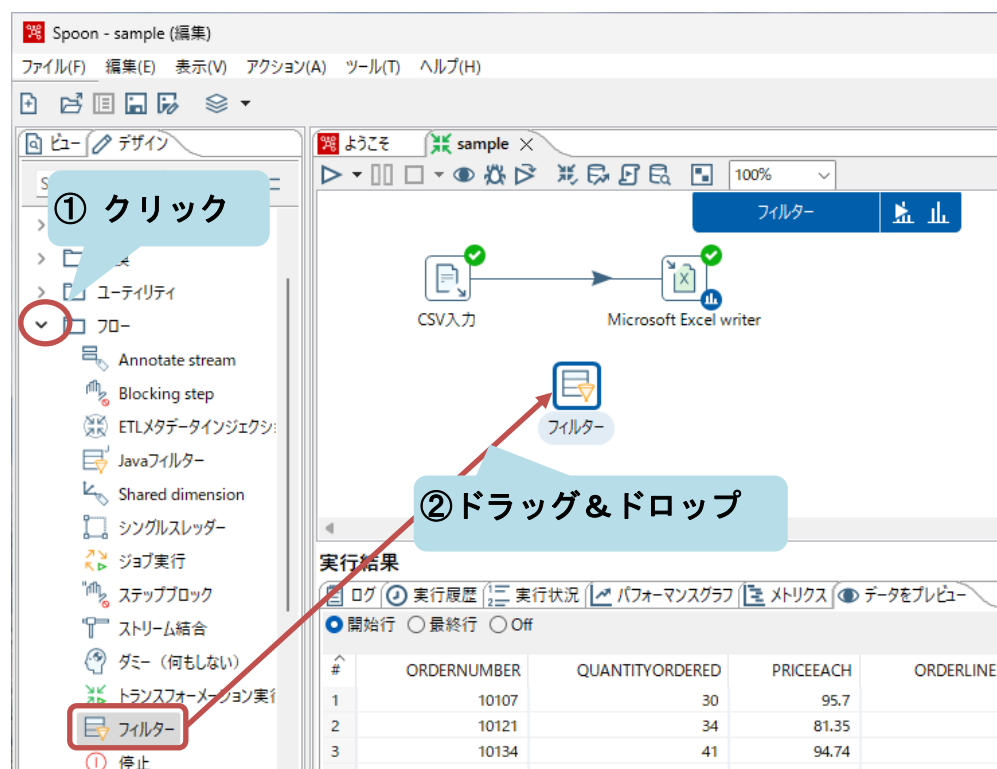
5.  をクリックし、データ変換定義画面に戻ります。



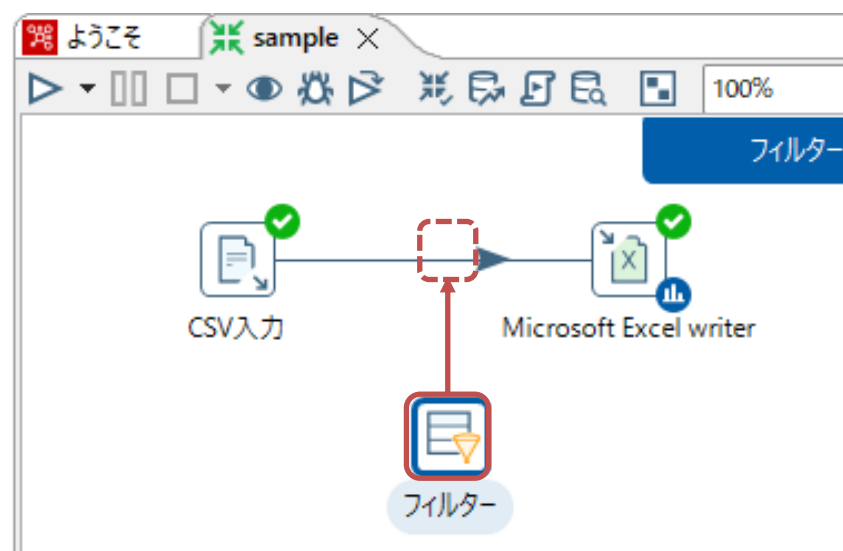
3.2 不要なデータを除外する

［フィルター］ステップで条件式を設定し、不要なデータを除外すれば、欠損を回避できます。

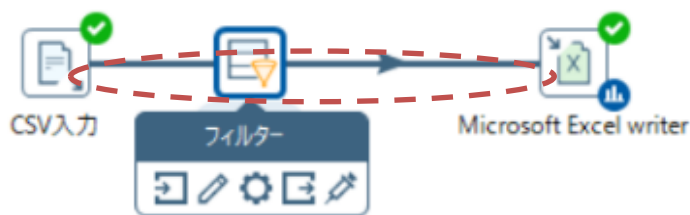
1. 部品一覧の［フロー］－［フィルター］をキャンバスにドラッグ&ドロップします。



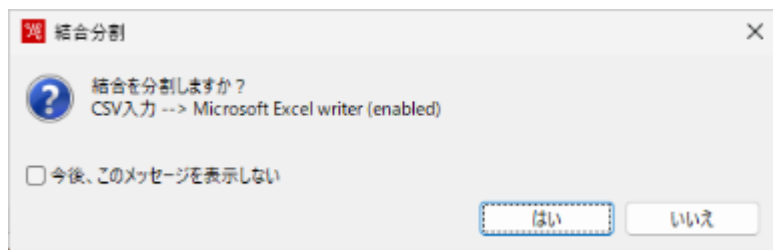
2. キャンバスに配置した［フィルター］ステップを、［CSV 入力］ステップから［Microsoft Excel writer］へのホップの上にドラッグします。



ホップの上にステップをドラッグすると次の図のようにホップが太くなる場所があります。そこで［フィルター］ステップをドロップします。



次の図のようなメッセージが表示されますので、[はい] ボタンをクリックします。



そうすると次の図のように、PDI が自動でホップを分割し、つなぎ直します。



[注意]

次の図のように [フィルター] ステップの左に → が付いていない場合は、分割できていません。



3. [フィルター] ステップをダブルクリックし設定画面を開きます。

設定画面にはステップ名のほか、次の内容を設定します。

#	設定項目	設定内容	備考
1	条件式が真 (True) のときに実行する処理	Microsoft Excel writer	
2	フィルタ条件 (左)	POSTALCODE	
3	フィルタ条件 (中)	IS NOT NULL	

設定後の設定画面は次のとおりです。

フィルター

ステップ名 フィルター

条件式が真 (True) のときに実行する処理 Microsoft Excel writer

条件式が偽 (False) のときに実行する処

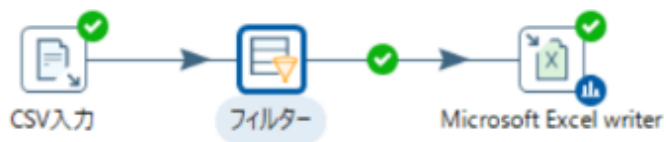
フィルタ条件

POSTALCODE IS NOT NULL

ヘルプ OK(O) キャンセル(C)

[補足]

設定画面を閉じると次の図のように「フィルター」ステップと「Microsoft Excel writer」のホップに✔が付きます。



3.3 データ変換を再実行する

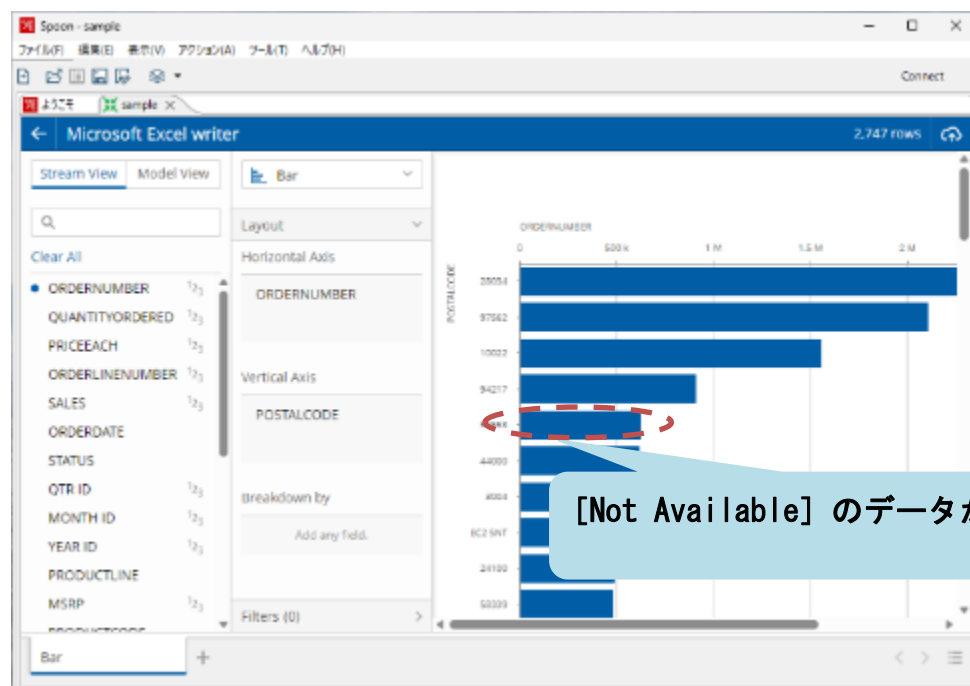
次に示す手順で加工後のデータを変換し、問題ないことを確認します。

1. 「2.4 データ変換を実行する」を参照し、データ変換を実行します。

[補足]

「2.4 データ変換を実行する」で作成した Excel ファイルを開いたまま実行すると手順 2 でエラーになります。Excel ファイルを閉じてから実行してください。

2. 「3.1.2 データをグラフ化する」を参照し、再度グラフ化します。
3. [Not Available] のデータがないことを確認します。



4. 付録

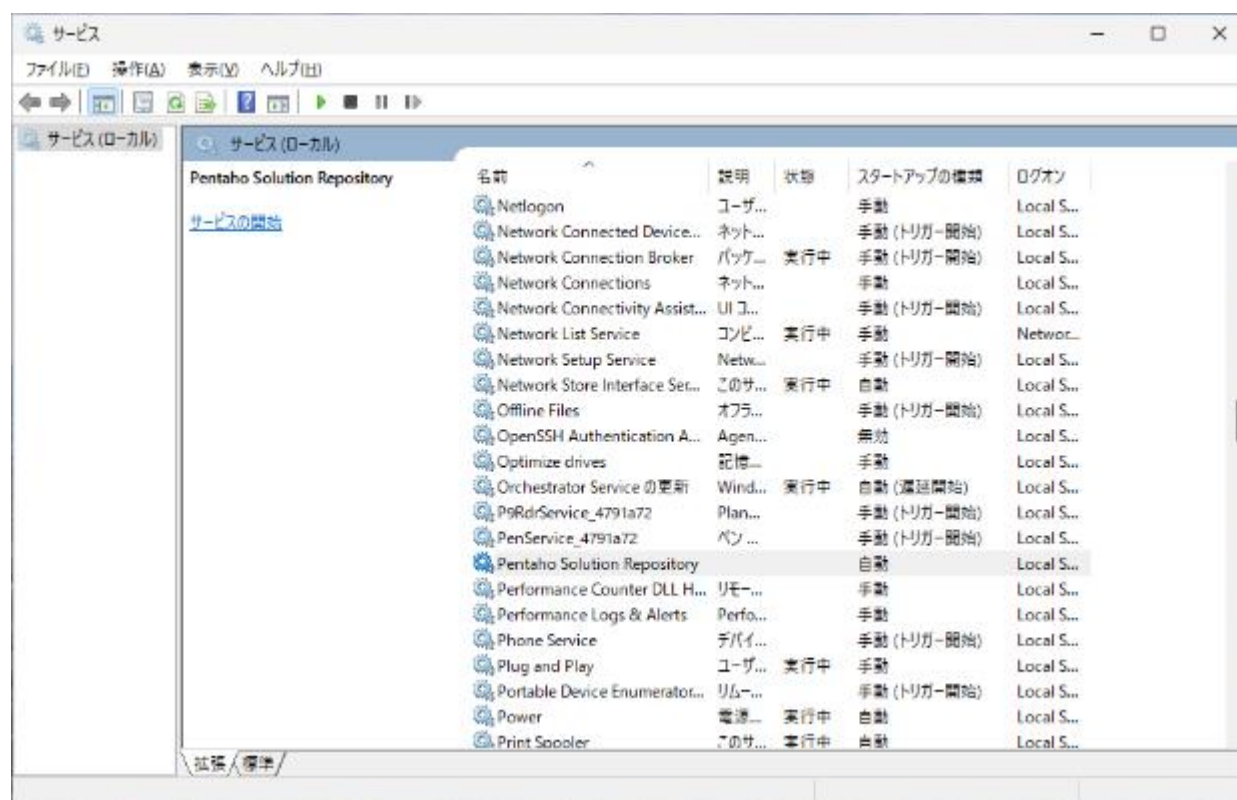
4.1 Pentaho Server の起動・停止方法

サーバ用コンポーネントでのスケジュール実行などに必要な Pentaho Server について、起動方法と停止方法を説明します。

4.1.1 起動

Pentaho Server の起動手順を次に示します。

1. Windows のスタートメニューの「すべて」－「Windows ツール」－「サービス」を選択し、「サービス」を開きます。

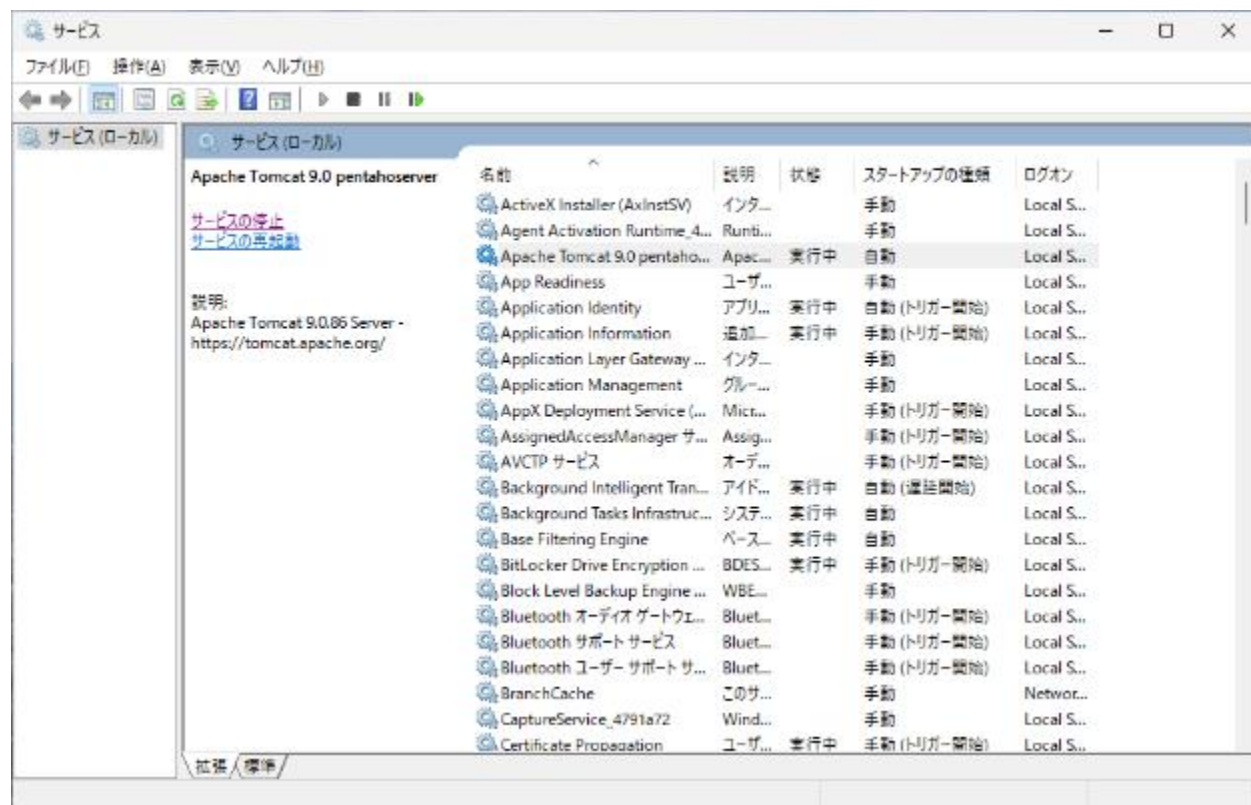


2. 次の順でサービスを開始します。
 1. Pentaho Solution Repository
 2. Apache Tomcat 9.0 pentahoserver

4.1.2 停止

Pentaho Server の停止手順を次に示します。

1. Windows の「サービス」を開きます。



2. 次の順でサービスを停止します。

1. Apache Tomcat 9.0 pentahoserver
2. Pentaho Solution Repository