

Pentahoでデータ連携：便利に自動化

著者	大曾根 康裕
雑誌名	技術報告
巻	19
ページ	5-8
発行年	2014-03-10
出版者	静岡大学技術部
URL	http://doi.org/10.14945/00008033

Pentaho でデータ連携 ～便利に自動化～

大曾根 康裕

名古屋工業大学 技術部 研究基盤課

1. サービス(システム)の連携について

本社と支社が違う経理システムを導入しているものだから、フロッピーでデータ交換しないと業務が回らなくて…。こうした作業、身に覚えがあるだろうか。複数のシステムがうまく回るようにデータ交換する事を、システム連携という。大規模に運用されているサービスの多くは、利用者の要望や状況に合うよう定期的にデータを交換する事でうまく回っている。

2. 無償版の Pentaho でサービス連携

本学は2010年9月より Pentaho (Data Integration)を導入、サービスの連携を行って来た[1]。Pentahoは2013年現在、無償版(CE)と、3種類の有償版(EE)が提供されている[2]。有償版は代理店の KSK アナリティクスと年間契約(約 170 万円/年～)で、自動更新や技術サポートが享受できる。ただ、システム連携をするだけであれば無償版で充分である。ソフトの更新も活発で、非常に有用である。

3. どんな事ができるのか

3.1 データベースの取り扱い

最も頻用しているのは、データベースを取り扱う事である。データベースは普段の生活では直接触れないようなものであるが、非常によく活用されている。身近な所ではツイッターやブログ・google や図書館等の検索システム・列車や航空機の予約システム・銀行の預金情報システムなど、大量の情報をリアルタイムかつ正確に取り扱うシステムで多く使われている。こうしたデータベースに対し、情報を取り出したリ整理したり、移し替えたりする作業を良く行っている。

3.2 ファイルの転送

手元の書類を遠隔のパソコンに送り届けたり、取得したりする事をファイル転送という。これは普段の生活の中でも活用しているような事である。FTP・ウィンドウズ共有を主に用いている。

3.3 ファイルの読み書き

ファイルに書き込まれている情報を読み出したり書き込んだりできる。CSV やエクセル等の情報を利用できる。本学で利用した事はないが、科学の分析機器で書き出されるバイナリーデータ等も、書類の書式が分かっていたら利用できると思われる。

3.4 ウェブサイトの利用

実態としては上記のファイル転送に近い。ウェブサイトから、単に書類をダウンロードする事もできるし、アカウントを使ったログインした上で取得する事もできる。

3.5 メールを送受信

電子メールを送受信する事も可能である。添付書類を付けて送るといった事もお手の物である。定期的にメールを確認し、送られて来たメールに応じて作業を変えるなどという仕組みも作れる。

3.6 つまり…

デスクワークで行う主な作業は、だいたい扱える事ができる。毎日の決まったルーチンワークを、Pentaho にさせてみるのはいかがだろうか。個人的に使用する事もできる。

4. どうやって使うのか

Pentaho には設計ツール(Spoon)と、実行専用ツール(Pan / Kitchen)が同梱されている。連携の工程表のようなものを作成するわけだが、Spoon を使って流れ図を描くように行う(図 3)。

先に述べたように、一つ一つの『できること』が部品として用意されており、これらを適宜繋ぎ合わせる。作成したものは書類に保管し、必要な時に実行できる。

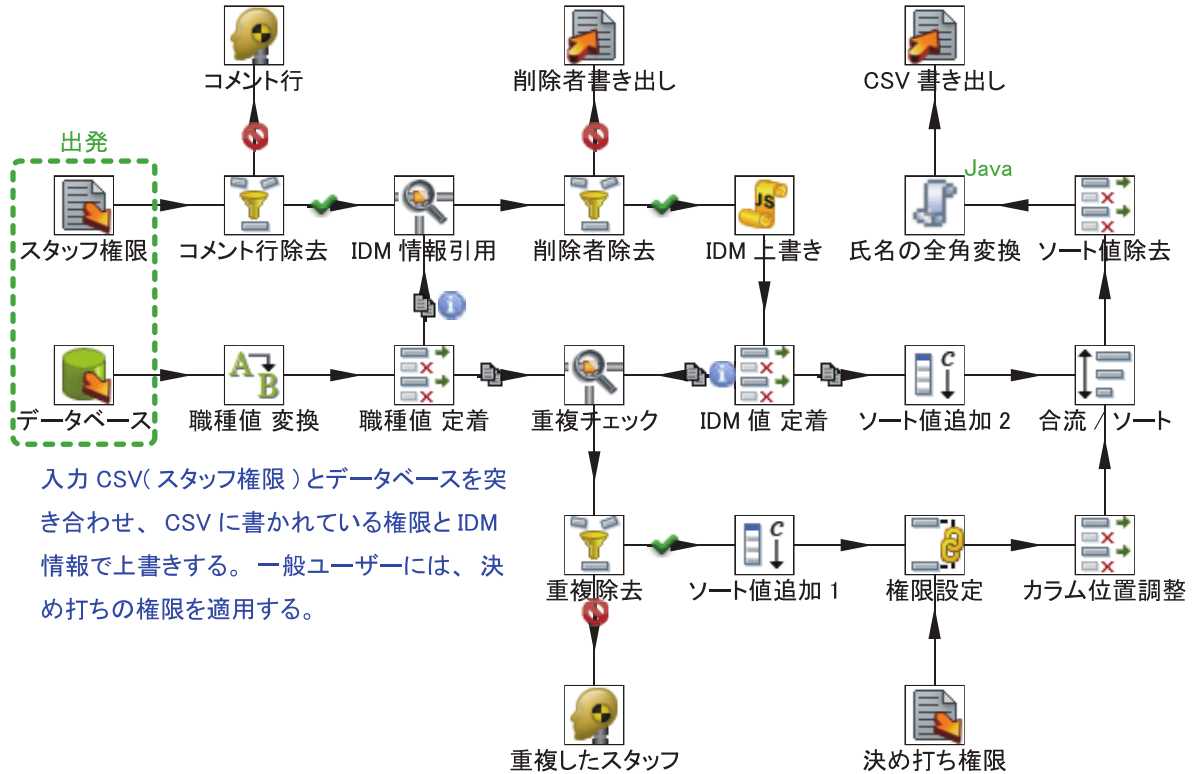


図 3 設計の例

5. 自動実行

自分の代わりに毎日のルーチンワークをしてもらう為には、指定した時間に働いてもらう必要がある。無償版には自動実行システムは同梱されていないため、ここは自前で組み立てる。とは言っても基本を押さえればそれほど難しくない。

5.1 実行専用ツールの活用

自動実行の際にはウィンドウが勝手に開く仕組みが邪魔になるので、実行専用ツール(Pan / Kitchen)を使う。“pan /file:some.ktr”のように書類を指定して起動すると、画面上に進行状況が表示される。定期的にこの指令をシステムから呼び出してもらえば、自動実行となる。

5.2 タスクスケジューラ

定期呼び出しに何をを使うかは OS 環境次第だが、Windows ならタスクスケジューラが良いだろう。Mac OSX や Linux なら launchctl や crontab 等が利用できる。本学では Windows Server 2008 R2 Enterprise で運用しているため、タスクスケジューラを使っている(図 4)。

タスクスケジューラ

名前	状態	トリガー
pentaho - SharePoint	準備完了	毎日 0:04 に起動
pentaho - SIP Server	無効	毎日 0:00 に起動 - トリガーされた後、1 日間の間 1 時間毎に繰り返します。
pentaho - test	実行中	
pentaho - サテライト入室	準備完了	毎日 9:20 に起動
pentaho - 卒業生DB	準備完了	毎日 8:15 に起動 - トリガーされた後、12 時間の間 30 分間毎に繰り返します。
pentaho - 安全管理室	準備完了	毎日 5:00 に起動
pentaho - 情報工学科CSE	準備完了	毎日 2:20 に起動
pentaho - 携帯IDm	準備完了	毎日 0:02 に起動
pentaho - 教務DB → 出欠DB	準備完了	毎日 22:00 に起動
pentaho - 研究者DB	準備完了	毎日 5:00 に起動
pentaho - 統一DB → 教務DB	準備完了	毎日 2:15 に起動

図 4 タスクスケジューラ

6. 使い勝手を良くするコツ

前述のように Pentaho には作業の部品が手広く用意されており、簡単に組み合わせるだけでも多くの事が実現できる。だが万全ではない。例えばメールでの日本語の扱いが不完全であったり、FTP 転送時に特殊な処理をさせたい場合に対応できなかったりする。よく使われそうな作業や、致命的な欠陥を優先的に開発されているようで、一部に具合が悪い所もある。代理店と契約を締結していれば、改善を要求できるが、無償版なので自分で対応する。対応方法としては、簡単なものから難易度の高いものまで幾つかある。

6.1 バッチファイル・Visual Basic スクリプト等の利用

作業内容をバッチファイルに丸投げする事もできる。Text File Output という部品にその機能が備わっている訳だが、スクリプト作成の知識があれば、欲しい機能を簡単に追加できる。

6.2 プラグインの利用

Pentaho の plugin フォルダ内にプログラムを設置する事で機能が追加される。これは非常に有用で強力な拡張方法である。本学独自の要望に応える為にいくつか作成した。

ただ、プラグインを作るのは少々敷居が高い。Pentaho が利用している基本システムをある程度理解し、開発ソフト(Eclipse)を活用して開発する事になる。この初期設定は大変だが、設定の手順書もある。Eclipse の開発機能がかなり強力なので、コツが掴めれば難しくはない。

7. どんな連携をしてきたか

7.1 IC カード入室システム

図 3 の事例に相当するが、全学的に個人情報管理しているデータベースから IC カード情報を引用し、教室の解錠システムに反映させる仕組みである。教室の入り口で職員証をかざすだけで入室できる便利なシステムである。かつては担当者が手作業で IC カードの情報を集計して登録していた。設定する条

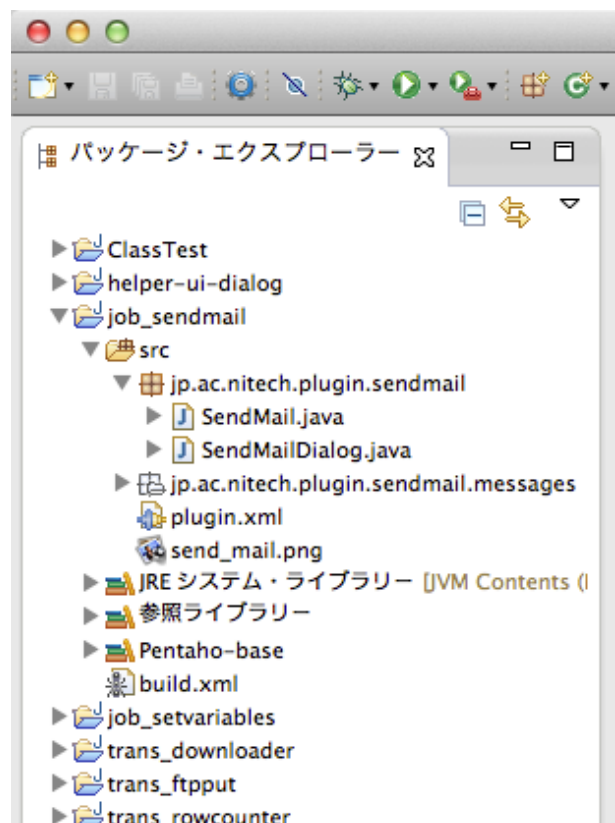


図 5 Eclipse 開発画面の例

件や項目に細かい決まりがあり，これを全て手作業で行うのは大変であった．今では大部分が自動的に処理されるようになった．

なお，IC カードが絡む連携は他にも，カードの再発行や携帯電話の FeliCa に関する連携も扱っている．カード情報が一元的に管理できるようになると，複数のサービスを一つのカードで済ませられるようになる．職員証がコピーカードがわりに使えるようになると便利そうである．

7.2 アカウントの保全

アカウントとパスワードを入力するようなウェブサービスを利用された事はあるだろうか．ブログのようなシステムには独自のアカウント管理システムがある．こうしたアカウントの保全も取り扱っている．本学では全学的な在籍状況も一元的に管理されているため，これらの中から利用権限のあるアカウントを取り出して反映させる．それぞれのアカウントの登録方法に癖があり，手動で指定された書式に加工するのは大変であった．

8. 連携の要望は変化する

連携対象のサービスは時折変化するものである．サービスを請け負っている会社との契約更新で事業者が変わったり，ソフトのバージョンが上がったりする事がある．こうした場合も，要求される書式が変わるだけであれば，入出力担当を交換すれば事足りる．

Pentaho では作業内容を細かく分類し，担当を繋ぎ合わせる事で全体の作業を成している．処理すべき内容に変動があった場合，どの担当作業に変更を加えるべきかが分かりやすい．

9. おわりに

これまでの3年間の運用で，“かなりの事が Pentaho で実現できる”というのを幾度も経験して来た．これは先の強力な拡張性に依る所が大きい．無償版にもかかわらず，強力な機能が利用できるのは，オープンソース(LGPL)で提供されているおかげである．

また，部品の役割分担が明確で，視覚的にもよく取りまとめられている．この事は作業内容の把握を容易にし，作業の引き継ぎにも有利である．連携プログラムが分散せずに一カ所に集約して運用すれば，組織としても連携の把握がしやすくなるだろう．

毎日のルーチンワークの中に Pentaho で解決できる事があるかも知れない．また，本学の連携で作成したプラグインの中には他にも転用ができるものもあるかもしれない．書類を読み出し，加工・転送する仕組みは，科学・データ分析用途にも応用できるのではないかとの提案もある．色々と情報交換ができれば幸いである．

参考資料

[1] <http://sourceforge.net/projects/pentaho/files/>

[2] <http://pentaho-partner.jp/purchase/func>