

データ整備サービス

Powered By

SOFIT *Super* **REALISM**

データ移行はデータ整備サービスで

日本ソフト開発株式会社
BIG DATA R&D本部

日本ソフト開発株式会社 BIG DATA R&D本部です。

データ整備サービスを活用したデータ移行をご紹介します。

データ整備サービスは、超高速データ処理システム「SOFIT Super REALISM」を利用したサービスです。

01

データ移行の課題と解決方法

© 2022 Nippon Software Knowledge corp.

1

データ移行の課題と解決方法について説明します。

マイグレーションにおけるデータ移行 その現状

最後の難関 システム移行 **iSYSTEMS**

第1回 移行は全工数の4割を費やす

2008/03/17
実森 仁志
出典：日経SYSTEMS 2007年5月16ページより
(記事は執筆時の情報に基づいており、現在では異なる場合があります)

0 15 4 連載目次へ>>

シェア フックマーク ツイート 保存する

移行関連で全工数の35~40%

8割以上が移行トラブルを経験

8割以上が移行トラブルを経験

マイグレーションを成功させる大きなポイントはデータ移行!!

難易度が高いだけに、移行のプロセスでトラブルが生じることは多い。本誌読者100人を対象にしたアンケートでは、「移行でトラブルが生じたことがない」と回答したのは有効回答72件のうち2割弱に過ぎない。残る8割強は何らかのトラブルに遭っている。

単純なサーバー・マシンのリプレースから、データの抽出作業を頼んだ新システム構築に匹敵する工数がかかる。みずほ情報総研の調査によると、システムをSAP R/3で再構築する際にかかる工数を工程別に概算した(図1)。その割合を見ると、移行関連の工数がいかに大きなウェートを占めているかが分かる。

ケース3: SIベンダーのSEである陣内丈一郎氏(仮名)は、移行手順書の記載通りに作業を実施したところ、移行手順書の記述内容に遅れが

© 2022 Nippon Software Knowledge corp. 2

システムマイグレーションを実施される企業が多くあります。
マイグレーションにおけるデータ移行は軽視されがちですが、その現状としては、
移行関連でマイグレーション全体工数の35~40%を占めると言われています。
また、実に8割以上が移行トラブルを経験していると言われています。
つまり、マイグレーションを成功させる大きなポイントはデータ移行になります。

データ移行での失敗事例

8割以上が移行トラブルを経験…その内容は

ケース1

ドキュメントに記載された現行システムの仕様に基づいて移行に臨んだが、**現行システムはドキュメントとまったく異なる仕様**であることが分かり、そのままでは移行困難であるとの結論に達した。スケジュールを仕切り直すしかなかった。

ケース2

現行システムの開発ベンダーに移行データの抽出作業を頼んだが法外な見積もりを提示された。やむなく自分でデータを抽出することにしたが、**現行システムの仕様把握に時間**がかかり、移行延期を余儀なくされた。

ケース3

抽出データによる移行リハーサルを重ねて本番に臨んだが、**移行対象データには特殊なデータが含まれており**障害を引き起こした。

共通原因：移行対象となるデータを把握できていない

© 2022 Nippon Software Knowledge corp.

3

データ移行での失敗事例をいくつか紹介します。

ドキュメントに記載された現行システムの仕様に基づいて移行に臨んだが、現行システムはドキュメントとまったく異なる仕様であることが分かり、そのままでは移行困難であるとの結論に達した。スケジュールを仕切り直すしかなかった。

現行システムの開発ベンダーに移行データの抽出作業を頼んだが法外な見積もりを提示された。やむなく自分でデータを抽出することにしたが、現行システムの仕様把握に時間がかかり、移行延期を余儀なくされた。

抽出データによる移行リハーサルを重ねて本番に臨んだが、移行対象データには特殊なデータが含まれており障害を引き起こした。

いずれの場合も、移行対象となるデータを把握できていないことが主原因であり、全件データの内容を確認することで、ドキュメントと現状の差異を明確にした上で、データ移行に望む必要があります。

データ移行の課題とSOFIT Super REALISMによる解決



データ移行の流れは、

従来手法では、データ確認、移行仕様作成、移行PG作成、サンプルデータでのテスト、そして「全件データ」でのテストという手順になり、

初めのデータ確認がコンサルティング、移行仕様作成から移行システム構築に分けられます。ところが、

コンサルティング作業には、多大な時間とコストが必要となるため、ドキュメントとサンプルデータによる確認にとどまることが多く、不十分なデータ確認の状態に移行システム構築へと進みます。

その結果、イレギュラーデータによる手戻りや気付かないテーブルの存在など多くの問題が発生し、移行仕様から見直すこととなります。

SOFIT Super REALISMを活用したデータ移行作業の流れを説明します。

はじめにドキュメントをもとに移行仕様作成、次にデータ確認、移行仕様修正、「全件データ」を用いたデータ処理という手順となり、工程そのものが従来手法とは異なります。大きな違いは、プログラミングではなく、「全件データ」を用いたデータ処理で移行の仕組みを構築することです。

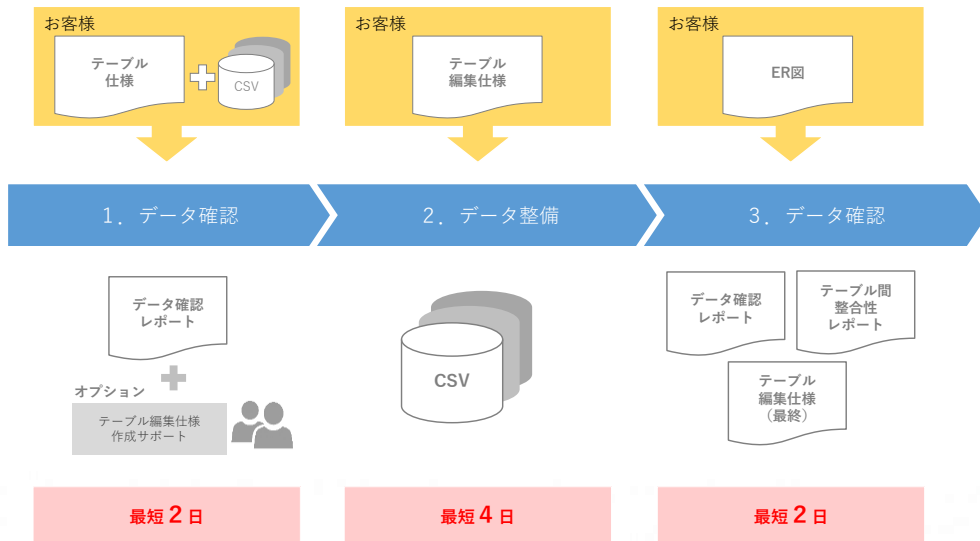
SOFIT Super REALISMを利用することで、短期間で「全件データ」の確認が可能となり、従来手法では6ヶ月必要であったデータ確認を1ヶ月で実現しました。

ノンプログラミングで「全件データ」を処理しながら処理手順を定義していきます。超高速処理を実現していますので、従来手法では10時間必要な処理を1時間に短縮することができました。

作成した移行データの確認を行うことで、より精度の高いデータの提供を実現します。

このように、十分なデータ確認と超高速処理で確実なデータ移行を実現します。

データ整備サービスは、REALISMを用いることでデータ確認やデータ整備を短期間で実施するサービスです。



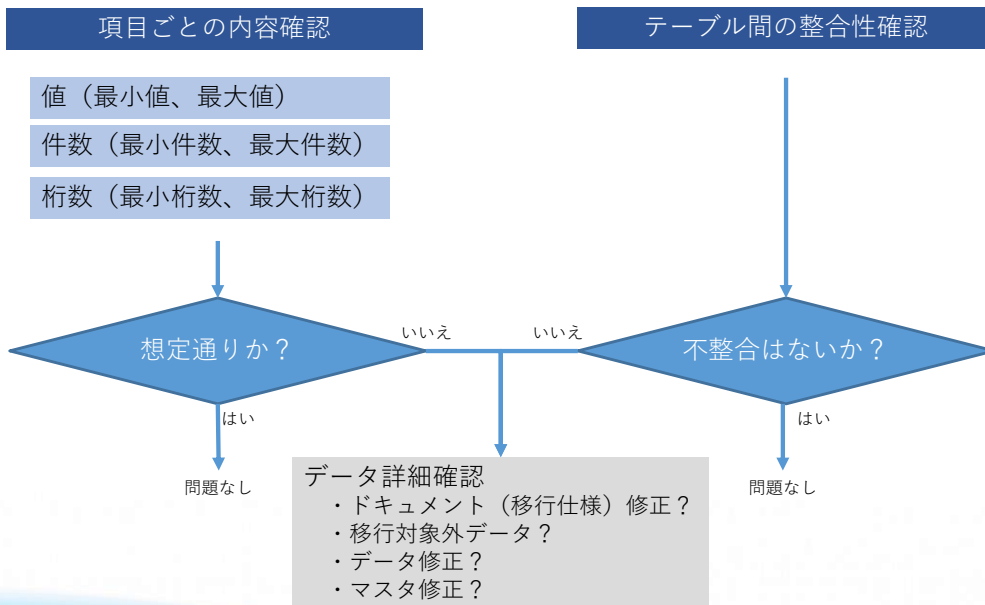
データ整備サービスの流れをご説明します。

まず「データ確認：データ概観」で、
お客様からCSVファイルとテーブル仕様を預かり、データ確認レポートを最短2日で作成します。
お客様によるデータ整備のためのテーブル編集仕様の作成が難しい場合は、作成サポートも行います。

次に「データ整備」で、
お客様が作成したテーブル編集仕様をもとに、
お預かりしたCSVファイルを整備/加工して、最短4日でCSVファイルをお返しします。

最後に「データ確認」で、
データ整備後のCSVファイルがテーブル編集仕様に沿ったデータであることをデータ確認レポートで
確認します。
ER図をいただければ、テーブル間整合性レポートも作成します。

データ確認すべきこと



© 2022 Nippon Software Knowledge corp.

6

データ確認では、どのようなことをすべきか説明します。
まずは、項目ごとの内容、データ概観および基本統計量を調べます。

具体的には、各テーブル各項目に格納されている値 件数 桁数を調べることです。

そして、もう一つはテーブル間の整合性を確認することです。
主に明細データの内容がマスタに存在しているかを調べます。
特に項目ごとの内容確認は必ず実行する必要があります。

確認結果を項目の特性に応じてチェックするとともに、テーブル間に不整合がないか確認します。
確認結果が想定通りではない場合やテーブル間に不整合がある場合にはデータの詳細確認を行います。
ドキュメントが誤っている場合には、ドキュメントを訂正します。
状況によっては移行対象外のデータの場合もあるでしょう。
あるいは、データそのものが誤っており修正を行うべき場合もあります。
テーブル間に不整合がある場合にはマスタの修正が必要な場合もあります。

データ確認レポート

テーブル名称	セーレスマン
項目数	9
レコード件数	1,000
文字コード	MS932

CSVファイルを「テーブル内容確認ツール」で読み込み、マクロ作成し実行すると、項目ごとに、「最小値、最大値」「最小件数、最大件数」「最小桁数、最大桁数」を素早く簡単に確認できる。

主な着眼点

No	1	2	3	4	5	6	9
項目名称	セーレスマンID	名前	氏名	名	年齢	性別	地方
型式	文字列	文字列	文字列	文字列	数字	文字列	文字列
値	値	値	値	値	値	値	値
件数	件数	件数	件数	件数	件数	件数	件数
最小値	1 セーレスマンID-0001	1	2	2 <NULL>	3	18 23 女	67
	2 セーレスマンID-0002	1 阿形 薫	1 阿形 薫	1 あけみ	2	19 23 男	933
	3 セーレスマンID-0003	1 阿部 幸子	1 阿部 幸子	3 まな	1	20 24	
	4 セーレスマンID-0004	1 阿部 秀子	1 愛川 さとみ	1	21 28		
	5 セーレスマンID-0005	1 阿部 隆夫	1 栗野 はいり	1	22 29		
最大値	1 セーレスマンID-489	1 森 栄治	1 森 栄治	1 豊二	1	1899 1 男	933
	2 セーレスマンID-1000	1 齋藤 和義	1 齋藤 和義	1	60 22 女	67	
	3 セーレスマンID-0999	1 齋藤 友	1 齋藤 友	1 廣末	1	59 22	
	4 セーレスマンID-0998	1 齋藤 信一	1 齋藤 信一	1 初部	1	58 28	
	5 セーレスマンID-0997	1 齋藤 健太	1 楠山 隆洋	1	57 21		
最小件数	1 セーレスマンID-0001	1 阿形 薫	1 阿形 薫	1 さかな	1	1899 1 女	67
	2 セーレスマンID-0002	1 阿形 薫	1 阿形 薫	1 さとみ	1	27 12 男	933
	3 セーレスマンID-0003	1 阿部 幸子	1 栗野 はいり	1	41 16		
	4 セーレスマンID-0004	1 阿部 秀子	1 安西 ひとみ	1	25 17		
	5 セーレスマンID-0005	1 阿部 隆夫	1 安藤 まゆみ	1	23 18		
最大件数	1 セーレスマンID-0516	2 鈴木 真雄	2 鈴木 真雄	2 明	7	38 32 男	933
	2 セーレスマンID-489	1 中山 隆	2 高橋 20	7	29 32 女	67	
	3 セーレスマンID-0999	1 初部	2 青木 9	6	48 30		
	4 セーレスマンID-0998	1 守村 弘之	2 山田 9	6	47 30		
	5 セーレスマンID-0997	1 吉井 和彦	2 伊藤 9	6	46 30		
最小桁数	1 セーレスマンID-489	12	0	0	0	18 2 女	2
	2 セーレスマンID-0001	13 小沢	4 岡	2 綾	2	19 2 男	2
	3 セーレスマンID-0002	13 岡 寛	6 角	2 宗	2	20 2 女	
	4 セーレスマンID-0003	13 岡 正	6 藤	2 雅	2	21 2 女	
	5 セーレスマンID-0004	13 岡 明	6 岡	2 海	2	22 2 女	
最大桁数	1 セーレスマンID-054111	15 佐々木 浩一郎	14 木佐貫	6 雄一郎	6	1899 4 男	2
	2 セーレスマンID-1000	13 坂田 さとみ	14 木ノ内	6 由比子	6	60 2 女	2
	3 セーレスマンID-0999	13 坂尾 伊織	12 西島田	6 由香子	6	59 2 女	
	4 セーレスマンID-0998	13 坂尾 伊織	6 勇太郎	6	58 2 女		
	5 セーレスマンID-0997	13 坂尾 伊織	6 勇太郎	6	57 2 女		
備考	キー項目の確認	名前の登録状況	氏の登録状況	名の登録状況	年齢の登録状況	性別の登録状況	地方の登録状況

© 2022 Nippon Software Knowledge corp.

7

データ確認レポートについて説明します。

データの確認には、「テーブル内容確認ツール」を利用します。
 確認したいテーブル名を指定すると、REALISM上にデータ確認処理が保存されます。
 この処理を実行すると、このような「データ確認レポート」が出力されます。

項目ごとに、

「最小値、最大値」で、名前が未登録のデータや、年齢が1899と不正なデータが存在する等のデータ不備

「最小件数、最大件数」で、同じセールスマンIDが2つ存在している等のキーの重複。

「最小桁数、最大桁数」で、セールスマンIDの桁不足や桁あふれ等のデータ不備を素早く簡単に確認することができます。

データ確認対処例

No	2		5	
項目名称	名前		年齢	
型式	文字列		数字	
	値	件数	値	件数
値数		993		44
最小値	1		2	18
	2	阿形 富男	1	19
	3	阿部 幸子	1	20
	4	阿部 秀子	1	21
	5	阿部 隆夫	1	22
最大値	1	藪 栄治	1	1899
	2	齋藤 和義	1	60
	3	齋藤 友	1	59
	4	齋藤 信一	1	57
	5	齋藤 健太	1	57

データ不備 (名前が未登録のデータが2件ある)

データ不備 (年齢が1899と不正なデータが存在する)



データ特定 ⇒ 修正

テーブル編集仕様 (一部抜粋)

項目名	型	必須or任意	詳細仕様
名前	文字列	必須 (NULL値不可)	「名前」がNULL値の場合、「不明」を設定
年齢	整数	任意 (NULL値可)	「年齢」が100以上の場合、NULL値を設定

© 2022 Nippon Software Knowledge corp.

8

データ確認結果からの対処例です。

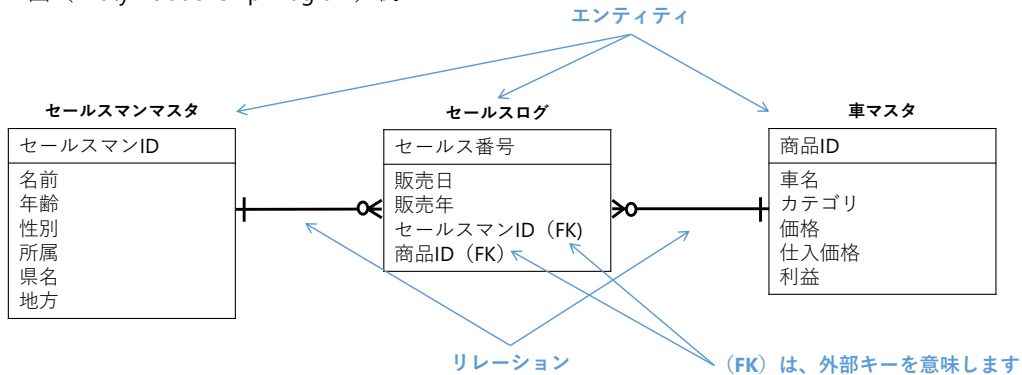
データ確認の結果、名前が未登録のデータが2件、年齢が1899と明らかに不正なデータが1件存在することがわかりました。

該当データを特定し、データ修正を行うことが原則ですが、調査に時間が掛かることも想定されるため、

テーブル編集仕様に、名前がNULL値の場合は「不明」という文字を設定する、年齢が100以上の場合はNULL値を設定するという詳細仕様を追記して、データ移行を実施しました。

テーブル間整合性確認 (ER図)

ER図 (Entity Relationship Diagram) 例



リレーション書き方例



1対1



1対多 (0以上)



1対多 (1以上)

※ER図にはIE記法とIDEFIX記法がありますが、上記はIE記法を紹介しています。

© 2022 Nippon Software Knowledge corp.

9

ER図がどのようなものか簡単に紹介します。

ER図とはデータベース設計における代表的な設計図のことです。

これが、車販売のER図になります。

ER図のEはエンティティ (Entity) の略で、Rはリレーションシップ (Relationship) の略であり、データ (テーブル) 間の関係性を表す図のことです。

リレーションは、1対1、1対多0以上、1対多1以上など関係性によって書き方が定められています。データ整備サービスでは、このようなER図をお預かりすることで、ER図から外れたデータが存在しないかを確認するテーブル間の整合性確認を行います。

データ移行の事例を説明します。

データ整備「自治体データ移行」

■ データ移行構築

- ・従来方法：COBOL：62本



- ・REALISM：105本



■ データ移行処理

- ・従来方法



- ・REALISM



© 2022 Nippon Software Knowledge corp.

11

自治体様でのデータ移行の事例です。

汎用機からオープン系へのマイグレーションに伴うデータ移行で、当初はCOBOLで移行システムを構築する予定でしたが、REALISMを利用することで約4人月の工数削減ができました。また、REALISMでは「実データ全件」を使った処理構築を行うので、事前にイレギュラーデータも把握しており、その品質は非常に高いものになります。

データ移行処理では、本番データ移行に向け、毎月テスト移行を計9回繰り返し行うことで、移行精度を上げ本番移行を安全に迎える計画でしたが、汎用機上では1回あたり10時間掛かる処理が、REALISMを使うことで2時間で終わることができましたので、移行処理時間も80時間削減することができました。

また、10時間掛かる処理が2時間で終わることができるので、本番移行時に不測の事態が発生した場合にも再処理時間を確保できる余裕を持ったスケジュールを立てることができました。

データ整備「医療DWH」

■ データ移行：170床 10年分

- ・従来方法：データ連携機能



■ 実績

病院名	SM	TG	KF	TS	HN	SU	YD	TD
病床数	170	1,407	1,065	476	555	426	737	1,154
稼働年月	2013年08月	2014年01月	2014年01月	2014年09月	2014年11月	2014年11月	2015年04月	2016年01月
検歴データ	○	○	○	○	○	○	○	○
看護経過表	○		○	○	○	○	○	○
XMLINDEX	○		○			○	○	○
オータ数	19		14	20	17	16	20	
処理日数	4日	3日	16日	5日	5日	4日	4日	3日

© 2022 Nippon Software Knowledge corp.

12

こちらは医療の電子カルテシステムリプレースに伴うDWHへのデータ移行の事例になります。

従来は、稼働後に利用するデータ連携機能を使って「当初データ」も移行をされておりました。確かにその精度は高く安心できるのですが、処理速度が遅く170床10年分のデータを移行するには4.5ヶ月必要、つまり電子カルテシステム稼働後4.5ヶ月後にならないとDWH側に全てのデータが揃わないという状況で、大きな病院になると1年以上かかっていました。

そこで、REALISMを使って「当初データ」を短期間で全て移行する「高速データ移行サービス」として展開させていただきました。おおむね半年程度かかるところを5日程度で終えられ、DWHの早期稼働を実現することができました。

 **日本ソフト開発株式会社**
Nippon Software Knowledge corp.



日本ソフト開発株式会社
【SOFIT Super REALISM】
YouTubeチャンネル



DX推進サービス
SOFIT Super REALISM
Webサイト



データ整備サービス
Webサイト

□ 本社 BIG DATA R&D本部
〒521-0015
米原市米原西23番地
TEL : 0749-52-3811 (代)
E-mail : realism@nihonsoft.co.jp

□ 東京支店
〒105-0013
港区浜松町2-7-17
イーグル浜松町ビル10階
TEL : 03-5473-0036 (代)

□ 大阪支店
〒532-0003
大阪市淀川区宮原4-1-14
住友生命新大阪北ビル8階
TEL : 06-6350-3505 (代)

以上で、データ整備サービスを活用したデータ移行の紹介を終わります。