

Informatica Data Quality (DQ)

全社レベルでデータ品質を管理する包括的なアプローチを実現

『Informatica Data Quality』を採用することで、データ品質の問題を発見・解消し、データから最大の価値を獲得します。

データの品質課題とは？

- ✓ データ品質は、完全性、適合性、関連性、一貫性、重複度などさまざまな観点で評価し、正確に課題を把握する必要があります。
- ✓ 外部データなどの信頼できるソースであっても、「株式会社」と「株」や電話番号等の表記フォーマットが異なる可能性があるためマッピング(名寄せ)の精度を高めるためにも、事前にプロファイリングし課題を把握、各課題に応じたクレンジングをすることが重要です。

【データ品質課題例】

顧客ID	会社ID	姓	名	業種	郵便番号	住所1	2	3	4	電話番号
A010010002	22001	園田	茂樹	20	324-42	栃木県大田原市大豆田3-15-17大田原センタービル				0375515395
A010010003	33001	馬場	清隆	20	370-2311	群馬県富岡市下高尾3-15				0307137216
A010020001	44001	井口	玲子	50	3670106	埼玉県児玉郡美里町南十条3-2-6				0369875175
A010020003	66001	浜崎	登	40	160-0005	東京都新宿区愛住町4-10-5				0323213105
	55001	清水	善次	60	375-0012	群馬県藤岡市下戸塚4-11-3				0387829328
...

顧客IDはプライマリーキーの役目を果たしているはずが、**データ移行時にNot Null 制約に違反**してロードに失敗

業種は10、20のようなID値ではなく、「金融業」「製造業」などの実データとして**業種マスタ表**と連携しなければならなかった。さらに**60の値は業種マスタ表に存在しない**

フォーマットが異なるデータ
正しいフォーマット：999-9999

住所の分かち書きがしたかったが、住所データの突合が技術的にクリアできずに住所データは妥協した

電話番号は10ケタのはずが、**データ移行時に桁あふれ**してロードに失敗

データの品質を高める4つの主要機能

1. プロファイリング

現在のデータの品質を分析・評価



2. クレンジング

ルールを定義し、クレンジング部品でデータを標準化

顧客顧客分類	顧客顧客タイプ
<input type="checkbox"/> Special	1
<input type="checkbox"/> Normal	5
<input type="checkbox"/> Platinum	2
<input type="checkbox"/> Gold	3
<input type="checkbox"/> Silver	4

4. スコアリング

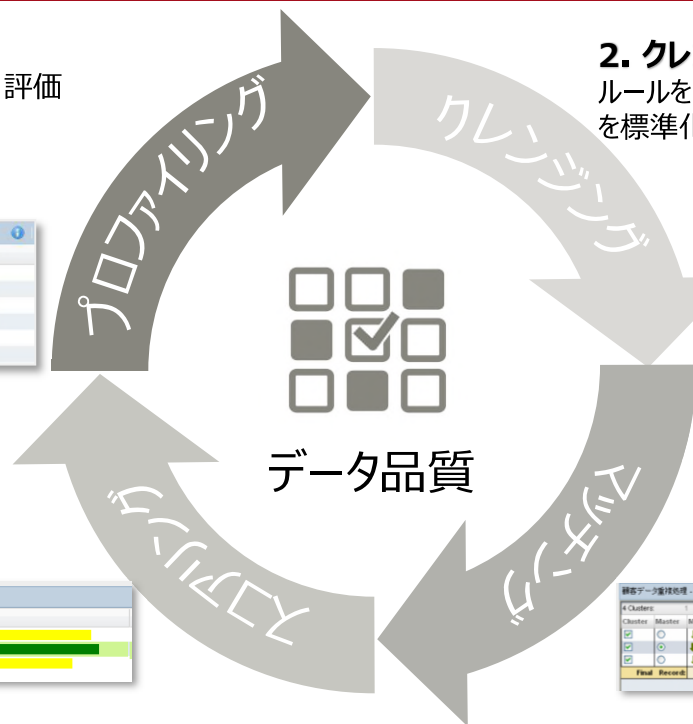
定義したルールに基づき定常的にデータ品質を測定

NAME	SCORE	% VALID CURRENT
CustomerNum_Vaildity	87	
StateCd_Vaildity	90	
ZipCD_Accuracy	80	

3. マッチング

重複レコードの特定とゴールデンレコードの生成

Cluster	Matched	Match Score	Customer ID	Full Name	Address
1	0.99	10001	10001	小宮 聡	埼玉県 さいたま市
2	0.98	11000	11000	小宮 聡	埼玉県 さいたま市
Final Record	0.99	10001	10001	小宮 聡	埼玉県 さいたま市



データ品質を支える高度なクレンジング処理機能

- 単純な数値変換、フォーマット変換では対応できないクレンジング処理を実現します。
- 住所クレンジング、重複データの検出、辞書を利用したデータ変換などに対応しています。

高度なクレンジング例 住所クレンジング

日本の住所クレンジング前

住所
東京都新宿区市谷本村町1-1
福岡県太宰府市国分二の二七

- 住所の分割、修正、整合性検査
- 郵便番号の引当、逆引き対応
- 住所の分かち書き

クレンジング後

郵便番号	都道府県	市区郡	町村	丁目	番地	その他
1620845	東京都	新宿区	市谷本村町		1番地	1号
6340111	福岡県	太宰府市	国分	2丁目	2番地	7号

海外住所クレンジング例

100 Cardinal Street
Ledwood City, CA 94063-4755



100 Cardinal Way
Redwood City, CA 94063-4755

高度なクレンジング例 重複データの検出・排除

- 住所、氏名、生年月日などから同一と思われるデータを識別
- それらのデータを照合して、マージ

ソースデータ1

ID	姓	名	会社名	生年月日	郵便番号	都道府県	市区郡	町村
777	鈴木	一郎	インフォマティカ・ジャパン株式会社		1620845	東京都	新宿区	市谷本村町
11	山田	花子	株式会社 架空	1990年1月1日	8914207	鹿児島県	熊毛郡	屋久島町

ソースデータ2

ID	氏名	会社名	生年月日	郵便番号	都道府県	市区郡	町村
A500	鈴木一郎	インフォマティカ株式会社		1020094	東京都	千代田区	紀尾井町
B333	佐藤花子	株式会社 架空	1990年1月1日	8914207	鹿児島県	熊毛郡	屋久島町



マッチスコア

合計	氏名	会社名	生年月日	郵便番号	都道府県	市区郡	町村
70%	100%	80%	0%	0%	100%	0%	0%
90%	60%	100%	100%	100%	100%	100%	100%

データの品質を分析評価、監視するデータプロファイリングとスコアリング

データ品質管理を実現する上で重要なデータ内容の現状把握、継続的な分析機能を業務部門の方に使いやすいブラウザベースのツールとして提供します。

- プロファイリング : データの中身を分析し、迅速なデータ品質問題解決を支援
- スコアカード : 継続的に高いデータ品質を維持するためのデータ品質監視機能

高度なデータ分析 プロファイリング

データの内容を分析し、問題のあるデータの特定を支援します。またデータ品質ルールを設定することにより、ルールに合致しないデータを特定することを可能にします。

データ品質監視 スコアカード

任意の閾値でデータ品質を監視し、グラフ化することにより、データ品質管理の可視化を実現します。また傾向グラフを確認することで、時間と共に劣化したデータ品質や、クレンジング効果などを把握することが可能です。

富士通の強み

- データ統合関連ビジネスに関して20年以上の実績を有し、経験豊富な技術者が多数在籍
- インフォマティカ再販パートナーとして15年以上のInformaticaソリューション取り扱い実績
- DWH/BI/BAといった情報利活用とともにお客様への最適なソリューションをご提案ご提供可能
- 「Informatica Partner Award」を2017年度、2018年度、2019年度の3年連続で受賞

※記載の会社名、商品名は、各社の商標または登録商標です。
※記載された情報は、予告なく変更することがあります。
※記載の内容は、2020年8月現在のものです。

お問い合わせ先

富士通コンタクトライン（総合窓口）
0120-933-200

受付時間 9:00～12:00 および 13:00～17:30
（土・日・祝日・当社指定の休業日を除く）