



Informatica®  
10.5.6

# データ検出ガイド

## Informatica データ検出ガイド

10.5.6

2024 年 5 月

© 著作権 Informatica LLC 2011, 2024

本ソフトウェアおよびマニュアルは、使用および開示の制限を定めた個別の使用許諾契約のもとでのみ提供されています。本マニュアルのいかなる部分も、いかなる手段（電子的複製、写真複製、録音など）によっても、Informatica LLC の事前の承諾なしに複製または転載することは禁じられています。

Informatica および Informatica ロゴは、米国およびその他の国における Informatica LLC の商標または登録商標です。Informatica の商標の最新リストは、Web (<https://www.informatica.com/trademarks.html>) にあります。その他の企業名および製品名は、それぞれの企業の商標または登録商標です。

オプトアウトの権利の制限の下、本ソフトウェアによって、本ソフトウェアがデプロイされているコンピューティングおよびネットワーク環境に関する情報、デプロイメントのデータ使用状況およびシステム統計情報が米国の Informatica に自動的に送信されます。この送信は Informatica のプライバシーポリシーの下にサービスの一部と見なされ、Informatica は、<https://www.informatica.com/in/privacy-policy.html> のプライバシーポリシーに従って、この情報を使用または処理します。使用状況の収集は、Administrator ツールで無効にすることができます。

米政府の権利プログラム、ソフトウェア、データベース、および関連文書や技術データは、米国政府の顧客に配信され、「商用コンピュータソフトウェア」または「商業技術データ」は、該当する連邦政府の取得規制と代理店固有の補足規定に基づきます。このように、使用、複製、開示、変更、および適応は、適用される政府の契約に規定されている制限およびライセンス条項に従うものとし、政府契約の条項によって適当な範囲において、FAR 52.227-19、商用コンピュータソフトウェアライセンスの追加権利を規定します。

本ソフトウェアまたはドキュメンテーション（あるいはその両方）の一部は、第三者が保有する著作権の対象となります。必要な第三者の通知は、製品に含まれています。

特許については、<https://www.informatica.com/legal/patents.html> を参照してください。

本マニュアルの情報は、予告なしに変更されることがあります。このドキュメントで問題が見つかった場合は、[infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com) までご報告ください。

Informatica 製品は、それらが提供される契約の条件に従って保証されます。Informatica は、商品性、特定目的への適合性、非侵害性の保証等を含めて、明示的または黙示的ないかなる種類の保証をせず、本マニュアルの情報を「現状のまま」提供するものとします。

発行日: 2024-09-26

# 目次

<b>序文</b> .....	13
Informatica のリソース.....	13
Informatica Network.....	13
Informatica ナレッジベース.....	13
Informatica マニュアル.....	14
Informatica 製品可用性マトリックス.....	14
Informatica Velocity.....	14
Informatica Marketplace.....	14
Informatica グローバルカスタマサポート.....	14
 <b>第 1 部 : データ検出の概要</b> .....	15
 <b>第 1 章 : プロファイリングの概要</b> .....	16
プロファイリングの概要.....	16
プロファイリングアーキテクチャ.....	17
データ検出プロセス.....	19
 <b>第 2 章 : データ検出</b> .....	21
データ検出の概要.....	21
プロファイルおよび分析タイプ.....	21
プロファイリングのコンポーネント.....	22
プロファイル結果.....	23
 <b>第 3 章 : カラムプロファイルの概念</b> .....	25
カラムプロファイルの概念の概要.....	25
カラムプロファイルオプション.....	26
リポジトリプロファイルのロックおよびバージョン管理されたプロファイルの管理.....	26
スコアカード.....	27
 <b>第 4 章 : データドメイン検出の概念</b> .....	28
データドメイン検出の概念の概要.....	28
データドメイン.....	29
データドメイングループ.....	29
データドメイングロッサリ.....	30
データドメイン検出のプロセス.....	30
Spark エンジンでのデータドメイン検出.....	30
 <b>第 5 章 : キュレーションの概念</b> .....	32
キュレーションの概念の概要.....	32
解析者と開発者のキュレーション.....	32

キューレーションタスク.....	33
<b>第 II 部 : Informatica Analyst を使用したデータ検出.....</b>	<b>35</b>
<b>第 6 章 : Informatica Analyst のカラムプロファイル.....</b>	<b>36</b>
Informatica Analyst のカラムプロファイルの概要.....	36
カラムプロファイリングプロセス.....	37
プロファイルオプション.....	37
サンプリングのオプション.....	38
ドリルダウンのオプション.....	39
ランタイム環境.....	39
ネイティブ環境.....	39
Hadoop 環境.....	39
Informatica Analyst のオペレーティングシステムプロファイルの概要.....	40
オペレーティングシステムプロファイルの選択.....	41
リポジトリアセットのロックとチームベース開発の概要.....	41
Informatica Analyst でのカラムプロファイルの作成.....	42
カラムプロファイルの編集.....	43
プロファイルの実行.....	44
Spark エンジンでのプロファイルの実行.....	44
同期オプション.....	44
Informatica Analyst でのフラットファイルデータオブジェクトの同期.....	45
Informatica Analyst でのリレーショナルデータオブジェクトの同期.....	47
<b>第 7 章 : Informatica Analyst のルール.....</b>	<b>48</b>
Informatica Analyst のルールの概要.....	48
定義済みルール.....	48
定義済みルールのプロセス.....	49
定義済みルールの適用.....	49
式ルール.....	50
式ルールの作成.....	50
ルール仕様を使用した式ルールの作成.....	51
<b>第 8 章 : Informatica Analyst のフィルタ.....</b>	<b>53</b>
Informatica Analyst のフィルタの概要.....	53
フィルタの作成.....	53
簡易フィルタの作成.....	54
詳細フィルタの作成.....	55
SQL フィルタの作成.....	56
フィルタの管理.....	56
<b>第 9 章 : Informatica Analyst のカラムプロファイル結果.....</b>	<b>58</b>
Informatica Analyst のカラムプロファイル結果の概要.....	58

サマリビュー.....	59
サマリビューのプロパティ.....	60
サマリビューのデフォルトフィルタ.....	60
詳細ビュー.....	61
[詳細ビュー] ペイン.....	62
統計.....	63
データプレビュー.....	64
データ型.....	64
異常値.....	65
パターン.....	66
値.....	67
プロファイル実行のタイプ.....	70
最新のプロファイル実行.....	70
履歴プロファイル実行.....	70
統合済みプロファイル実行.....	70
プロファイル実行の選択.....	71
複数のプロファイル結果の比較の概要.....	72
複数のプロファイル結果の比較.....	72
プロファイル結果の比較のサマリビュー.....	73
プロファイル結果の比較の詳細ビュー.....	75
カラムプロファイルのドリルダウン.....	76
行データのドリルダウン.....	76
ドリルダウンデータへのフィルタの適用.....	76
Analyst ツールでのキュレーション.....	77
データ型とデータドメインの承認.....	77
データ型とデータドメインの拒否.....	77
Informatica Analyst のカラムプロファイルのエクスポートファイル.....	78
CSV ファイルのプロファイルエクスポート結果.....	78
Microsoft Excel のプロファイルエクスポート結果.....	78
Informatica Analyst からのプロファイル結果のエクスポート.....	79
<b>第 10 章 : Informatica Analyst のビジネス用語、コメント、タグ.....</b>	<b>80</b>
Informatica Analyst でのビジネス用語、コメント、タグの概要.....	80
ビジネス用語.....	80
ビジネス用語のカラムへの割り当て.....	81
コメント.....	81
プロファイルまたはカラムへのコメントの追加.....	81
タグ.....	82
プロファイルまたはカラムへのタグの追加.....	82
<b>第 11 章 : Informatica Analyst のスコアカード.....</b>	<b>83</b>
Informatica Analyst のスコアカードの概要.....	83
Informatica Analyst のスコアカードのプロセス.....	84

Informatica Analyst でのスコアカードの作成. . . . .	85
既存のスコアカードへのカラムの追加. . . . .	86
既存のスコアカードへカラムの追加. . . . .	86
スコアカードの実行. . . . .	87
スコアカードの表示. . . . .	88
スコアカードの編集. . . . .	88
メトリック. . . . .	89
メトリックウェイト. . . . .	89
データ品質の値. . . . .	89
しきい値の定義. . . . .	90
[メトリックグループ] . . . . .	90
メトリックグループの作成. . . . .	91
メトリックグループへのスコアの移動. . . . .	91
メトリックグループの編集. . . . .	91
メトリックグループの削除. . . . .	92
カラムのドリルダウン. . . . .	92
傾向グラフ. . . . .	93
スコアの傾向グラフ. . . . .	93
コストの傾向グラフ. . . . .	93
傾向グラフの表示. . . . .	94
傾向グラフのエクスポート. . . . .	95
Informatica Analyst のスコアカードダッシュボード. . . . .	96
プロジェクト別のスコアカード. . . . .	97
スコアカード実行傾向. . . . .	98
スコアカード付きデータオブジェクト. . . . .	99
累積メトリック傾向. . . . .	100
Informatica Analyst のスコアカードエクスポートファイル. . . . .	102
Informatica Analyst からのスコアカード結果のエクスポート. . . . .	102
Microsoft Excel のスコアカードエクスポート結果. . . . .	103
スコアカード通知. . . . .	103
通知電子メールメッセージテンプレート. . . . .	104
スコアカード通知の設定. . . . .	105
スコアカード通知のグローバル設定の実行. . . . .	105
スコアカードリネージュ. . . . .	106
Informatica Analyst でのスコアカードリネージュの表示. . . . .	106
<b>第 12 章 : Informatica Analyst でのデータドメイン検出. . . . .</b>	<b>107</b>
Informatica Analyst でのデータドメイン検出の概要. . . . .	107
Informatica Analyst のデータドメインングロッサリ. . . . .	107
Informatica Analyst でのデータドメイングループの作成. . . . .	108
Informatica Analyst でのデータドメインの作成. . . . .	108
Informatica Analyst でのプロファイル結果からのデータドメインの作成. . . . .	109
Informatica Analyst でのデータドメインとデータドメイングループの検索. . . . .	109

Informatica Analyst のデータドメイン検出オプション. . . . .	110
Informatica Analyst でのデータドメインカラムの選択. . . . .	110
Informatica Analyst でのデータドメインの選択. . . . .	111
Informatica Analyst のデータドメイン推測オプション. . . . .	111
Informatica Analyst でのデータドメイン検出実行カラムプロファイルの作成. . . . .	113
Informatica Analyst でのカラムプロファイルとデータドメイン検出の編集. . . . .	114
データドメイン検出実行プロファイルの実行. . . . .	115
Informatica Analyst のデータドメイン検出結果. . . . .	115
データドメインの承認. . . . .	115
データドメインの却下. . . . .	116
Informatica Analyst のデータドメイン検出エクスポートファイル. . . . .	116
Microsoft Excel でのデータドメイン検出結果. . . . .	116
Informatica Analyst からのデータドメイン検出結果のエクスポート. . . . .	117
<b>第 13 章 : Informatica Analyst でのエンタープライズ検出. . . . .</b>	<b>118</b>
Informatica Analyst でのエンタープライズ検出の概要. . . . .	118
Informatica Analyst でのエンタープライズ検出プロセス. . . . .	119
エンタープライズ検出の設定オプション. . . . .	119
データドメイン検出の設定. . . . .	119
カラムプロファイル設定. . . . .	120
Informatica Analyst でのエンタープライズ検出プロファイルの作成. . . . .	121
エンタープライズ検出のオプションの編集. . . . .	122
<b>第 14 章 : Informatica Analyst でのエンタープライズ検出の結果. . . . .</b>	<b>124</b>
Informatica Analyst でのエンタープライズ検出結果の概要. . . . .	124
サマリビュー. . . . .	124
サマリビューのプロファイル結果. . . . .	125
データドメイン検出結果の表示. . . . .	126
カラムプロファイルの結果の表示. . . . .	126
データ型の競合. . . . .	127
データ型の競合の表示. . . . .	127
プロファイルビュー. . . . .	127
プロファイルプロパティの確認. . . . .	128
<b>第 15 章 : Informatica Analyst での検出検索. . . . .</b>	<b>129</b>
Informatica Analyst の概要での検出検索. . . . .	129
検出検索の前提条件. . . . .	130
Informatica Analyst での検出検索プロセス. . . . .	130
検出検索オプション. . . . .	130
検出検索基準. . . . .	131
アセットの選択. . . . .	132
Informatica Analyst での検出検索結果. . . . .	132
[検出検索結果] パネル. . . . .	132

検出検索結果のフィルタリング.....	133
一致タイプ.....	134
直接一致.....	134
間接一致.....	134
一致情報の表示.....	134
検出検索結果からアセットを開く.....	135
関連するアセット.....	135
各アセットタイプの関連アセット.....	135
関連アセットの表示.....	136
FAQ（よくある質問）.....	136
<b>第 16 章 : Informatica Analyst での Business Glossary Desktop.....</b>	<b>138</b>
ビジネス用語.....	138
Metadata Manager ビジネス用語集でビジネス用語を管理.....	139
ビジネス用語集デスクトップでのビジネス用語のルックアップ.....	139
<b>第 III 部 : Informatica Developer を使用したデータ検出.....</b>	<b>140</b>
<b>第 17 章 : Informatica Developer のプロファイル.....</b>	<b>141</b>
Informatica Developer のプロファイルの概要.....	141
Informatica Developer のプロファイルのビュー.....	142
リポジトリオブジェクトのロックおよびバージョン管理されたオブジェクトを使用したチームベ ース開発.....	143
<b>第 18 章 : データオブジェクトプロファイル.....</b>	<b>144</b>
データオブジェクトプロファイルの概要.....	144
Informatica Developer のカラムプロファイル.....	145
フィルタリングのオプション.....	145
サンプリングのオプション.....	147
ランタイム環境.....	148
ネイティブ環境.....	148
Hadoop 環境.....	148
プライマリキー検出.....	149
プライマリキー推測のプロパティ.....	149
推測されるプライマリキーのプロパティ.....	150
キー違反のプロパティ.....	150
機能依存性検出.....	151
機能依存性推測のプロパティ.....	151
推測される機能依存性のプロパティ.....	151
機能依存性違反のプロパティ.....	152
Informatica Developer のオペレーティングシステムプロファイル.....	152
オペレーティングシステムプロファイルの選択.....	152
Informatica Developer での単一のデータオブジェクトプロファイルの作成.....	153



Informatica Developer での複数のデータオブジェクトプロファイルの作成. . . . .	153
プロファイルの編集. . . . .	154
同期オプション. . . . .	155
Informatica Developer でのフラットファイルデータオブジェクトの同期. . . . .	155
Informatica Developer でのリレーショナルデータオブジェクトの同期. . . . .	157
コメント. . . . .	158
Informatica Developer でのコメントの追加. . . . .	158
<b>第 19 章 : 半構造化データソースのカラムプロファイル. . . . .</b>	<b>159</b>
半構造化データソースのカラムプロファイルの概要. . . . .	159
JSON および XML データオブジェクト. . . . .	160
JSON または XML データソースからのデータオブジェクトの作成. . . . .	160
HDFS 内の半構造化データソースの複合型ファイルデータオブジェクト. . . . .	161
HDFS 内の JSON または XML データソースからの複合型ファイルデータオブジェクト. . . . .	161
HDFS 内の Avro または Parquet データソースからの複合型ファイルデータオブジェクト. . . . .	161
HDFS 接続の作成. . . . .	162
HDFS 内の JSON または XML ファイルからの複合型ファイルデータオブジェクトの作成. . . . .	162
Avro または Parquet データソースからの複合型ファイルデータオブジェクトの作成. . . . .	163
半構造化データソースでのカラムプロファイルの作成. . . . .	164
<b>第 20 章 : Informatica Developer のルール. . . . .</b>	<b>166</b>
Informatica Developer のルールの概要. . . . .	166
Informatica Developer でのルールの作成. . . . .	167
Informatica Developer でのルールの適用. . . . .	167
<b>第 21 章 : マプレットとマッピングのプロファイリング. . . . .</b>	<b>168</b>
マプレットとマッピングのプロファイリングの概要. . . . .	168
マプレットまたはマッピングオブジェクトに対するプロファイルの実行. . . . .	168
マッピングまたはマプレットオブジェクトのプロファイルの比較. . . . .	169
プロファイルからのマッピングの生成. . . . .	169
<b>第 22 章 : Informatica Developer のカラムプロファイル結果. . . . .</b>	<b>171</b>
Informatica Developer のカラムプロファイル結果. . . . .	171
カラム値のプロパティ. . . . .	172
カラムパターンのプロパティ. . . . .	172
カラム統計のプロパティ. . . . .	172
カラムデータ型のプロパティ. . . . .	173
Informatica Developer でのキューレーション. . . . .	174
データ型の承認. . . . .	174
データ型の拒否. . . . .	174
Informatica Developer からのプロファイル結果のエクスポート. . . . .	175

<b>第 23 章 : Informatica Developer のスコアカード.....</b>	<b>176</b>
Informatica Developer のスコアカードの概要.....	176
スコアカードの作成.....	176
スコアカードリネージュのリソースファイルのエクスポート.....	177
Informatica Developer からのスコアカードリネージュの表示.....	177
 <b>第 24 章 : Informatica Developer でのデータドメイン検出.....</b>	 <b>179</b>
Informatica Developer でのデータドメイン検出の概要.....	179
Informatica Developer のデータドメイングロッサリ.....	180
Informatica Developer でのデータドメイングループの作成.....	180
Informatica Developer でのデータドメインの作成.....	180
Informatica Developer でのプロファイル結果からのデータドメインの作成.....	181
Informatica Developer でのデータドメインの検出.....	182
データドメインのインポート.....	182
データドメインのエクスポート.....	183
Informatica Developer のデータドメイン検出オプション.....	183
Informatica Developer でのデータドメインの選択.....	184
Informatica Developer でのデータドメインカラムの選択.....	184
Informatica Developer のデータドメイン推測オプション.....	185
Informatica Developer でのデータドメイン検出実行プロファイルの作成.....	186
Informatica Developer でのプロファイルの編集.....	187
Informatica Developer でのデータドメイン検出実行プロファイルの実行.....	187
Informatica Developer のデータドメイン検出結果.....	188
データドメイングループ別の表示.....	188
カラム別の表示.....	189
結果の検証.....	189
データドメインの承認.....	189
データドメインの却下.....	190
Informatica Developer からのデータドメイン検出結果のエクスポート.....	190
 <b>第 25 章 : Informatica Developer でのエンタープライズ検出.....</b>	 <b>191</b>
Informatica Developer でのエンタープライズ検出の概要.....	191
エンタープライズ検出プロセス.....	192
エンタープライズ検出のプロファイルオプション.....	193
エンタープライズ検出でのデータドメインの選択.....	193
エンタープライズ検出のカラムプロファイルのサンプリングオプション.....	194
ランタイム環境オプション.....	194
エンタープライズ検出のプライマリキー推測オプション.....	195
エンタープライズ検出の外部キー推測オプション.....	195
外部キー推測の自動キュレーションパラメータ.....	196
Informatica Developer でのエンタープライズ検出プロファイルの作成.....	198
プロファイルの編集.....	199

エンタープライズ検出プロファイルの実行. . . . .	200
外部キー検出. . . . .	201
オブジェクトの親子関係の定義. . . . .	201
データオブジェクト間の外部キー関係の検出. . . . .	201
外部キー分析結果. . . . .	202
結合分析. . . . .	202
結合プロファイルの作成. . . . .	202
結合分析結果. . . . .	203
結合プロファイル結果のファイルへのエクスポート. . . . .	204
重複検出. . . . .	204
重複検出の結果. . . . .	204
重複データの検出. . . . .	205
DDL スクリプトファイル. . . . .	206
エンタープライズ検出プロファイルからの DDL スクリプトの作成. . . . .	206
エンタープライズ検出プロファイルの同期. . . . .	206
エンタープライズ検出プロファイルの同期. . . . .	206
<b>第 26 章: エンタープライズ検出結果. . . . .</b>	<b>208</b>
エンタープライズ検出結果の概要. . . . .	208
[リレーション] ビュー. . . . .	209
データオブジェクトの検索. . . . .	210
[外部キープロファイリング] ビューへの移動. . . . .	210
外部キープロファイリングビュー. . . . .	210
データオブジェクトリレーションの表示. . . . .	210
ビューのズームインとズームアウト. . . . .	211
データオブジェクトの検索. . . . .	211
カラムリレーションの表示. . . . .	211
イメージとしてのエンティティリレーションダイアグラムの保存. . . . .	212
[外部キープロファイリング] ビューからのデータオブジェクトプロファイル結果の表示. . . . .	212
テーブルビュー. . . . .	213
[テーブル詳細] ペイン. . . . .	213
エンタープライズ検出結果の検証. . . . .	213
カラムのリレーションのキューレーション. . . . .	213
モデルリポジトリへの結果のコミット. . . . .	214
[データドメイン] ビュー. . . . .	215
データドメイン検出結果の表示. . . . .	215
データドメイン検出結果の検証. . . . .	215
行のドリルダウン. . . . .	215
[データドメイン] ビューからのデータオブジェクトプロファイル結果の表示. . . . .	216
[カラムプロファイル] ビュー. . . . .	216
データオブジェクトプロファイル結果の表示. . . . .	216
エンタープライズ検出の実行中のカラムプロファイル結果の表示. . . . .	216
エンタープライズ検出の実行中のデータドメイン検出結果の表示. . . . .	217

エンタープライズ検出のランタイムステータスの表示. . . . .	217
エンタープライズ検出エクスポートファイル. . . . .	218
エンタープライズ検出結果のエクスポート. . . . .	218
<b>第 27 章 : Informatica Developer での Business Glossary Desktop. . . .</b>	<b>219</b>
ビジネス用語集の検索. . . . .	219
ビジネス用語のルックアップ. . . . .	219
ビジネス用語のルックアップ用にホットキーをカスタマイズ. . . . .	220
<b>付録 A : プロファイリングウェアハウス接続に基づく機能のサポート. 221</b>	<b>221</b>
プロファイリング機能のサポート. . . . .	221
<b>索引. . . . .</b>	<b>223</b>

# 序文

*Informatica Data Discovery* ガイドを使用すると、プロファイルを作成および実行して、データソースのコンテンツ、品質、および構造を分析できます。データ検出を実行して、1 つ以上のデータソースでのカラム間の関係など、ソースシステムのメタデータを検出する方法を説明します。Developer tool および Analyst ツールを使用すると、カラムプロファイル、データドメイン検出プロファイル、またはエンタープライズ検出プロファイルを作成、管理および実行できます。

これは、データアナリストおよびデータ開発者向けのガイドです。

## Informatica のリソース

Informatica は、Informatica Network やその他のオンラインポータルを通じてさまざまな製品リソースを提供しています。リソースを使用して Informatica 製品とソリューションを最大限に活用し、その他の Informatica ユーザーや各分野の専門家から知見を得ることができます。

### Informatica Network

Informatica Network は、Informatica ナレッジベースや Informatica グローバルカスタマサポートなど、多くのリソースへの入口です。Informatica Network を利用するには、<https://network.informatica.com> にアクセスしてください。

Informatica Network メンバーは、次のオプションを利用できます。

- ナレッジベースで製品リソースを検索できます。
- 製品の提供情報を表示できます。
- サポートケースを作成して確認できます。
- 最寄りの Informatica ユーザーグループネットワークを検索して、他のユーザーと共同作業を行えます。

### Informatica ナレッジベース

Informatica ナレッジベースを使用して、ハウツー記事、ベストプラクティス、よくある質問に対する回答など、製品リソースを見つけることができます。

ナレッジベースを検索するには、<https://search.informatica.com> にアクセスしてください。ナレッジベースに関する質問、コメント、ご意見の連絡先は、Informatica ナレッジベースチーム ([KB\\_Feedback@informatica.com](mailto:KB_Feedback@informatica.com)) です。

## Informatica マニュアル

Informatica マニュアルポータルでは、最新および最近の製品リリースに関するドキュメントの膨大なライブラリを参照できます。マニュアルポータルを利用するには、<https://docs.informatica.com> にアクセスしてください。

製品マニュアルに関する質問、コメント、ご意見については、Informatica マニュアルチーム ([infa\\_documentation@informatica.com](mailto:infa_documentation@informatica.com)) までご連絡ください。

## Informatica 製品可用性マトリックス

製品可用性マトリックス (PAM) には、製品リリースでサポートされるオペレーティングシステム、データベースなどのデータソースおよびターゲットが示されています。Informatica PAM は、<https://network.informatica.com/community/informatica-network/product-availability-matrices> で参照できます。

## Informatica Velocity

Informatica Velocity は、Informatica プロフェッショナルサービスが開発したヒントとベストプラクティスのコレクションで、多数のデータ管理プロジェクトから得た実体験に基づいています。Informatica Velocity には、世界中の組織と連携してデータ管理ソリューションを計画、開発、デプロイ、管理する Informatica コンサルタントによる集合知を表しています。

Informatica Velocity リソースには、<http://velocity.informatica.com> からアクセスしてください。Informatica Velocity についての質問、コメント、またはアイデアがある場合は、[ips@informatica.com](mailto:ips@informatica.com) から Informatica プロフェッショナルサービスにお問い合わせください。

## Informatica Marketplace

Informatica Marketplace は、お使いの Informatica 製品を拡張したり強化したりするソリューションを検索できるフォーラムです。Marketplace で、Informatica デベロッパーやパートナーからの多数のソリューションを活用すれば、生産性を向上したり、プロジェクトでの実装時間を短縮したりできます。Informatica Marketplace は、<https://marketplace.informatica.com> からアクセスしてください。

## Informatica グローバルカスタマサポート

電話または Informatica Network を介してグローバルカスタマサポートに連絡できます。

各地域の Informatica グローバルカスタマサポートの電話番号は、Informatica Web サイト (<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>) を参照してください。

Informatica Network のオンラインサポートリソースを見つけるには、<https://network.informatica.com> にアクセスして eSupport オプションを選択します。

# パート I: データ検出の概要

この部には、以下の章があります。

- [プロファイリングの概要, 16 ページ](#)
- [データ検出, 21 ページ](#)
- [カラムプロファイルの概念, 25 ページ](#)
- [データドメイン検出の概念, 28 ページ](#)
- [キュレーションの概念, 32 ページ](#)

# 第 1 章

## プロファイリングの概要

この章では、以下の項目について説明します。

- [プロファイリングの概要, 16 ページ](#)
- [プロファイリングアーキテクチャ, 17 ページ](#)
- [データ検出プロセス, 19 ページ](#)

## プロファイリングの概要

プロファイリングは、アプリケーション、スキーマ、またはエンタープライズのデータソースの内容、品質、および構造を検出するために使用します。データソースの内容には、値の頻度やデータ型が含まれます。データソースの構造には、キーや機能依存性が含まれます。

検出プロセスの一環として、プロファイルを作成して実行できます。プロファイルは、エンタープライズのデータソース間のデータの不整合や、データオブジェクトを危険にさらす隠れたデータの問題を検出および分析するリポジトリオブジェクトです。企業内のデータソースに対してプロファイルを実行すると、エンタープライズのデータおよびメタデータの強みと弱みについて把握することができます。

Informatica Analyst と Informatica Developer を使用して、ソースデータやメタデータを分析できます。アナリストと開発者は、これらのツールを使用して、連携しながらデータ品質の問題を特定し、データのリレーションを分析することができます。職務に基づいて、Analyst ツールまたは Developer tool のいずれかの機能を使用できます。実行できるプロファイリングのレベルは、使用するツールによって異なります。

Developer tool と Analyst ツールのどちらでも実行できる作業は次のとおりです。

- カラムプロファイリングを実行する。これには、カラム内の一意の値と NULL 値の数やデータパターンの検出が含まれます。
- データドメイン検出を実行します。エンタープライズ内の重要なデータ特性を検出できます。
- データ型、データドメイン、プライマリキー、外部キーを含むプロファイル結果をキュレーションします。
- スコアカードを作成してデータ品質を監視します。
- オペレーティングシステムプロファイルを選択して、カラムプロファイル、エンタープライズ検出プロファイル、およびスコアカードを作成および実行します。この処理は、オペレーティングシステムプロファイルに定義した、オペレーティングシステムユーザーの権限に基づいて行われます。
- リポジトリアセットのロックを使用して、他のユーザーが作業を上書きしないようにします。
- バージョン管理システムを使用して、複数のバージョンのプロファイルを保存します。
- タグを作成してデータオブジェクトに割り当てる。
- Business Glossary Desktop のビジネス用語としてオブジェクト名の意味を調べる。例えば、カラムやプロファイルの名前の意味を参照してビジネス要件や現在の実装を調べることができます。



Developer tool で実行できる作業は次のとおりです。

- データソース内の 2 つのデータカラム間に見込まれる結合度を検出する。
- 1 つまたは複数のデータソース内のカラムのペア間で発生している重複データの割合を特定する。
- カラムプロファイリングの結果を比較する。
- プロファイルからマッピングオブジェクトを生成する。
- データソース内のプライマリキーを検出する。
- 1 つ以上のデータソースのセット内の外部キーを検出する。
- データソース内のカラム間の機能依存性を検出する。
- 複数の接続にまたがる大量のデータソース上でデータ検出作業を実行する。データ検出作業には、カラムプロファイル、プライマリキーのリレーションと外部キーのリレーションの推測、データドメイン検出、およびデータリレーションをグラフィックにまとめたサマリの生成が含まれます。

Analyst ツールで実行できる作業は次のとおりです。

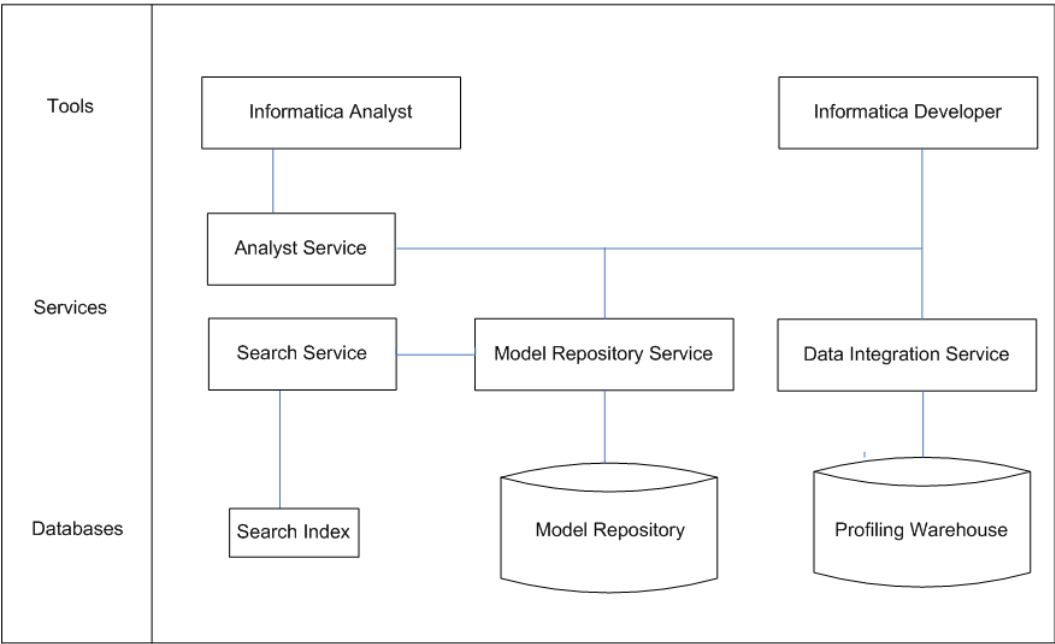
- 複数の接続にまたがる大量のデータソース上でエンタープライズ検出を実行する。カラムメタデータとデータドメインの検出結果をまとめたサマリを表示できます。
- 企業内のデータとメタデータの場所を検索するために、検出検索を実行する。データオブジェクト、ルール、プロファイルなどの特定のアセットを検索できます。検出検索はアセットを検索し、エンタープライズ内のデータベースおよびスキーマにある他のアセットとのリレーションを識別します。
- 履歴プロファイル実行のプロファイル結果を表示します。
- 1 つのカラムプロファイルの 2 回のプロファイル実行の結果を比較する。
- スコアカードのメトリックまたはメトリックグループごとにスコアカードリネージュを表示する。
- スコアカードダッシュボードを表示する。
- プロファイルまたはプロファイル内のカラムにコメントを追加します。
- プロファイルまたはプロファイル内のカラムにタグを割り当てます。
- プロファイル内のカラムにビジネス用語を割り当てます。

## プロファイリングアーキテクチャ

プロファイリングアーキテクチャは、ツール、サービス、およびデータベースで構成されます。ツールコンポーネントは、クライアントアプリケーションで構成されます。サービスコンポーネントには、ツールの管理、データ統合作業の実行、およびプロファイルオブジェクトのメタデータの管理に必要なアプリケーションサー

ビスが含まれます。データベースコンポーネントは、モデルリポジトリとプロファイリングウェアハウスで構成されます。

次の図は、プロファイリングのアーキテクチャコンポーネントを示しています。



プロファイルを実行すると、モデルリポジトリサービスからアナリストサービスまたは Developer tool にプロファイル定義が渡されます。次に、アナリストサービスまたは Developer tool からデータ統合サービスのプロファイリングプラグインが呼び出され、このプロファイリングプラグインによって、プロファイルジョブが処理されてデータ統合サービスにサブミットされます。その後、データ統合サービスでプロファイリング結果が生成され、データ統合サービスにより、プロファイリング結果がプロファイリングウェアハウスに書き込まれます。

検出検索では、検索サービスが使用されます。検索サービスは、モデルリポジトリまたはプロファイリングウェアハウスの代わりに、検索インデックスで各検索を実行します。検索サービスは、モデルリポジトリおよびプロファイリングウェアハウスのコンテンツに基づいて検索インデックスを生成します。検索サービスには、各リポジトリからコンテンツを抽出するためのエクストラクタが含まれています。

以下の表では、アーキテクチャコンポーネントについて説明します。

コンポーネント	説明
Informatica Analyst	データソースのデータやメタデータの検出、分析、およびレポートに使用できる Web ベースのクライアントアプリケーション。
Informatica Developer	プライマリキーの検出、外部キーの検出、およびエンタープライズ検出など、高度なデータ検出を実行する場合に使用するクライアントアプリケーション。
アナリストサービス	Analyst ツールを実行し、サービスコンポーネントと Analyst ツールユーザー間の接続を管理するアプリケーションサービス。
検索サービス	Analyst ツールで検索を管理するアプリケーションサービス。デフォルトでは、検索サービスはデータオブジェクト、プロファイル、マッピング仕様、参照テーブル、ルール、スコアカードなどのモデルリポジトリから検索結果を返します。

コンポーネント	説明
検索インデックス	検索サービスがモデルリポジトリおよびプロファイリングウェアハウスから抽出するインデックス付きコンテンツを格納する、カスタムディレクトリ内のファイルシステム。
モデルリポジトリサービス	モデルリポジトリを管理するアプリケーションサービス。
データ統合サービス	Analyst ツール、Developer tool、および外部クライアントに対するデータ統合作業を実行するアプリケーションサービス。
モデルリポジトリ	Analyst ツールや Developer tool で作成されるプロジェクトのメタデータを格納するリレーショナルデータベース。
プロファイリングウェアハウス	プロファイル結果やスコアカードの結果など、プロファイリング情報を格納するデータベース。

## データ検出プロセス

データ統合プロジェクトでは、多くの場合、最初にプロファイリングを行います。プロファイルを作成することで、データソースの内容、品質、および構造を分析できるようになります。プロファイリングプロセスの環境として、データソースのメタデータを検出します。

使用するプロファイルは、データ分析のタイプ（カラムプロファイル、プライマリキー検出、外部キー検出、データドメイン検出など）によって異なります。データ品質の問題を検出し、文書化します。データ検出を実行するには、以下のタスクを実行します。

1. データソースからデータの内容を検出して分析します。これには、データ型、値の頻度、パターンの頻度、データの統計（最小値や最大値など）が含まれます。
2. データの構造を検出します。これには、キー、機能依存性、外部キーが含まれます。
3. プロファイル結果を確認、検証します。
4. プロファイル結果をドリルダウンします。
5. プロファイル結果をキュレーションします。
6. 参照データを作成します。
7. データの問題を文書化します。
8. ルールを作成して実行します。
9. スコアカードを作成してデータ品質を監視します。

検出プロセスの管理には、次のツールを使用できます。

### Informatica Administrator

ユーザー、グループ、特権、およびロールを管理します。アナリストサービスを管理したり、Informatica Analyst のプロジェクトやオブジェクトに対する権限を管理したりできます。Informatica Developer のアクセス権限もこのツールで制御できます。

### Informatica Developer

1 つ以上のデータソースのメタデータを検出して分析するには（カラムの関係の検出など）、このツールでプロファイルを作成して実行します。プロファイルはウィザードを使用して作成します。

## Informatica Analyst

Analyst ツールでは、カラムプロファイルの実行、データドメイン検出の実行、およびデータオブジェクトへのエンタープライズ検出の実行を行うことができます。プロファイルの実行後、データソースのデータ行をドリルダウンすることができます。

## 第 2 章

# データ検出

この章では、以下の項目について説明します。

- [データ検出の概要, 21 ページ](#)
- [プロファイルおよび分析タイプ, 21 ページ](#)
- [プロファイリングのコンポーネント, 22 ページ](#)
- [プロファイル結果, 23 ページ](#)

## データ検出の概要

データ検出は、ソースシステムのメタデータ（内容と構造）を検出するプロセスです。コンテンツは、データ値、頻度、データ型を参照します。構造には、候補キー、プライマリキー、外部キー、および機能依存性が含まれます。プロファイルを作成して実行することで、データソースの内容と構造を検出することができます。

単一のデータオブジェクト内で、または複数のデータオブジェクト全体でデータを分析するようにプロファイルを定義できます。プロファイルにコメントを追加すると、プロファイリングプロセスを効果的に追跡することができます。

データ構造を評価し、データカラムに必要なタイプの情報が含まれていることを確認するには、プロファイルを実行します。プロファイリングしたデータのデータ行をドリルダウンできます。プロファイル結果でデータに問題が見つかった場合は、ルールを適用して結果セットを修正できます。ルールの適用前と適用後にスコアカードを作成して、データ品質を追跡および測定することができます。プロファイルまたはスコアカードの外部ソースメタデータに変更が発生する場合は、変更をそのデータオブジェクトに同期させることができます。

## プロファイルおよび分析タイプ

プロファイルは、実行する必要がある分析のタイプに基づいて作成します。作成するプロファイルのタイプは、実行する分析のタイプに対応しています。例えば、プライマリキー分析を実行する場合は、プライマリキープロファイルを作成します。

以下のプロファイルを作成して、データの分析および検出を実行できます。

### カラムプロファイル

テーブルまたはファイル内の選択したカラムのデータ品質を分析します。カラム分析用のプロファイルは、Analyst ツールおよび Developer ツールで定義できます。

### データドメイン検出

エンタープライズ内の重要なデータ特性を検出します。データドメイン検出では、カラム値またはカラム名に基づいて、カラムに関連付けられているすべてのデータドメインが特定されます。検出プロセスの一環として、データーールやカラム名ルールを手動で作成して、値やカラム名がデータドメインに属するかどうかを確認できます。データドメインを作成する際に、これらのルールを関連付けることができます。また、データドメインは、カラムプロファイル結果の値およびパターンから作成することもできます。

### プライマリキープロファイル

テーブルまたはファイル内のカラム間のプライマリキー関係を検出します。プライマリキー分析用のプロファイルは、Developer ツールで定義できます。

### 機能依存性プロファイル

テーブルまたはファイル内のカラム間の機能依存性を検出します。機能依存性分析用のプロファイルは、Developer ツールで定義できます。

### 外部キープロファイル

複数のテーブルまたはファイル全体のカラム間の外部キー関係を検出します。外部キー分析用のプロファイルは、Developer ツールで定義できます。

### 結合プロファイル

単一のデータソース内、または複数のデータソース全体のカラム間に見込まれる結合度を特定します。結合分析用のプロファイルは、Developer ツールで定義できます。結果はベン図に表示されます。

### 重複検出

1 つまたは複数のデータソース内のカラムのペア間で発生している重複データの割合を特定します。重複検出タスクは、Developer ツール内のエディタから実行できます。結果の検証と表示は、ベン図で行うことができます。

### エンタープライズ検出

大量のデータソースに対して複数のデータ検出タスクを実行し、プロファイル結果の統合サマリを生成します。カラムプロファイルの実行、データドメイン検出の実行、およびプライマリキー関係と外部キー関係の検出が含まれます。エンタープライズ検出は、大量のデータソースのプロファイルプロセスを自動化します。

**注:** Analyst ツールでプロファイルに加えた変更は、Developer tool のモデルリポジトリへの接続を更新するまで Developer tool には反映されません。

## プロファイリングのコンポーネント

プロファイルには複数のコンポーネントがあり、それらを使用することでデータソースの内容と構造を効果的に分析することができます。

プロファイルのコンポーネントを次に示します。

### フィルタ

元のデータソースから特定の条件を満たすサブセットを作成します。その後、このサンプルデータに対してプロファイルを実行できます。

### ルール

プロファイルの実行時にデータに適用される条件を定義するビジネスロジック。データを検証するには、プロファイルにルールを追加します。

## タグ

ビジネス上の用途に基づいてモデルリポジトリ内のオブジェクトを定義するメタデータ。タグを作成すると、ビジネス用途に従ってオブジェクトをグループ化できます。タグは、Analyst ツールで、プロファイルまたはプロファイル内のカラムに割り当てます。

## コメント

プロファイルに関する説明。コメントは、プロファイルに関する情報を Analyst ツールおよび Developer tool の他のユーザーと共有する場合に使用します。コメントは、Analyst ツールで、プロファイルまたはプロファイル内のカラムに追加します。

## スコアカード

カラムの有効値またはプロファイル結果内のルール of の出力を視覚的に表したものの。データ品質の進捗を測定するには、スコアカードを使用します。

# プロファイル結果

プロファイル結果はプロファイルの実行後に表示できます。プロファイルのカラムとルールについて、値の概要、パターン、および統計を確認できます。プロファイルのカラムとルールのプロパティが表示され、プロファイルデータをプレビューできます。

以下の表に、各プロファイルタイプに対するプロファイル結果を示します。

プロファイルタイプ	結果
カラムプロファイル	<ul style="list-style-type: none"><li>- カラム内の NULL 値、重複なしの値、および重複ありの値の数と割合、カラム値の推測されたデータ型。</li><li>- 選択したカラム内のデータ値の頻度と文字パターン、そのカラムの統計の概要。</li><li>- カラムデータを分析して推測するデータタイプ。</li><li>- 文書化された、データのデータタイプ。</li><li>- 最大値と最小値。</li><li>- プロファイルの実行日時。</li></ul>
プライマリキープロファイル	<ul style="list-style-type: none"><li>- 推測されるプライマリキー候補の一意の値、重複する値、および NULL 値の数と割合。</li><li>- 推測されるプライマリキー候補内のキー違反の数。</li></ul>
機能依存性プロファイル	<ul style="list-style-type: none"><li>- 推測される機能依存性。</li><li>- 機能依存性違反の数。</li></ul>
外部キープロファイル	<ul style="list-style-type: none"><li>- 定義したプライマリキーと外部キーの推測条件を満たすプライマリキーカラムと外部キーカラム。</li><li>- プライマリキーと外部キーの間で一致するデータ値の数（割合で表される）。</li><li>- プロファイルの実行前にプライマリキーカラムと外部キーカラムに対して定義したリレーションのタイプ。</li></ul>
結合プロファイル	<ul style="list-style-type: none"><li>- カラム間のリレーションを示すベン図。</li><li>- カラム内の孤立値、NULL 値、および結合値の数と割合。</li></ul>
重複検出	<ul style="list-style-type: none"><li>- 2 つのカラム間の重複の割合。</li><li>- カラム間の重複を示すベン図。</li></ul>

プロファイルタイプ	結果
データドメイン検出	<ul style="list-style-type: none"> <li>- 事前定義されたデータドメインと一致するカラム名およびデータ。</li> <li>- カラムが所属するデータドメイングループとそのデータタイプ。</li> </ul>
エンタープライズ検出	<ul style="list-style-type: none"> <li>- カラムプロファイル結果。</li> <li>- データドメイン検出結果。</li> <li>- プライマリキー検出結果。</li> <li>- 外部キープロファイル結果のグラフィカルビューとテーブルビュー。</li> </ul>

サードパーティのレポートツールを使用して、プロファイリングウェアハウスからプロファイル結果を読み取ることができます。Informatica が用意したプロファイルビューのセットで、読み取るプロファイル統計に合わせてカスタマイズできます。これらのビューは、一般的なタイプのプロファイル統計とプロファイル結果分析に基づいています。



## 第 3 章

# カラムプロファイルの概念

この章では、以下の項目について説明します。

- [カラムプロファイルの概念の概要, 25 ページ](#)
- [カラムプロファイルオプション, 26 ページ](#)
- [リポジトリプロファイルのロックおよびバージョン管理されたプロファイルの管理, 26 ページ](#)
- [スコアカード, 27 ページ](#)

## カラムプロファイルの概念の概要

カラムプロファイルによって、値の頻度、割合、パターンなど、データソースに含まれるカラムの特性が決まります。

カラムプロファイリングでは、データに関する以下の情報が検出されます。

- カラムごとの NULL 値、重複なしの値、重複ありの値の数（数値および割合で表現される）。
- カラムごとのデータのパターン、およびそれらの値が出現する頻度。
- カラム値に関する統計。値の最大長さと最小長さや、各カラムの最初の値と最後の値など。
- 文書化されたデータ型、推測されたデータ型、および文書化されたデータ型と推測されたデータ型の間で起こりうる競合。
- パターンと値頻度の異常値。

プロファイルの作成または編集時に、次のオプションを設定できます。

- カラムプロファイルオプション。プロファイルの実行、サンプリングオプションの選択、およびドリルダウンオプションの選択を行うカラムを選択できます。
- フィルタおよびルールの追加、編集、または削除。

プロファイル結果で、プロファイルおよびプロファイルのカラムにコメントおよびタグを追加できます。ビジネス用語をカラムに割り当てられます。

モデルリポジトリでは、ユーザーがリポジトリプロファイルロックで作業を上書きしないように、プロファイルがロックされます。バージョン管理システムでは、複数バージョンのプロファイルが保存され、各バージョンにバージョン番号が割り当てられます。プロファイルをチェックアウトして、変更後にチェックインできます。プロファイルを再度チェックインする前に、プロファイルのチェックアウトアクションを取り消すことができます。

データ品質を定期的に確認するには、スコアカードを作成します。カラムの有効値を視覚的に表示できるように、プロファイルへのルールの適用前と適用後にスコアカードを作成します。

スケジューラサービスを使用して、プロファイルの実行やスコアカードの実行が特定の時刻または間隔で実行されるように、スケジュールを設定します。スケジューラサービスでは、プロファイル、スコアカード、デプロイ済みのマッピング、およびデプロイ済みのワークフローのスケジュールを管理します。スケジュールの作成、管理、および実行は、Informatica Administrator で行えます。

## カラムプロファイルオプション

プロファイルを作成するときに、プロファイルウィザードを使用してフィルタ、ルール、ドリルダウンオプション、サンプリングオプション、および接続を定義できます。これらのオプションにより、プロファイルがソースデータから行を読み取る方法が決定されます。

カラムプロファイル、データドメイン検出プロファイル、またはエンタープライズ検出プロファイルの次のオプションを定義できます。

- フィルタ。フィルタを作成してプロファイルに適用できます。
- ルール。プロファイルの作成時にルールを追加できます。Analyst ツールまたは Developer tool で作成したルールは、利用できます。
- ドリルダウンオプション。データソース内の現在のデータを読み取るか、プロファイリングウェアハウスにステージングされたプロファイルデータを読み取るかを選択できます。
- サンプリングオプション。サンプリングオプションのいずれかを選択して、プロファイルを実行する行の数を決定できます。
- 接続。プロファイルは、ネイティブまたは Hadoop ランタイム環境で実行できます。Hadoop ランタイム環境では、Blaze または Spark エンジンを選択できます。

## リポジトリプロファイルのロックおよびバージョン管理されたプロファイルの管理

モデルリポジトリでは、ユーザーが作業を上書きしないように、プロファイルがロックされます。プロファイルの編集を開始すると、他のユーザーがプロファイルに対する変更を保存しないように、プロファイルがロックされます。プロファイルを保存すると、ロックは解除されます。バージョン管理されたプロファイルの管理ではプロファイルのバージョンが作成され、バージョン履歴を表示できます。

Developer tool または Analyst ツールでプロファイルを編集するときに、モデルリポジトリはプロファイルをロックします。ツールが予期せずに停止した場合、ロックは保持されて、モデルリポジトリに再度接続したときにロックしたプロファイルを表示できます。プロファイルの編集を継続するか、プロファイルのロックを解除することができます。

モデルリポジトリがバージョン管理システムと統合されている場合は、プロファイルのバージョンを管理できます。例えば、プロファイルのチェックアウトとチェックイン、チェックアウトの取り消し、プロファイルの特定の履歴バージョンの表示、チェックアウトしたプロファイルの表示などが可能です。Analyst ツールでのリポジトリアセットのロックおよびバージョン管理されたアセットの管理については、『Analyst ツールガイド』を参照してください。Developer tool でのリポジトリオブジェクトのロックおよびバージョン管理されたオブジェクトの管理については、『Developer Tool ガイド』を参照してください。

# スコアカード

スコアカードは、カラムの有効値またはプロファイル結果内のルールの出力を視覚的に表したものです。データ品質の進捗を測定するには、スコアカードを使用します。プロファイルからスコアカードを作成し、データ品質の進捗を時系列で監視することができます。

スコアカードには、メトリック、メトリックグループ、およびしきい値など、複数のコンポーネントがあります。プロファイルの実行後に、スコアカードにメトリックなどのソースカラムを追加し、メトリックの有効値を設定できます。スコアカードは、メトリックおよびスコアカードレベルで不良データのコストを追跡してデータ品質の値を測定するのに役立ちます。メトリックごとに不良データのコストを測定するには、メトリックにコスト単位を割り当て、固定費または変動費を設定します。スコアカードを実行すると、スコアカードの結果には、各メトリックの不良データのコストと、すべてのメトリックのコスト合計値が含まれます。

スコアカード内の関連し合うスコアを1つのまとまりとして分類するには、メトリックグループを使用します。しきい値では、レコード内のカラムで許容できる不良データの範囲を割合で指定します。正常、許容可能、または許容不可の範囲のデータに対してしきい値を設定できます。

スコアカードを実行する際、実データまたはステージングされたデータで、ドリルダウンするスコアメトリックを設定します。スコアカードを実行してスコアを表示した後、各メトリックをドリルダウンして、有効なデータレコードと無効なレコードを識別します。スコアカードリネージュは、スコアカード内のメトリックまたはメトリックグループごとにも表示できます。データ品質を効率的に追跡するには、スコアの傾向グラフとコストの傾向グラフを使用します。これらのグラフで、不良データのスコアとコストが、特定の期間にわたって変化する様子を監視します。

スコアカードのメトリックと設定情報はプロファイリングウェアハウスに保存されます。スコアカードの結果を取得してレポートを実行するようにサードパーティのアプリケーションを設定できます。スコアカードの結果は、Web アプリケーション、ポータル、またはビジネスインテリジェンスレポートなどのレポートにも表示できます。

## 第 4 章

# データドメイン検出の概念

この章では、以下の項目について説明します。

- [データドメイン検出の概念の概要, 28 ページ](#)
- [データドメイン, 29 ページ](#)
- [データドメイングループ, 29 ページ](#)
- [データドメイングループサリ, 30 ページ](#)
- [データドメイン検出のプロセス, 30 ページ](#)
- [Spark エンジンでのデータドメイン検出, 30 ページ](#)

## データドメイン検出の概念の概要

重要なソースデータに対して効果的に作業するための対策を取ることができるように、そのソースデータの意味を特定して理解する必要があります。データドメイン検出は、データのセマンティクスに基づいてデータソース内でのデータの機能的な意味を検出するプロセスです。

プロファイルを作成してデータドメイン検出を実行すると、エンタープライズ内の重要なデータ特性を特定できます。そのうえで、データ品質やデータマスキングなどのデータ管理ポリシーをさらにデータに適用できます。例えば、製品コードや説明を検出し、データを有用で信頼できるものにするためには、どのデータ品質の標準化ルールや解析ルールを適用する必要があるかを分析します。もう 1 つの例として、クレジットカード番号、電子メールアドレス、電話番号などの機密性の高い顧客データの検出があります。その後、検出した情報をマスクして保護することができます。

Analyst ツールと Developer tool の両方で、プロファイルを作成して実行し、データドメイン検出を実行することができます。データドメイン検出を実行するプロファイルは、次のルールに基づいて定義できます。

- データルール。ルールに定義された特定のロジックに一致するデータを含むカラムを検出します。
- カラム名ルール。ルールに定義されたカラム名ロジックに一致するカラムを検出します。

カラムプロファイル結果の値およびパターンから、データドメインを作成できます。次に、このデータドメインを使用して、複数のデータシステムまたはエンタープライズにわたって重要なデータを検出できます。

サンプリングオプションやフィルタを設定したプロファイルを作成して、データドメイン検出を実行できます。プロファイルを実行すると、データソースにサンプリングオプションとフィルタが適用されて、データセットが生成されます。データドメイン検出プロセスは、このデータセットを使用してデータドメインを検出します。

# データドメイン

データドメインは、カラムデータまたはカラム名のセマンティクスに基づいた、定義済みまたはユーザー定義のモデルリポジトリオブジェクトです。例えば、社会保障番号、クレジットカード番号、電子メール ID、および電話番号をそれぞれデータドメインにすることができます。

データドメインは、データソース内で未検出のままになっている重要なデータの検出に役立ちます。例えば、レガシーデータシステムでは社会保障番号が「コメント」フィールドに含まれているとします。新しいデータシステムに移す前に、この情報を検出して保護する必要があります。

データドメインの一致の適合条件として、ソース行の最小割合またはソース行の最小数を選択できます。また、カラムプロファイルでデータドメイン検出を実行する際に、NULL 値を除外することもできます。

論理データドメインはデータドメイングループにグループ化できます。データドメイングロッサリに、すべてのデータドメインとデータドメイングループが一覧表示されます。Developer tool の「設定」メニューを使用して、データをデータドメイングロッサリからインポート、またはデータドメイングロッサリにエクスポートします。

ソースデータとメタデータに一致するデータとカラム名のパターンを定義するには、ルールを使用します。データドメインを作成すると、Analyst ツールまたは Developer tool により、関連付けられているルールや他の依存オブジェクトがデータドメイングロッサリにコピーされます。Developer tool を使用して、データドメインのデータドメイングロッサリからのインポートやエクスポートなど、データドメインを管理します。Developer tool を使用して、データドメインのルールロジックを管理することもできます。

**注:** すべてのデータドメインルールを 1 つのプロジェクトまたはフォルダーに保存することができます。これは、データドメインをエクスポートした後で、ルールや関連付けられている他のデータオブジェクトを編集する必要がある場合に役立ちます。

## データドメイングループ

データドメイングループは、データドメインを特定のグループに分類する際に役立ちます。例えば、データドメインの first\_name、last\_name、および account\_number を Personal Health Information (PHI) データドメイングループにグループ化できます。

社会保障番号、名、および姓を含む、Personally Identifiable Information (PII) データドメイングループを作成できます。1 つのデータドメインを、複数のデータドメイングループに含めることができます。例えば、社会保障番号を Payment Card Industry (PCI) データドメイングループと PII データドメイングループの両方に追加できます。データドメイングループには、データドメインを含めることができますが、他のデータドメイングループを含めることはできません。

**注:** インストール後、データドメインファイル Informatica\_IDE\_DataDomain.xml をインポートすると、データドメイングロッサリに定義済みのデータドメイングループとデータドメインが表示されます。その後、必要に応じてデータドメイングループを追加で作成できます。データドメインに関連付けられているルールを表示したり変更したりするには、Informatica\_IDE\_DataDomainRule.xml ファイルをインポートします。

# データドメイングロッサリ

データドメイングロッサリは、すべてのドメイングループとデータドメインのコンテナです。データドメイングロッサリを使用して、データドメインとデータドメイングループを作成、管理、および削除することができます。

特定のドメインとドメイングループをデータドメイングロッサリ内で検索することができます。データドメインを XML ファイルにエクスポートしたり、XML ファイルからデータドメイングロッサリにインポートしたりすることもできます。データドメイングロッサリには、コピーされたルールとデータドメインに関連付けられているすべての参照データが含まれます。データドメイングロッサリ内のルールは編集できません。

データドメインは、Developer ツールの「設定」メニュー、および Analyst ツールの「管理」メニューから表示できます。データドメインとデータドメイングループを作成、編集、および削除するユーザーを指定するには、Model Repository Service の特権 **Manage Data Domains** を使用します。

## データドメイン検出のプロセス

職務に応じ、Analyst ツールまたは Developer ツールでプロファイルを定義して実行し、データドメインの検出を実行できます。データドメイン検出オプションを設定してプロファイルを実行したら、検出結果を検証してドリルダウンすることができます。エディタ内からデータドメイン検出を実行した場合は、検出結果をデータモデルに追加できます。

データドメイン検出を実行するには、以下の手順を実行します。

1. データドメインとドメイングループを作成またはインポートします。
2. 必要に応じて、適切なドメイングループ内にデータドメインを統合します。
3. プロファイルを作成してデータドメイン検出を実行します。最初に、データドメイン検出とともにカラムプロファイルを実行するか、データドメイン検出のみを実行するかを選択します。
4. カラム、データドメイン、および適切なサンプリングオプションを選択します。
5. プロファイルを実行します。
6. プロファイル結果を検証、ドリルダウンし、必要に応じて結果をデータモデルに追加します。

## Spark エンジンでのデータドメイン検出

Spark エンジンでデータ検出を実施するためにプロファイルを実行する際に、参照テーブルは Hadoop クラスタにステージングされます。すべてのデータドメインの参照テーブルがクラスタにステージングされたことを確認するには、次の手順を実行できます。

### 前提条件:

データドメイン検出の実行時に、HDFS ユーザーを偽装する権限がある必要があります。

### JDBC .JAR ファイルのダウンロード

1. 使用する参照データベースの JDBC .jar ファイルを取得します。ファイルはデータベースベンダの Web サイトからダウンロードできます。
2. ダウンロードしたファイルを <INFA\_HOME>/externaljdbcjars にコピーします。

## カスタムプロパティのデータ統合サービスでの設定

1. Informatica Administrator を起動し、**データ統合サービス**を**ドメインナビゲータ**で選択します。
2. **【プロパティ】** タブで **【カスタムプロパティ】** オプションをクリックします。
3. データドメインの参照テーブルをステージングするには、次のカスタムプロパティを設定します。

プロパティ名	プロパティ値
AdvancedProfilingServiceOptions.ProfilingSparkReferenceDataHDFSDir	/tmp/cms
ExecutionContextOptions.SparkRefTableHadoopConnectorArgs	--connect <JDBC シンドライバ接続 URL>

4. /tmp/cms ディレクトリがクラスタ上にあることを確認します。このディレクトリがない場合、/tmp/cms ディレクトリまたはデータのステージング先にするカスタムディレクトリを作成します。参照データはデフォルトで/tmp/cms ディレクトリにステージングされます。
5. データ統合サービスをリサイクルします。
6. Analyst ツールまたは Developer tool を開いて、参照データをステージングするすべてのデータドメインで最初のプロファイルを実行に実行します。

**注:** 最初のプロファイル実行ですべてのデータドメインを選択せず、追加のデータドメインを次のプロファイル実行で選択すると、プロファイル実行に失敗する可能性があります。

## 第 5 章

# キュレーションの概念

この章では、以下の項目について説明します。

- [キュレーションの概念の概要, 32 ページ](#)
- [解析者と開発者のキュレーション, 32 ページ](#)
- [キュレーションタスク, 33 ページ](#)

## キュレーションの概念の概要

キュレーションとは、データソースで検出されたメタデータを検証し管理することで、使用とレポートに適したメタデータにするプロセスです。

キュレーションできる推測されたプロファイル結果を、次に示します。

- データ型
- データドメイン
- プライマリキー
- 外部キー

推測されたプロファイル結果をキュレーションすることで、カラム、データドメイン、およびデータベースとスキーマでのデータオブジェクト関係に関するメタデータを正確にします。そして、検出検索を使用して複数のリポジトリ全体で情報を検索した場合に、最も関連性の高いメタデータを探することができます。また、エンタープライズ検出の結果に外部キーの関係図を表示した場合にも、最も関連性の高いメタデータを探することができます。

プロファイル実行中にプロファイルによって生成された特定のメタデータの推測をキュレーションできます。例えば、カラムプロファイルの結果とデータドメイン検出の結果の中の推測されたデータ型を承認または拒否できます。また、エンタープライズ検出の結果の中の推測されたプライマリキーや外部キーも承認または却下できます。

## 解析者と開発者のキュレーション

データ解析の担当者、すなわちデータスチュアートであれば、Analyst ツールを使用してカラムのプロファイルの結果とデータドメイン検出の結果をキュレーションすることができます。プロファイル結果をキュレーシ



ョンすることで、検出検索とデータアセットの詳細検証に対応できる正確なプロファイル情報を準備することができます。

開発者やデータ構築担当の方なら、Developer ツールでカラムプロファイルの結果、データドメイン検出の結果、プライマリキー検出の結果、外部キー検出の結果をキュレーションできます。

### キュレーションの例

開発者としてエンタープライズ検出を実行する場合、Developer ツールはデータセット全体に対して選択したデータドメインを処理します。この実行によって、電話番号データが社会保障番号データドメインとして推測されるなど、複数のデータドメインに推測されることがあります。複数のデータドメインに推測される現象は、カラム内のデータの一部が異なるデータドメインと一致する場合に発生します。例えば、10 桁の電話番号の 1 桁が欠けると、社会保障番号と同じパターンになる可能性があります。この現象は、カラム内、または複数のデータドメイン全体の一致パターンの中に、潜在的なデータ品質の問題があることを示唆しています。この場合、Developer ツールは電話番号データドメインと社会保障番号データドメインの両方に推測する可能性があります。最適なデータドメインを選択して承認できるように、プロファイル結果をキュレーションすることができます。この例では、データドメインが社会保障番号であるという推測がデータ品質問題によって発生するため、電話番号が関連するデータドメインです。

エンタープライズ検出を実行する場合、Developer tool が日付カラムに対して、Date、String、Varchar など、複数のデータ型を推測する可能性があります。データアーキテクトは、例えば、日付カラムで最も関連性の高いデータ型である Date データ型を選択して承認できます。

Developer tool 内のエンタープライズ検出では、データオブジェクト関係をすべてカラムデータに基づいて推測する可能性があります。このようなデータオブジェクト関係の一部には、検出された候補キーに不必要なデータオブジェクト関係が含まれています。例えば、Developer tool がシーケンスを表すカラムを考えられるキーとして推測し、似たカラムを持つ他のテーブルとのリレーションを検出するという可能性があります。これらのデータオブジェクトの関係は、データベース内で有効な関係を形成しない場合があります。そのような場合、キュレーションの中で、最適な推測されたプロファイル結果を評価、検証、および承認することができます。

## キュレーションタスク

プロファイルの実行後にプロファイル結果をキュレーションすることができます。前回プロファイルを実行したときのキュレーションの決定を撤回することもできます。

Analyst ツールで実行できるキュレーションタスクは次のとおりです。

- 複数のカラムまたはデータドメインについて推測されたデータ型を承認または拒否します。
- [承認済み] または [拒否済み] のデータ型を [推測済み] ステータスに戻します。
- [承認済み] または [却下済み] のデータドメインを [推測済み] ステータスに戻します。
- 却下された結果の行を表示または非表示にします。
- 承認されたデータ型とデータドメインなど、メタデータの特定の設定に基づいてカラムをプロファイルの実行から除外します。

Developer ツールで実行できるキュレーションタスクは次のとおりです。

- 複数のカラムについて、推測されたデータ型を承認または拒否します。
- [承認済み] または [拒否済み] のデータ型を [推測済み] ステータスに戻します。
- [承認済み] または [却下済み] のデータドメインを [推測済み] ステータスに戻します。
- 却下された結果の行を表示または非表示にします。

- プライマリキー検出の結果の中のデータオブジェクトを承認または却下します。
- 外部キー検出の結果など、エンタープライズ検出の結果を承認または却下します。
- 承認されたデータ型とデータドメインなど、メタデータの特定の設定に基づいてカラムをプロファイルの実行から除外します。

# パート II: Informatica Analyst を使用したデータ検出

この部には、以下の章があります。

- [Informatica Analyst のカラムプロファイル, 36 ページ](#)
- [Informatica Analyst のルール, 48 ページ](#)
- [Informatica Analyst のフィルタ, 53 ページ](#)
- [Informatica Analyst のカラムプロファイル結果, 58 ページ](#)
- [Informatica Analyst のビジネス用語、コメント、タグ, 80 ページ](#)
- [Informatica Analyst のスコアカード, 83 ページ](#)
- [Informatica Analyst でのデータドメイン検出, 107 ページ](#)
- [Informatica Analyst でのエンタープライズ検出, 118 ページ](#)
- [Informatica Analyst でのエンタープライズ検出の結果, 124 ページ](#)
- [Informatica Analyst での検出検索, 129 ページ](#)
- [Informatica Analyst での Business Glossary Desktop, 138 ページ](#)

## 第 6 章

# Informatica Analyst のカラムプロファイル

この章では、以下の項目について説明します。

- [Informatica Analyst のカラムプロファイルの概要, 36 ページ](#)
- [カラムプロファイリングプロセス, 37 ページ](#)
- [プロファイルオプション, 37 ページ](#)
- [ランタイム環境, 39 ページ](#)
- [Informatica Analyst のオペレーティングシステムプロファイルの概要, 40 ページ](#)
- [リポジトリアセットのロックとチームベース開発の概要, 41 ページ](#)
- [Informatica Analyst でのカラムプロファイルの作成, 42 ページ](#)
- [カラムプロファイルの編集, 43 ページ](#)
- [プロファイルの実行, 44 ページ](#)
- [Spark エンジンでのプロファイルの実行, 44 ページ](#)
- [同期オプション, 44 ページ](#)

## Informatica Analyst のカラムプロファイルの概要

プロファイルを作成するには、プロファイルを実行するデータオブジェクトのカラムを選択します。サンプリングおよびドリルダウンのオプションを設定することによって、プロファイリングの高速化を図ることができます。ランタイム環境を選択できます。プロファイルを作成するには、ルールおよびフィルタをプロファイルに追加できます。プロファイルの実行後に、プロファイリングの統計を調べてデータを把握できます。

最大 1000 のカラムを含む幅広いテーブルおよびフラットファイルをプロファイリングできます。プロファイルを作成または実行する際には、すべてのカラムを選択するか、プロファイルで使用する各カラムを選択できます。ドリルダウン対象としてすべてのカラムを選択し、これらのカラムの値の頻度を表示することができます。カラム名が 245 文字を超えると、区切りファイルでプロファイルを実行するカラムを選択できません。

Spark エンジンでは、半構造化データソースでプロファイルを実行できません。

Informatica Analyst の次の方法を使用してカラムプロファイルを作成できます。

- **[ライブラリ]** ワークスペースでデータオブジェクトを右クリックして、プロファイルを作成します。
- デフォルトのオプションを使用してデフォルトのカラムプロファイルを作成します。
- プロファイルの設定をカスタマイズしてカスタムプロファイルを作成します。

**注:** Avro、JSON、Parquet、および XML データソースに対して、プロファイルを表示および実行できます。Informatica Developer で、Avro、JSON、Parquet および XML データソースに対してカラムプロファイルを作成および編集できます。

## カラムプロファイリングプロセス

カラムプロファイリングプロセスの一部として、プロファイリングにすべてのソースカラムを含めるか、特定のカラムを選択することができます。また、デフォルトのプロファイルオプションをそのまま使用したり、サンプリングオプション、ドリルダウンオプション、およびランタイム環境を設定したりすることもできます。

以下に、カラムプロファイリングプロセスの手順を示します。

1. カラムプロファイルの名前、説明、場所を選択します。
2. プロファイルを実行するインポートされたデータオブジェクトまたは外部ソースを選択します。
3. 必要に応じて、ソースデータをプレビューします。
4. プロファイルを実行するカラムを選択します。
5. デフォルトオプションを使用してプロファイルを作成するか、デフォルトのオプションを変更するかを指定します。設定できるオプションには、サンプリングオプション、ドリルダウンオプション、およびランタイム環境があります。
6. 必要に応じて、プロファイルの作成時にルールおよびフィルタを追加します。
7. プロファイルを実行します。

**注:** カラム名、および多言語データや Unicode データのプロファイリングについては、以下のルールとガイドラインに注意してください。

- さまざまなソースから多言語データをプロファイリングしたり、ブラウザのロケール設定に基づいてプロファイル結果を表示したりすることができます。Analyst ツールでは、ブラウザのロケールに基づいて、Datetime、Numeric、Decimal の各データ型が変更されます。
- 多言語データに基づくソート: 多言語データに基づいてソートすることができます。Analyst ツールでは、ソート順がブラウザのロケールに基づいて表示されます。
- DB2 データベースの Unicode データをプロファイリングするには、データベースで DB2CODEPAGE データベース環境変数を設定し、データ統合サービスを再起動します。

## プロファイルオプション

プロファイルオプションには、データサンプリングオプションとデータドリルダウンオプションがあります。データオブジェクトのカラムプロファイルを作成または編集するときに、これらのオプションを設定できます。

プロファイルオプションを設定するには、**[検出]** ワークスペースを使用します。カラム、サンプリング、およびドリルダウンのデフォルトのオプションを使用してプロファイルを作成できます。実データとステージングされたデータの選択を切り替えるには、ドリルダウンオプションを使用します。

## サンプリングのオプション

Analyst ツールでプロファイルの対象として選択される行の数は、サンプリングのオプションで決まります。サンプリングのオプションは、プロファイルの定義または実行するときに設定することができます。

以下の表に、プロファイルのサンプリングのオプションを示します。

オプション	説明
すべての行	データオブジェクト内のすべての行でプロファイルを実行します。 ネイティブ、Blaze、Spark、および Databricks ランタイム環境でサポートされます。
最初の<number>行をサンプリング	データオブジェクト内の行のうち、先頭からのサンプル行でプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
<number>行のランダムサンプリング	データオブジェクトでランダムに選択された数の行に対してプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
ランダムサンプリング (自動)	データオブジェクト内の行数に基づいて計算されたサンプル行でプロファイルを実行します。 ネイティブおよび Blaze ランタイム環境でサポートされます。
Limit n <number>行	データオブジェクトの行数に基づいてプロファイルを実行します。Hadoop 検証環境でプロファイルを実行するように選択すると、Spark エンジンはデータオブジェクトの複数のパーティションからサンプルを収集し、これらのサンプルを単一のノードにプッシュしてサンプルサイズを計算します。Limit n のサンプリングオプションは、Oracle、SQL Server および DB2 データベースをサポートします。Limit n のサンプリングオプションには、詳細フィルタは適用できません。 Spark ランタイム環境でサポートされます。
ランダムパーセンテージ	データオブジェクト内の行の割合に対してプロファイルを実行します。 Spark および Databricks ランタイム環境でサポートされます。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

ランダムな行のサンプルでプロファイルを実行するように選択すると、ランダムサンプルアルゴリズムでは、データオブジェクト内の行をランダムに選択して、プロファイルを実行します。カラムプロファイルのランダムサンプリングオプションを選択すると、Analyst ツールは、ステージングされたデータに対してドリルダウンを実行します。これは、ドリルダウンのパフォーマンスに影響を与える可能性があります。データドメイン検索プロファイルのランダムサンプリングオプションを選択すると、Analyst ツールはライブデータのドリルダウンを実行します。

## ドリルダウンのオプション

ドリルダウンのオプションは、プロファイルを定義または編集するときに設定できます。

以下の表に、プロファイルのドリルダウンのオプションを示します。

オプション	説明
Live	実データをドリルダウンして、データソースの現在のデータを読み取ります。
Staged	ステージングされたデータをドリルダウンして、プロファイリングウェアハウスでステージングされているプロファイルデータを読み取ります。
カラム選択	プロファイリング対象として選択していないカラムをドリルダウン対象として指定します。

## ランタイム環境

カラムプロファイルのランタイム環境として、ネイティブ、Hadoop、または Databricks を選択できます。Hadoop ランタイム環境では、Blaze または Spark エンジンを選択できます。Databricks ランタイム環境では、Databricks Spark オプションを選択できます。Informatica Analyst では、ランタイム環境の選択後に、プロファイル定義にランタイム環境が設定されます。

### ネイティブ環境

ネイティブのランタイム環境でプロファイルを実行すると、Analyst ツールがプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、データ統合サービスが実行する同じマシン上でこれらのマッピングを実行し、プロファイル結果をプロファイリングウェアハウスに書き込みます。デフォルトでは、すべてのプロファイルがネイティブのランタイム環境で実行されます。

ネイティブソースを使用し、ネイティブ環境でプロファイルを作成して実行できます。ネイティブのデータソースは、フラットファイル、リレーショナルソース、メインフレームソースなど、Hadoop 以外のソースのことです。また、ネイティブ環境で Hive または HDFS データソースを使用してマッピング仕様または論理データソースに対してプロファイルを実行することもできます。

### Hadoop 環境

Hadoop ランタイム環境でプロファイルを実行する場合、Blaze エンジンまたは Spark エンジンを選択できます。

[Blaze] または [Spark] の選択後に、Hadoop 接続を選択できます。データ統合サービスがプロファイルのロジックを Hadoop クラスタの Blaze または Spark エンジンにプッシュして、プロファイルを実行します。

Hadoop 環境でプロファイルを実行すると、Analyst ツールがプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、これらのマッピングを Hadoop 接続で Hadoop 環境にプッシュします。Blaze エンジンまたは Spark エンジンはマッピングを処理し、データ統合サービスはプロファイル結果をプロファイリングウェアハウスに書き込みます。

## Sqoop データソースのカラムプロファイル

Sqoop を使用するデータオブジェクトでカラムプロファイルを実行できます。Hadoop を検証環境として選択したら、Hadoop 接続で Blaze エンジンまたは Spark エンジンを選択して、カラムプロファイルを実行します。

論理データオブジェクトまたはカスタマイズデータオブジェクトのカラムプロファイルを実行すると、num-mappers 引数を設定して並行処理を実現しパフォーマンスを最適化することができます。また、Sqoop で作業単位を分割する際の基準のカラムを指定するように、split-by 引数も設定しなければなりません。

以下の構文を使用します。

```
--split-by <column_name>
```

プライマリキーの範囲の上限と下限の間で値の均等分布がない場合、データの均等分布がある別のカラムを指定して作業単位を分割するように、split-by 引数を設定できます。

split-by カラムを定義しない場合、Sqoop では次の条件に基づいて作業単位が分割されます。

- データオブジェクトに単一プライマリキーが含まれる場合、Sqoop はプライマリキーを split-by カラムとして使用します。
- データオブジェクトにコンポジットプライマリキーが含まれる場合、Sqoop のデフォルトは、split-by 引数を使用せずにコンポジットプライマリキーを処理する Sqoop の動作になります。詳細については、Sqoop のマニュアルを参照してください。
- データオブジェクトに同一のカラムを持つ 2 つのテーブルが含まれる場合、split-by カラムをテーブル修飾名で定義する必要があります。例えば、テーブル名が CUSTOMER であり、カラム名が FULL\_NAME の場合、split-by カラムを次のように定義します。  
--split-by CUSTOMER.FULL\_NAME
- データオブジェクトにプライマリキーがない場合、m 引数の値と num-mappers 引数は、デフォルトで 1 になります。

Cloudera Connector Powered by Teradata または Hortonworks Connector for Teradata を使用し、Teradata テーブルにプライマリキーがない場合、split-by 引数は必須です。

## Informatica Analyst のオペレーティングシステムプロファイルの概要

Analyst ツールでオペレーティングシステムプロファイルを選択できます。オペレーティングシステムプロファイルを選択すると、データ統合サービスによってカラムプロファイル、エンタープライズ検出プロファイル、およびスコアカードが作成および実行されます。この処理は、オペレーティングシステムプロファイルユーザーに定義された権限に基づいて行われます。

Analyst ツールでは、デフォルトのプロファイルを使用して、プロファイルとスコアカードが実行されます。オペレーティングシステムプロファイルが 1 つしかない場合は、そのオペレーティングシステムプロファイルがデフォルトで選択されます。オペレーティングシステムプロファイルが複数ある場合は、それらの中から 1 つを選択できます。



## オペレーティングシステムプロファイルの選択

Informatica Analyst でオペレーティングシステムプロファイルを選択できます。データ統合サービスでは、プロファイリングジョブを実行するために、オペレーティングシステムプロファイルユーザーの権限を使用します。

1. Informatica Analyst のヘッダー領域で、[<ユーザー名>] > [設定] をクリックします。  
[設定] ダイアログボックスが表示されます。
2. オペレーティングシステムプロファイルを選択します。[保存] をクリックします。

## リポジトリアセットのロックとチームベース開発の概要

モデルリポジトリでは、ユーザーが他のユーザーの作業を上書きしないように、プロファイルがロックされます。モデルリポジトリがバージョン管理システムと統合されている場合は、複数のバージョンのアセットが保存され、バージョンにバージョン番号が割り当てられます。プロファイルのチェックアウトとチェックイン、チェックアウトの取り消しなどを実行できます。チェックアウトしたプロファイルの特定のバージョンを表示できます。

Analyst ツールでプロファイルの編集を開始すると、他のユーザーがそのプロファイルを編集できないようにプロファイルがロックされます。プロファイルを保存すると、ロック状態が保持されます。プロファイルを閉じると、プロファイルのロックは解除されます。

モデルリポジトリでは、バージョン管理されたアセット管理を使用してチームの他のメンバがプロファイルを上書きできないように、プロファイルが保護されます。別のユーザーによってチェックアウトされたプロファイルを編集しようとする、そのプロファイルをチェックアウトしたユーザーを示す通知を受け取ります。チェックアウト済みのプロファイルは、読み取り専用モードで開くことができるほか、別の名前でも保存することもできます。

[プロファイルのプロパティ] ダイアログボックスでプロファイルのバージョンを選択し、そのバージョンのプロファイル定義を表示できます。[アクション] メニューで [プロファイルのプロパティ] オプションにアクセスできます。リポジトリアセットのロックおよびバージョン管理されたアセットの管理の詳細については、『Analyst ツールガイド』を参照してください。

# Informatica Analyst でのカラムプロファイルの作成

カスタムプロファイルまたはデフォルトプロファイルを作成できます。カスタムプロファイルを作成する際には、カラム、サンプル行、ドリルダウンオプションを設定できます。デフォルトプロファイルを作成すると、カラムプロファイルとデータドメイン検出がすべてのデータドメインの全データセットに対して実行されます。

1. **【検出】** ワークスペースで、**【プロファイル】** をクリックするか、ヘッダ領域から **【新規】** > **【プロファイル】** を選択します。

**注:** **【ライブラリ】** ワークスペースでデータオブジェクトを右クリックして、プロファイルを作成できます。このプロファイルでは、データオブジェクトのプロパティからプロファイル名、場所名、およびデータオブジェクトが抽出されます。デフォルトのプロファイルを作成するか、設定をカスタマイズして、カスタムプロファイルを作成できます。

**新規プロファイルウィザード**が表示されます。

2. **【単一ソース】** オプションがデフォルトで選択されています。**【次へ】** をクリックします。
3. **【全般プロパティの指定】** 画面で、プロファイルの名前と説明（省略可能）を入力します。**【場所】** フィールドで、プロファイルの作成先のプロジェクトまたはフォルダを選択します。**【次へ】** をクリックします。
4. **【ソースの選択】** 画面で、**【選択】** をクリックしてデータオブジェクトを選択するか、**【新規】** をクリックしてデータオブジェクトをインポートします。**【次へ】** をクリックします。

- **【データオブジェクトの選択】** ダイアログボックスで、データオブジェクトを選択します。**【OK】** をクリックします。

**【プロパティ】** ペインに、選択したデータオブジェクトのプロパティが表示されます。**【データプレビュー】** ペインに、データオブジェクトのカラムが表示されます。

- **【新しいデータオブジェクト】** ダイアログボックスで、接続、スキーマ、テーブル、またはビューを選択し、プロファイルの作成、場所の選択、データオブジェクトのインポート先フォルダの作成ができます。**【OK】** をクリックします。

5. **【ソースの選択】** 画面で、プロファイルを実行するカラムを選択します。**【名前】** を選択すると、すべてのカラムを選択できます。**【次へ】** をクリックします。

デフォルトでは、すべてのカラムが選択されています。Analyst ツールによって、名前、データ型、精度、スケール、NULL 可能、および各カラムがプライマリキーに含まれているかどうかなど、カラムのプロパティがリストされます。

6. **【設定の指定】** 画面で、カラムプロファイル、データドメイン検出、またはカラムプロファイルとデータドメイン検出のうちどれを実行するかを選択します。デフォルトでは、カラムプロファイルオプションが選択されています。

- **【カラムプロファイルを実行】** を選択して、カラムプロファイルを実行します。

- **【データドメイン検出を実行】** を選択して、データドメイン検出を実行します。**【データドメイン】** ペインで、検出するデータドメインを選択して、適合条件を選択し、**【データドメイン検出のカラム選択の編集】** ダイアログボックスでデータドメイン検出対象カラムを選択します。

- **【カラムプロファイルを実行】** と **【データドメイン検出を実行】** を選択して、カラムプロファイルとデータドメイン検出を実行します。**【データドメイン】** ペインでデータドメインオプションを選択します。

**注:** デフォルトでは、選択したカラムはカラムプロファイルとデータドメイン検出の対象となります。**【編集】** をクリックして、データドメイン検出のカラムを選択または選択解除します。

- データドメイン検出の実行対象として、**【データ】**、**【カラム】**、**【データとカラム】** を選択します。

- サンプルングオプションを選択します。**【次に対してプロファイルを実行】** ペインで、サンプルングオプションとして **【すべての行（完全分析）】**、**【サンプリング最初の】**、**【ランダムサンプリング】**、**【ランダムサンプリング（自動）】**、**【Limit N】**、または **【ランダムパーセンテージ】** を選択できます。このサンプリングオプションはカラムプロファイルとデータドメイン検出に適用されます。

- ドリルダウンオプションを選択します。[ドリルダウン] ペインでは、[ライブ] または [ステージングされた] ドリルダウンオプションを選択したり、[オフ] を選択してドリルダウンを無効にしたりできます。必要に応じて、[カラムの選択] をクリックして、ドリルダウンするカラムを選択します。承認済みのデータ型またはデータドメインを持つカラムの場合は、データ型およびデータドメインの推測を省略できます。
  - ランタイム環境として、[ネイティブ]、[Blaze]、[Spark]、または [Databricks] を選択します。[Blaze] または [Spark] を選択した場合は、[選択] をクリックして、[Hadoop 接続の選択] ダイアログボックスで Hadoop 接続を選択します。[Databricks] を選択した場合は、[選択] をクリックして、Databricks 接続を選択します。
7. [次へ] をクリックします。  
[ルールとフィルタの指定] 画面が開きます。
  8. [ルールとフィルタの指定] 画面では、次のタスクを実行できます。
    - ルールを作成、編集、または削除します。既存のルールをプロファイルに適用できます。
    - フィルタを作成、編集、または削除します。

注: このプロファイルでスコアカードを作成した場合、このプロファイル用に作成したフィルタを再利用できます。
  9. [保存して完了] をクリックしてプロファイルを作成するか、[保存して実行] をクリックしてプロファイルを作成および実行します。

## カラムプロファイルの編集

プロファイル実行後にカラムプロファイルに変更を加えることができます。

1. [ライブラリ] ワークスペースでプロファイルを含むプロジェクトを選択するか、[アセット] ペインでプロファイルを選択します。
2. プロファイル名をクリックします。  
サマリビューが [検出] ワークスペースに表示されます。
3. バージョン管理システムが有効になっている場合、[アクション] > [チェックアウト] をクリックしてプロファイルをチェックアウトします。
4. [アクション] > [プロファイルの編集] をクリックします。  
[プロファイル] ウィザードが表示されます。
5. 変更する内容に基づいて、次のページオプションのいずれかを選択します。
  - **全般プロパティの指定。** 名前、説明、場所などの基本プロパティを変更します。
  - **ソースの選択。** プロファイルの実行対象として、別の一致するデータソースとカラムを選択します。
  - **設定の指定。** カラムプロファイルまたはカラムプロファイルとデータドメイン検出を実行するように選択します。検出対象のデータドメインを選択し、データドメイン検出、サンプリング、およびドリルダウンのオプションを変更します。
  - **ルールとフィルタの指定。** ルールとフィルタを作成、編集、または削除します。
6. [保存して完了] をクリックしてプロファイルの編集を完了するか、[保存して実行] をクリックしてプロファイルを編集および実行します。
7. バージョン管理システムが有効な場合、次のタスクを実行する必要があります。
  - [保存して完了] をクリックしてプロファイルの編集を完了します。

- [サマリ] ビューで、[チェックイン] をクリックしてプロファイルをチェックインします。
- [アクション] > [プロファイルの実行] をクリックしてプロファイルを実行します。

## プロファイルの実行

プロファイルを実行してデータソースの内容および構造を分析し、ドリルダウン対象のカラムとルールを選択します。カラムとルールの実データまたはステージングされたデータをドリルダウンできます。初期プロファイルの実行後は、すべてのソースカラムに対してプロファイルを実行せずに、特定のカラムまたはルールに対してのみプロファイルを実行できます。

1. [ライブラリ] ワークスペースから、[プロジェクト] ペインで目的のプロファイルが格納されているプロジェクトまたはフォルダを選択するか、[アセット] ペインでプロファイルを選択します。

2. [アクション] > [開く] をクリックします。

サマリビューが [検出] ワークスペースに表示されます。

3. [アクション] > [プロファイルの実行] をクリックします。

Analyst ツールによってプロファイルが実行され、サマリビューにプロファイル結果が表示されます。プロファイルサマリログファイルおよびマッピングログファイルを表示して、Analyst ツールによって実行されたタスクの詳細を取得できます。

**注:** オペレーティングシステムプロファイルを使用してプロファイルを実行した場合、サマリログはデータ統合サービス用に構成されたログディレクトリにあり、マッピングログはオペレーティングシステムのプロファイル用に構成されたログディレクトリにあります。

## Spark エンジンでのプロファイルの実行

Spark エンジンで JDBC 接続を使用してプロファイルを実行すると、プロファイルの実行が失敗します。

Spark エンジンでプロファイルを実行する前に、次の手順を実行します。

1. JDBC ウェアハウス接続を作成します。
2. データを抽出するデータベースの Data Direct JAR ファイルを取得します。
3. ファイルを次の場所にダウンロードします。<INFA\_HOME>/externaljdbcjars

## 同期オプション

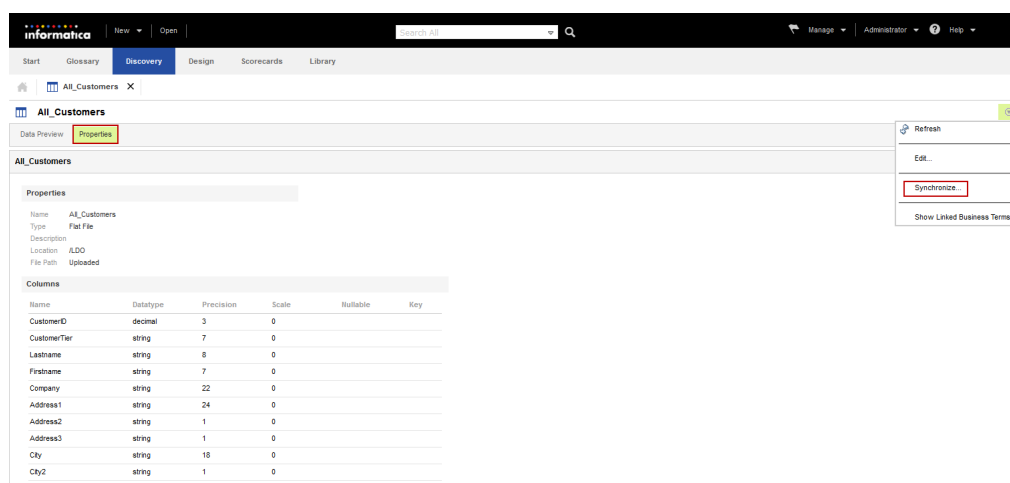
外部データソースのメタデータを変更する場合、モデルリポジトリ内のデータオブジェクトのメタデータはデフォルトでは更新されません。データオブジェクトのメタデータをデータソースのメタデータと同期させるには、[同期] オプションを使用します。[同期] オプションは、カラムプロファイル、エンタープライズ検出プロファイル、およびスコアカードに使用できます。外部データソースは、リレーショナルデータソースまたはフラットファイルデータソースにすることができます。

# Informatica Analyst でのフラットファイルデータオブジェクトの同期

Analyst ツールで、外部フラットファイルデータソースに加えられた変更をそのデータオブジェクトに同期させることができます。データオブジェクトを同期させるには、**フラットファイルの同期**ウィザードを使用します。

1. **【ライブラリ】** ワークスペースを開きます。
2. **【プロジェクト】** セクションで、フラットファイルデータオブジェクトをプロジェクトから選択します。  
Analyst ツールは、**【プロパティ】** タブにテーブルのプロパティを表示します。
3. **【アクション】** メニューから **【同期】** をクリックします。

次の図は、**【アクション】** メニューの **【プロパティ】** タブと **【同期】** オプションを示しています。



**【フラットファイルの同期】** ウィザードが表示されます。

4. 場所を参照するか、フラットファイルをインポートするネットワークパスを入力します。
  - 場所を参照するには、**【ファイルの選択】** をクリックして、マシンからアクセスできるディレクトリからフラットファイルを選択します。
  - ネットワークパスを入力するには、**【ネットワークパスの入力】** を選択して、ファイルパスとファイル名を設定します。

次の図は、フラットファイルの同期ウィザードを示しています。

### Synchronize Flat File: Step 1 of 5

Specify a location to import the flat file from and specify how to import the flat file.

☒ Browse and Upload:

Browse...

No file selected.

☐ Enter a Network Path:

☐ Hadoop File System

#### Description

Upload files from a local machine. Recommended for smaller files up to 10 MB. The Analyst tool uploads a copy of the file to the node on which the Analyst Service runs. Upload the file again if you modify the file.



Back

Next

Finish

Cancel

5. **【次へ】** をクリックします。
6. 区切られたフラットファイルまたは固定幅フラットファイルのインポートを選択します。
  - 区切られたフラットファイルをインポートするには、**【区切り記号付き】** オプションを選択します。
  - 固定幅フラットファイルをインポートするには、**【固定幅】** オプションを選択します。
7. **【次へ】** をクリックします。
8. 区切りフラットファイルまたは固定幅フラットファイルのオプションを設定します。
9. **【次へ】** をクリックします。
10. 必要に応じて、カラムの属性を変更します。
11. **【次へ】** をクリックします。
12. デフォルト名をそのまま使用するか、フラットファイルの別の名前を入力します。
13. 必要に応じて、説明を入力します。
14. **【完了】** をクリックします。

操作の確認を求める同期メッセージが表示されます。
15. **【はい】** をクリックして、フラットファイルを同期します。

同期が完了したことを示すメッセージが表示されます。メタデータの変更の詳細を表示するには、**【詳細を表示】** をクリックします。
16. **【OK】** をクリックします。

## Informatica Analyst でのリレーショナルデータオブジェクトの同期

外部リレーショナルデータソースに加えられた変更をそのテーブルデータオブジェクトに同期させることができます。外部データソースの変更には、ソースカラムとルールカラムの追加、変更、削除が含まれます。

1. **【ライブラリ】** ワークスペースを開きます。
2. **【プロジェクト】** セクションで、テーブルデータオブジェクトをプロジェクトから選択します。  
Analyst ツールは、**【プロパティ】** タブにテーブルのプロパティを表示します。
3. **【アクション】** メニューから **【同期】** をクリックします。  
操作の確認を求めるメッセージが表示されます。
4. 同期プロセスを完了するには、**【はい】** をクリックします。  
同期ステータスメッセージが表示されます。
5. 同期が完了したことを示すメッセージが表示されます。  
メタデータの変更の詳細を表示するには、**【詳細を表示】** をクリックします。
6. **【OK】** をクリックします。

## 第 7 章

# Informatica Analyst のルール

この章では、以下の項目について説明します。

- [Informatica Analyst のルールの概要, 48 ページ](#)
- [定義済みルール, 48 ページ](#)
- [式ルール, 50 ページ](#)

## Informatica Analyst のルールの概要

ルールは、カラムプロファイルの実行時にソースデータに適用される条件を定義するビジネスロジックです。ルールをプロファイルに追加して、データを検証できます。

ルールはさまざまな状況で使用します。例えば、1 つ以上のデータカラムをクレンジングするルールを追加したり、ソースデータでは提供されない情報を提供するルックアップルールを追加できます。また、データ品質プロジェクトやデータ統合プロジェクト用のクレンジングルールを検証するルールを追加することもできます。

カラムプロファイルを作成または編集する場合は、ルールを作成してプロファイルに追加したり、既存のルールをプロファイルに適用したりすることができます。カラムプロファイルでは、式ルールまたは定義済みのルールを使用できます。

プロファイルを実行すると、Analyst ツールによって、[概要] ビューの [ルール] 列にプロファイル結果が表示されます。特定のルールのカラム結果は、詳細ビューで確認できます。ルールの出力を 1 つ以上の仮想カラムにすることができます。仮想カラムはプロファイル結果に存在します。Analyst ツールは仮想カラムに対してプロファイルを実行します。例えば、姓と名が含まれている 1 つのカラムを FIRST\_NAME と LAST\_NAME の仮想カラムに分割する定義済みルールを使用するとします。Analyst ツールは、FIRST\_NAME カラムと LAST\_NAME カラムに対してプロファイルを実行します。

**注:** 他のオブジェクトタイプによって参照されているルールオブジェクトを削除しようとすると、それらのオブジェクトタイプの一覧を示すメッセージが表示されます。ルールを削除する前に、その削除の影響を判断するようにします。

## 定義済みルール

定義済みルールは、Developer tool で作成されたルール、または Developer tool と Analyst ツールに付属しているルールです。定義済みルールをカラムプロファイルに適用して、ソースデータを変更または検証します。

定義済みルールでは、トランスフォーメーションを使用してルールロジックを定義します。定義済みルールは複数のプロファイルで使用できます。モデルリポジトリでは、定義済みルールは入力グループ、出力グループ、およびルールロジックを定義するトランスフォーメーションが含まれるマプレットです。



## 定義済みルールのプロセス

プロファイルに定義済みルールを適用するには、**新しいルール**ウィザードを使用します。

定義済みルールは、次の手順に従って適用できます。

1. プロファイルを開きます。
2. 定義済みルールを選択します。
3. ルールのパラメータを確認します。
4. 入力カラムを選択します。ルールを複数のカラムに適用する必要がある場合は、複数のカラムを選択してもかまいません。
5. プロファイリングのオプションを設定します。

## 定義済みルールの適用

定義済みルールを適用するときは、ルールを選択し、そのルールの入力カラムと出力カラムを設定します。再利用可能なルールとして格上げされたルールを使用するか、開発者が作成したルールを使用するには、定義済みルールを適用します。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトを選択するか、**【アセット】** ペインでプロファイルを選択します。
2. **【アクション】** > **【開く】** をクリックしてプロファイルを開きます。  
サマリビューが**【検出】** ワークスペースに表示されます。
3. **【アクション】** > **【プロファイルの編集】** をクリックします。  
**【プロファイルウィザード】** が表示されます。
4. **【ルールとフィルタの指定】** をクリックします。
5. **【ルールとフィルタの指定】** 画面の**【ルール】** パネルで、**【アクション】** > **【既存ルールの適用】** をクリックします。  
**【ルールの適用ウィザード】** ダイアログボックスが表示されます。
6. ルールを選択し、**【次へ】** をクリックします。
7. **【追加】** をクリックします。  
**【入力ポートのカラムの選択】** ダイアログボックスが表示されます。
8. フィールドと入力カラムを選択します。**【OK】** をクリックします。  
入力カラムと出力カラムが**【ルールの適用ウィザード】** ダイアログボックスに表示されます。
9. **【ルールの適用ウィザード】** ダイアログボックスで、**【OK】** をクリックします。  
**【ルールとフィルタの指定】** 画面にルールが表示されます。

# 式ルール

式ルールでは、式関数およびカラムを使用してルールロジックを定義します。式ルールを作成して、Analyst ツールのカラムプロファイルに追加します。

式ルールを使用して、カラムプロファイルのカラムの値を変更または検証します。1 つ以上の式ルールを作成して 1 つのプロファイルで使用できます。式関数は、ソースデータの変換に使用される SQL のような関数です。式ルールロジックを作成するときは、次のタイプの関数を使用できます。

- 文字
- 変換
- データクレンジング
- 日付
- エンコーディング
- 財務
- 数値
- 科学
- 特殊
- テスト

次のメソッドを使用して、式ルールを作成できます。

- プロファイルウィザード。列プロファイルを作成または編集するときは、プロファイルウィザードで式ルールを作成して適用できます。ルールを再利用可能なルールに昇格させ、複数のプロファイルで使用することができます。
- ルール仕様。Analyst ツールでルール仕様を構成し、カラムプロファイルでルール仕様を使用することができます。ルール仕様を設定するときは、ビジネスルールの要件を 1 つ以上のルール文に変換します。ルール文は、データセットがビジネスルールに準拠しているかどうかを判定するロジックです。ルール仕様からマップレットを生成し、Developer tool で作成するカラムプロファイルでマップレットを使用します。

式エディタを使用して式関数を追加し、カラムを関数への入力として設定して式を検証し、戻り型、精度、およびスケールを設定できます。式のルールを作成および検証したら、出力ルールカラムの精度値を編集できます。デフォルトでは、出力ルールカラムの精度値は 10 に設定されています。出力ルールカラムが設定精度値を超えている場合は、精度値が切り詰められます。

式ルールの出力は、ルールの名前をカラム名として使用する仮想カラムです。Analyst ツールが仮想カラムに対してカラムプロファイルを実行します。例えば、式ルールを使用して郵便番号を検証するとします。このルールでは、郵便番号が有効なら 1 が、無効なら 0 が返されます。ルールの出力値 1 と 0 に対して、Informatica Analyst がカラムプロファイルを実行します。

## 式ルールの作成

**【プロファイル】** ウィザードを使用すると、式ルールを作成してプロファイルに追加できます。プロファイルのカラムの値を検証するための式ルールを作成します。

1. プロファイルを開きます。
2. サマリビューで、**【アクション】** > **【プロファイルの編集】** をクリックして、**【プロファイル】** ウィザードを開きます。
3. **【ルールとフィルタの指定】** をクリックします。
4. **【ルール】** ペインで、**【アクション】** > **【ルールの追加】** をクリックします。



【新しいルール】 ダイアログボックスが表示されます。

5. 【新しいルール】 ダイアログボックスで、ルールの名前と説明（省略可能）を入力します。ルールは、[関数] パネルまたは[カラム] パネルで作成できます。
  - [関数] パネルでは、関数カテゴリを選択して、右矢印ボタン（>>）をクリックします。ダイアログボックスでパラメータを指定して、[OK] をクリックします。  
関数、およびカラムと値が、[式] パネルに表示されます。
  - [カラム] パネルでカラムを選択し、右矢印ボタン（>>）をクリックします。[式] パネルにカラムが表示されます。関数、式、値を追加してルールを作成します。
6. ルールを確認するには、[検証] をクリックします。
7. 必要に応じて、ルールを再利用可能なルールとして格上げするように選択し、プロジェクトとフォルダーの場所を設定します。ルールを再利用可能なルールに格上げすると、他のユーザーもこのルールを別のプロファイルで定義済みルールとして使用できるようになります。
8. [OK] をクリックします。  
【ルールとフィルタの指定】 画面で、[ルール] ペインにルールが表示されます。

## ルール仕様を使用した式ルールの作成

ルール仕様を使用して、Informatica Analyst で式ルールを作成できます。ルールを列プロファイルに追加して、データを検証できます。

1. ヘッダー領域で、[新規] > [ルール仕様] をクリックします。  
【新しいルール仕様】 ウィザードが表示されます。

2. **【新しいルール仕様】** ダイアログボックスで、ルールの名前と説明（省略可能）を入力します。
3. **【場所】** フィールドで、**【参照】** をクリックして、ルールを保存するプロジェクトまたはフォルダを選択します。
4. **【続行】** をクリックします。  
ルール仕様は、**【デザイン】** ワークスペースに表示されます。
5. ルールのプロパティを入力するには、ルールで最上位の八角形の図形を選択し、**【プロパティ】** をクリックします。
6. プライマリルールセットを構成するには、ルールの **【次のレベルの四角形】** をクリックします。
7. ルールセットの入力を入力するには、**【プロパティ】** > **【入力】** をクリックします。  
**【入力管理】** ダイアログボックスが表示されます。
8. **【入力管理】** ダイアログボックスで、**【入力の追加】** をクリックし、入力の名前、データ型、最大長、および説明を入力します。必要に応じて、複数の入力を入力できます。
9. **【OK】** をクリックします。  
入力は、**【プロパティ】** セクションに表示されます。
10. ルールロジックを定義するには、**【ルールロジック】** をクリックし、演算子、条件を入力して、**【アクション】** リストでアクションを選択します。
11. 必要に応じて、複数のルールセットを入力します。
12. ルールを検証するには、**【検証】** () アイコンをクリックします。
13. 列プロファイルにルール仕様を保存して使用するには、**【保存して終了】** をクリックします。
14. ルールを保存して作業を続行するには、**【保存して続行】** をクリックします。
15. Developer tool でルールの仕様を使用するには、**【ルールの生成】** () アイコンをクリックしてマップレットを生成します。  
Analyst ツールによりモデルリポジトリにマップレットが作成されます。マップレットをルールとして検証し、Developer tool で作成したカラムプロファイルでマップレットを使用します。

## 第 8 章

# Informatica Analyst のフィルタ

この章では、以下の項目について説明します。

- [Informatica Analyst のフィルタの概要, 53 ページ](#)
- [フィルタの作成, 53 ページ](#)
- [フィルタの管理, 56 ページ](#)

## Informatica Analyst のフィルタの概要

フィルタを作成すると、フィルタ条件を満たす、元のデータソースのサブセットを作成することができます。その後、フィルタリングされたデータに対してプロファイルを実行できます。

フィルタを作成すると、プロファイル結果のうちフィルタ条件を満たすものを表示できます。サマリビューで使用可能なデフォルトフィルタを使用してプロファイル結果を表示できます。

## フィルタの作成

フィルタを作成すると、フィルタ条件を満たす、元のデータソースのサブセットを作成することができます。

1. プロファイルを開きます。
2. サマリビューで、**[アクション]** > **[プロファイルの編集]** をクリックします。  
**[プロファイル]** ウィザードが表示されます。
3. **[ルールとフィルタの指定]** をクリックします。
4. **[フィルタ]** ペインで、**[アクション]** > **[フィルタの追加]** をクリックします。  
**[新規フィルタ]** ダイアログボックスが表示されます。
5. 簡易フィルタ、詳細フィルタ、または SQL フィルタを作成します。

**注:** 日付カラムに対する簡易フィルタまたは詳細フィルタでは、YYYY/MM/DD HH:MM:SS 形式で条件を指定してください。

**[データプレビュー]** ペインには、フィルタ条件を満たす、元のデータソースのサブセットが表示されます。

6. **[OK]** をクリックします。  
**[ルールとフィルタの指定]** 画面で、**[フィルタ]** ペインにフィルタが表示されます。

## 簡易フィルタの作成

次の条件演算子を使用して簡易フィルタを作成できます: =、!=、>、<。簡易フィルタを使用すると、元のデータソースのサブセットを作成できます。

1. **【新規フィルタ】** ダイアログボックスで **【簡易】** をクリックします。

次の画像は、**【新規フィルタ】** ダイアログボックスで簡易フィルタを作成するために使用できるオプションを示しています。

New Filter

Create a filter. The filter is used to create a subset of the data rows before profiling.

Name\*:

Description:

Choose the filter type\*: ☒ Simple ☐ Advanced ☐ SQL

Columns	Operator	Values(s)
-Select-	-Select-	+

Filter Preview

2. 名前を入力し、必要に応じて説明を入力します。
  3. カラムを選択します。
  4. 条件演算子を選択します。
  5. 値を入力します。
  6. 必要な場合、プラス (+) アイコンをクリックしてフィルタを追加できます。
  7. **【OK】** をクリックします。
- 【ルールとフィルタの指定】** ページで、**【フィルタ】** ペインにフィルタが表示されます。

## 詳細フィルタの作成

AND、OR、NOT などの式を使用して詳細フィルタを作成することで、元のデータソースのサブセットを作成できます。

1. **【新規フィルタ】** ダイアログボックスで **【詳細】** をクリックします。

次の画像は、**【新規フィルタ】** ダイアログボックスの詳細フィルタオプションを示しています。

2. 詳細フィルタの名前と説明（省略可能）を入力します。
3. 詳細フィルタは、**【関数】** パネルまたは **【カラム】** パネルを使用して作成できます。
  - **【関数】** パネルでは、関数カテゴリを選択して、右矢印ボタン（>>）をクリックします。ダイアログボックスでパラメータを指定して、**【OK】** をクリックします。関数、およびカラムと値が、**【式】** パネルに表示されます。
  - **【カラム】** パネルでカラムを選択し、右矢印ボタン（>>）をクリックします。**【式】** パネルにカラムが表示されます。関数、式、値を追加して詳細フィルタを作成します。
4. 詳細フィルタを確認するには、**【検証】** をクリックします。
5. **【OK】** をクリックします。

【ルールとフィルタの指定】画面で、**【フィルタ】** ペインにフィルタが表示されます。

## SQL フィルタの作成

SQL クエリを使用して SQL フィルタを作成できます。リレーショナルデータソースに対して SQL フィルタを作成できます。

1. **【新規フィルタ】** ダイアログボックスで **【SQL】** をクリックします。

次の画像は、**【新規フィルタ】** ダイアログボックスの SQL フィルタオプションを示しています。

The screenshot shows the 'New Filter' dialog box with the 'SQL' tab selected. The dialog has a title bar 'New Filter' with a close button. Below the title bar is a subtitle: 'Create a filter. The filter is used to create a subset of the data rows before profiling.' There are two text input fields: 'Name\*' and 'Description:'. Below these fields are three radio buttons for 'Choose the filter type\*': 'Simple', 'Advanced', and 'SQL' (which is selected). Below the radio buttons is a section titled 'Type or paste a SQL statement' with a 'Validate' button. At the bottom of the dialog is a 'Filter Preview' section with a refresh icon. The bottom of the dialog has a help icon (?) and 'Ok' and 'Cancel' buttons.

2. 必要に応じて SQL フィルタの説明を入力します。
3. テキストボックスに、SQL クエリを入力するか貼り付けます。
4. **【検証】** をクリックして、SQL クエリを検証します。
5. **【OK】** をクリックします。

**【ルールとフィルタの指定】** ページで、**【フィルタ】** ペインに SQL フィルタが表示されます。

## フィルタの管理

フィルタを編集および削除できます。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトを選択するか、**【アセット】** ペインでフィルタリングするプロファイルを選択します。
2. プロファイルを開きます。
3. サマリビューで、**【アクション】** > **【プロファイルの編集】** をクリックして、**【プロファイル】** ウィザードを開きます。
4. **【ルールとフィルタの指定】** をクリックします。



5. [フィルタ] ペインでフィルタを選択して、[アクション] > [フィルタの編集] をクリックします。  
[フィルタの編集] ダイアログボックスが表示されます。
6. フィルタを編集して、[OK] をクリックします。
7. フィルタを削除するには、フィルタを選択して、[アクション] > [フィルタの削除] をクリックします。

## 第 9 章

# Informatica Analyst のカラムプロファイル結果

この章では、以下の項目について説明します。

- [Informatica Analyst のカラムプロファイル結果の概要, 58 ページ](#)
- [サマリビュー, 59 ページ](#)
- [詳細ビュー, 61 ページ](#)
- [統計, 63 ページ](#)
- [プロファイル実行のタイプ, 70 ページ](#)
- [複数のプロファイル結果の比較の概要, 72 ページ](#)
- [カラムプロファイルのドリルダウン, 76 ページ](#)
- [Analyst ツールでのキュレーション, 77 ページ](#)
- [Informatica Analyst のカラムプロファイルのエクスポートファイル, 78 ページ](#)

## Informatica Analyst のカラムプロファイル結果の概要

プロファイル結果を表示して、データの内容、構造、品質を把握し分析できます。サマリビューでは、プロファイルのすべてのカラムとルールを確認できます。詳細ビューでは、特定のカラムまたはルールの詳細なプロパティを確認できます。

プロファイル結果は、**[検出]** ワークスペースで確認できます。ビューヘッダーには、プロファイルのタイプ、プロファイル内のカラム数、プロファイル内のルール数、サンプリングデータ、作成日時が表示されます。

サマリビューでは、各カラムのプロパティを、値、水平棒グラフ、またはパーセンテージとして表示できます。NULL 値、重複なしの値、重複ありの値、パターン、データ型、データドメインなどのカラムプロパティを表示できます。サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。

詳細ビューでは、NULL 値、重複なしの値、重複ありの値、推測されたデータ型、推測されたデータドメイン、推測されたパターン、値、ビジネス用語を表示し、各ペインでデータをプレビューできます。

最新の実行、履歴実行、統合済み実行のプロファイル結果を表示できます。2 つのプロファイル実行のプロファイル結果を比較し、サマリビューおよび詳細ビューで結果を確認できます。プロファイル統計情報の確認とデータのキュレーションができます。プロファイル統計情報には、値、データ型、異常値、カラムとルールの統計情報が含まれます。データの検出とドリルダウンを実行できます。

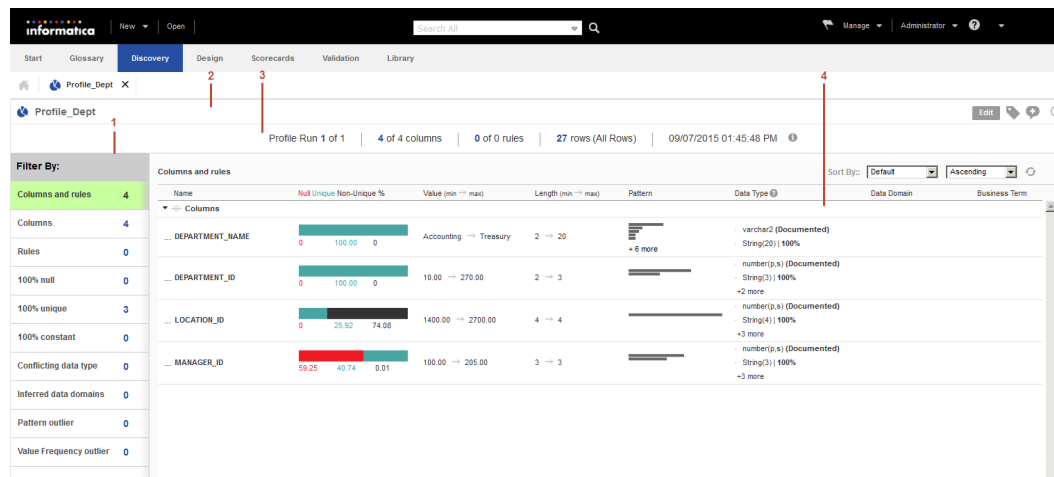
**注:** Avro、JSON、Parquet、および XML データソースに対して、プロファイルを表示および実行できます。最新の実行、履歴実行、統合済み実行のプロファイル結果を表示し、2 つのプロファイル実行のプロファイル結果を比較できます。

値の頻度、パターンの頻度、ドリルダウンデータ、コメント、タグ、およびビジネス用語を CSV ファイルにエクスポートできます。プロファイルのサマリ情報を Microsoft Excel ファイルにエクスポートすることで、すべてのデータをファイル内で確認して詳しく分析できます。プロファイル結果でルール情報を確認できます。表示されるプロファイル結果は、プロファイル設定およびサンプリングのオプションによって異なります。

## サマリビュー

プロファイル結果のサマリは、サマリビューにグリッド形式で表示されます。サマリビューでデフォルトのフィルタを使用して、特定の統計を表示できます。例えば、[ルール] を選択すると、サマリビューにプロファイルのすべてのルールが表示されます。

次の図は、サマリビューのグラフィカルビューの例を示しています。



1. デフォルトフィルタ。サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。
2. プロファイルヘッダー。ヘッダーでプロファイル名を表示できます。[編集] ボタンを使用してプロファイルの編集、タグおよびコメントアイコンを使用してタグやコメントの追加と編集をして、[アクション] メニューからオプションを選択します。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。プロファイルのプロファイル実行番号、プロファイル実行の総数、カラムおよびルールの数、および行数を表示できます。
4. サマリビュー。プロファイルのすべてのカラムとルールのプロパティを表示できます。

サマリビューでは、プロファイルの実行と編集、パターンまたは値頻度の異常値の検出、スコアカードへのカラムの追加、プロファイル実行の選択、2 つのプロファイル実行の比較、Microsoft Excel スプレッドシートへのプロファイル結果またはデータドメイン検出結果のエクスポート、複数カラムの推測結果の検証、コメントとタグの追加と削除、またはプロファイルプロパティの表示ができます。

## サマリビューのプロパティ

サマリビューには、プロファイルのすべてのカラムとルールのプロパティが表示されます。サマリビューには、プロパティを視覚的に表したものが表示されます。各サマリプロパティをクリックするとプロパティの値をソートできます。

次の表に、サマリプロパティのプロファイル結果を示します。

プロパティ	説明
名前	プロファイルのカラムまたはルールの名前が表示されます。
NULL、重複なし、重複ありの割合	カラムまたはルール出力について、NULL 値、重複なしの値、重複ありの値が割合で表示されます。値を水平棒グラフで表示できます。
パターン	カラムの複数のパターンが水平棒グラフで表示されます。マウスポインタを棒グラフの上に移動すると、パターンの特性やカラムに出現する類似パターンの数をパーセンテージとして確認できます。
値	カラムまたはルール出力の最小値と最大値が表示されます。
長さ	カラムまたはルール出力の値の最小長と最大長が表示されます。
データ型	<p>カラムまたはルールの文書化されたデータ型が表示されます。マウスをフィールド上に移動すると、推測されたデータ型が表示されます。Analyst ツールは次のデータ型を推測できます。</p> <ul style="list-style-type: none"><li>- String</li><li>- Varchar</li><li>- Decimal</li><li>- Integer</li><li>- Date</li></ul> <p>推測されたデータ型に基づいて一致率を表示することもできます。</p> <p><b>注:</b> Analyst ツールでは、精度が 38 より大きい数値カラムの値からデータ型を導出することはできません。また、精度が 255 より大きい文字列カラムの値からデータ型を導出することはできません。1800 年より前の年の値を含む日付カラムに対してカラムプロファイルを作成している場合、推測されたデータ型が固定長文字列として表示される可能性があります。その場合は、InferDateTimeConfig.xml の中で year-minimum パラメータのデフォルト値を必要に応じて変更してください。</p>
データドメイン	カラムに関連付けられているデータドメインの名前と、一致率と一致する行数が一緒に表示されます。
ビジネス用語	カラムに割り当てられたビジネス用語が表示されます。

## サマリビューのデフォルトフィルタ

サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。

サマリビューには、すべてのソースカラム、仮想カラム、およびルールカラムのプロファイル結果がデフォルトで表示されます。[フィルタ基準] ペインには、デフォルトフィルタを適用できるカラム数が表示されます。

サマリビューでは、次のデフォルトフィルタオプションを使用して、プロファイル結果を表示できます。

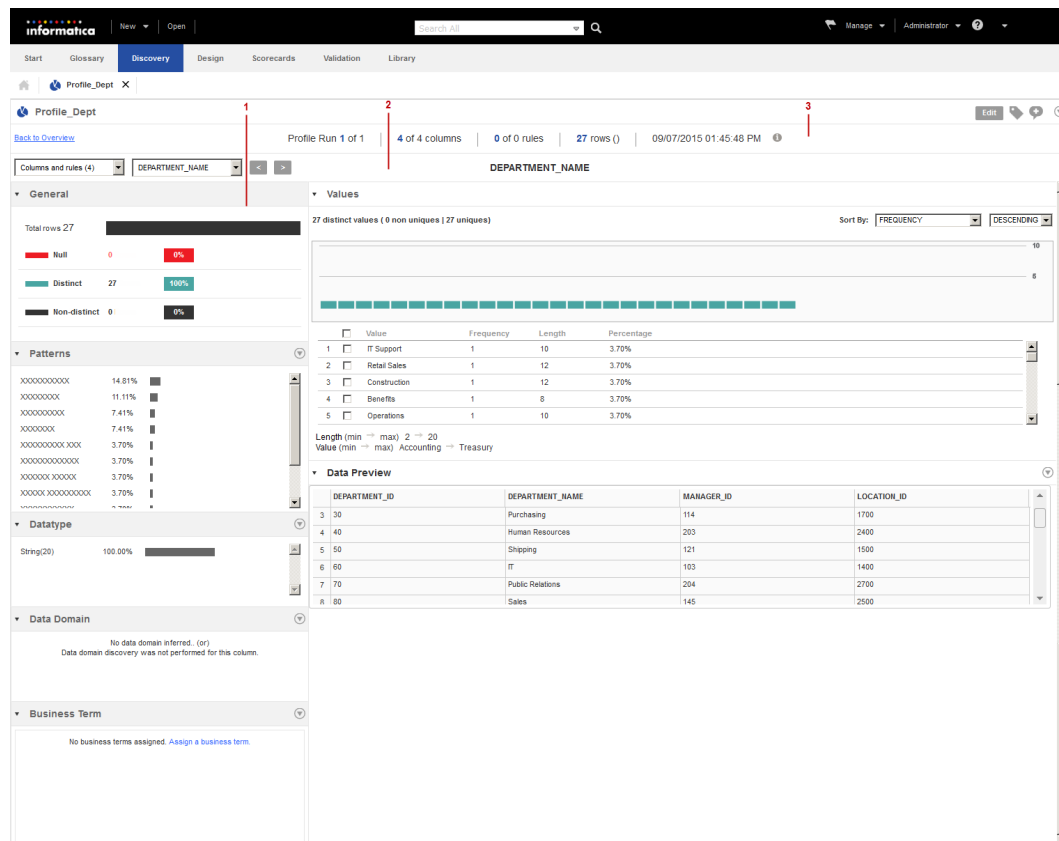
デフォルトフィルタオプション	説明
カラムとルール	ソースカラムとルールカラムのプロファイル結果を表示します。ソースカラムとルールカラムを展開および縮小して結果を表示できます。
カラム	ソースカラムのプロファイル結果を表示します。
ルール	ルールカラムのプロファイル結果を表示します。
100% NULL	100% NULL 値を持つカラムのプロファイル結果を表示します。
100%重複なし	100%重複のない値を持つカラムのプロファイル結果を表示します。
100%定数	すべてのレコードで同じ値を持つカラムのプロファイル結果を表示します。例えば、Country カラムのプロファイル結果が一律に"USA"である場合、100%定数フィルタには、Country カラムのプロファイル結果が含まれます。
競合しているデータ型	文書化されたデータ型と推測されたデータ型が一致しないカラムのプロファイル結果を表示します。例えば、CustomerTier カラムの文書化されたデータ型が Integer (2)、推測されたデータ型が String の場合、CustomerTier カラムはこのフィルタによって表示されます。
推測されたデータドメイン	推測されたデータドメインと設定済みデータドメインが同じカラムのプロファイル結果を表示します。
パターン異常値	パターン異常値を持つカラムのプロファイル結果を表示します。
値/頻度の異常値	値または頻度の異常値を持つカラムのプロファイル結果を表示します。

## 詳細ビュー

詳細ビューにカラム結果が表示されます。カラムプロパティの詳細を表示することができます。

カラムの詳細ビューは、サマリビューでカラムをクリックすると表示されます。

次の画像は、詳細ビューに表示されたカラムプロファイルのグラフィカルビューの例を示しています。



1. ペイン。全般プロパティ、カラム内の値、データプレビュー、推測されたパターン、推測されたデータ型、推測されたデータドメイン、およびビジネス用語を各ペインに表示できます。
2. カラムの詳細ヘッダ。ドロップダウンリストでカラムを選択するか、ナビゲーションボタンを使用して、カラム結果を表示できます。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。プロファイル実行と、プロファイル実行内のカラム数、ルール数、および行数のほか、プロファイル実行の日時を表示できます。

詳細ビューでは、プロファイルの実行または編集、スコアカードへのカラムの追加、プロファイル実行の選択、2つのプロファイル実行の比較、Microsoft Excel スプレッドシートへのプロファイル結果のエクスポートができます。また、値の頻度、パターンの頻度、データ型、選択した値のドリルダウンデータ、または選択したパターンのドリルダウンデータを csv ファイルにエクスポートしたり、カラムにコメントやタグを追加または削除したり、プロファイルプロパティを表示したりできます。

カラムのプロパティに対してさらなるアクションを実行するには、各ペインの [アクション] メニューを使用します。ペインは縮小または展開できます。

## [詳細ビュー] ペイン

詳細ビューの各ペインには、重複なしの値/重複ありの値/NULL 値の数と割合、パターン、推測されたデータ型、推測されたデータドメイン、値、データプレビュー、およびリンクされたビジネス用語などのカラムプロパティが表示されます。

カラムまたはルールをクリックできる場合は、そのカラムまたはビューの詳細ビューが開きます。

次の表に、詳細ビューの各ペインを示します。

ペイン	説明
全般	NULL 値、重複なしの値、重複ありの値を含む行の数が異なる色で表示されます。値をパーセンテージで表示できます。連続するプロファイル実行の各実行における全般値の増減をスパークラインとして表示できます。スパークラインには、最新の連続した 5 回のプロファイル実行について、NULL 値、重複なしの値、または重複ありの値の数の変動が折れ線グラフとして表示されます。スパークラインの各プロファイル実行にポインタを移動すると、値の数と割合を参照できます。カラムにはタグとコメントを追加できます。
パターン	カラム値のパターンが表示されます。カラムにパターンの出現する頻度が、水平棒グラフとパーセンテージで表示されます。パターンをドリルダウンしたり、参照テーブルにパターンを追加したり、選択したパターンを持つデータドメインを作成したりできます。
データ型	カラムの推測されたデータ型が表示されます。カラムのデータ型の頻度は、水平棒グラフとパーセンテージとして表示されます。データ型にドリルダウンすることができ、推測されたデータ型のうち選択したものを承認、拒否、またはリセットすることができます。 <b>【拒否済みを表示】</b> オプションには、拒否された推測データ型が表示されます。
データドメイン	カラムの推測データドメインが表示されます。適合する行、適合しない行、または NULL 値が含まれる行を対象に、データドメインでドリルダウンできます。データドメイン値を、承認、拒否、またはリセットできます。 <b>【拒否済みを表示】</b> オプションには、拒否されたデータドメインが表示されます。データドメイン値を確認できます。
ビジネス用語	カラムの割り当て済みビジネス用語が表示されます。カラムに対してビジネス用語の割り当てまたは割り当て解除を実行できます。
値	カラムのすべての値が、頻度、長さ、パーセンテージと共にグラフィカルに表示されます。それぞれの値にドリルダウンすることができます。値の参照テーブルへの追加、値の頻度ルールを作成、データドメインの作成を実行できます。
データプレビュー	選択したパターン、データ型、データドメイン、または値について、ドリルダウンデータが表示されます。

## 統計

統計情報（値、パターン、データ型、データドメイン、プロファイルのカラムとルールの異常値）を表示できます。

プロファイル統計情報はサマリビューで、カラム統計情報はサマリビューと詳細ビューで表示できます。最新のプロファイル実行、履歴プロファイル実行、統合済みプロファイル実行の統計情報を表示できます。2 つのプロファイル実行のプロファイル結果を比較し、プロファイルとカラムの統計情報をサマリビューおよび詳細ビューで確認できます。

## データプレビュー

[データプレビュー] ペインでは、選択したパターン、データ型、データドメイン、または値に対するドリルダウンデータを表示できます。

[データプレビュー] ペインは詳細ビューで表示することが可能です。サマリビューでカラムをクリックすると、詳細ビューが表示され、[データプレビュー] ペインは折りたたまれます（デフォルト設定）。カラムデータを表示するには、[アクション] > [プレビューの表示] をクリックします。

以下の表に、[データプレビュー] ペインの [アクション] メニューに含まれるオプションを示します。

オプション	説明
フィルタに追加	プロファイル結果のサブセットでデータの不整合を分析できるように、ドリルダウンデータをフィルタするためのドリルダウンフィルタを作成します。
フィルタの保存	ドリルダウンフィルタを保存します。
プレビューの表示	ソース行を表示します。
データのエクスポート	ドリルダウン結果を CSV ファイルまたは Microsoft Excel ファイルにエクスポートします。

## データ型

データ型には、プロファイル結果の各カラムの推測されたデータ型がすべて含まれています。

データ型は、サマリビューと詳細ビューで確認できます。サマリビューでは、文書化されたデータ型と推測されたデータ型を表示できます。[競合しているデータ型] フィルタは、文書化されたデータ型と推測されたデータ型が一致しないカラムを表示します。詳細ビューでは、カラムの推測されたデータ型を表示できます。カラムのデータ型の頻度は、水平棒グラフとパーセンテージとして表示されます。選択した推測データ型をドリルダウン、承認、拒否、またはリセットできます。[拒否済みを表示] オプションには、拒否された推測データ型が表示されます。

次の表に、データ型のプロパティを示します。

プロパティ	説明
データ型	プロファイルのカラムの文書化されたデータ型と推測されたデータ型の一覧を表示します。
頻度	カラムに特定のデータ型が出現する回数（数値）を表示します。
割合	カラムに特定のデータ型が出現するパーセンテージを表示します。



プロパティ	説明
ドリルダウン	<p>カラムのデータ型に基づいて特定のソース行にドリルダウンします。</p> <p><b>注:</b> 推測されたデータ型を複数選択した場合、ドリルダウンアクションは実行できません。</p>
ステータス	<p>データ型のステータスを示します。ステータスは推測済み、承認済み、または拒否済みです。</p> <p><b>推測済み</b></p> <p>カラムのデータ型が、Analyst ツールによって推測されたものであることを示します。</p> <p><b>承認済み</b></p> <p>カラムのデータ型が承認済みであることを示します。データ型を承認すると、そのデータ型がモデルリポジトリにコミットされます。</p> <p><b>拒否済み</b></p> <p>カラムのデータ型が拒否されたことを示します。</p>

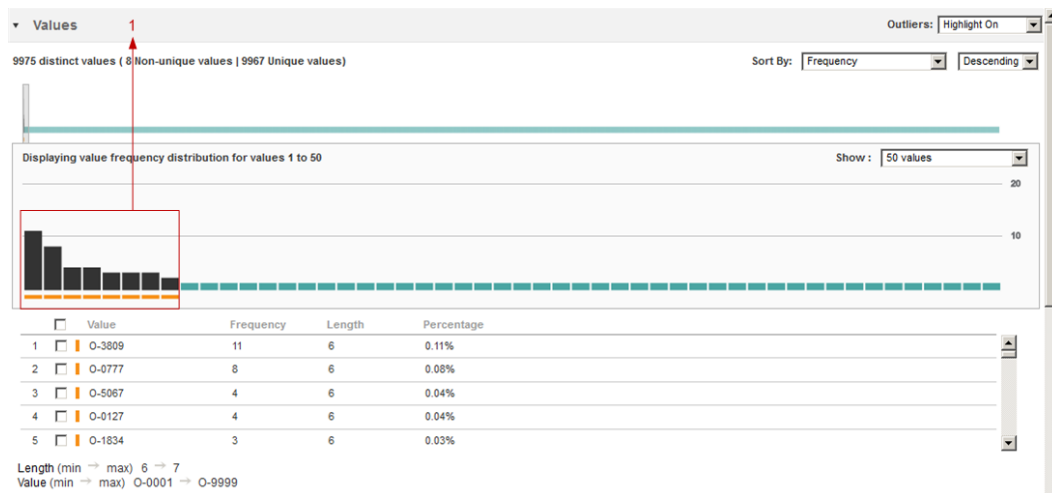
## 異常値

異常値とは、プロファイル結果の特定のカラムで、期待される値の範囲に収まらないパターン、値、頻度のことです。

データ統合サービスのプロファイリングプラグインは、カラムの大多数の値の範囲内に収まらない値を識別するアルゴリズムを実行します。特定のカラムに期待される大多数の値の範囲内に収まらないすべてのパターン、値、または頻度は異常値となります。

デフォルトでは、Analyst ツールはプロファイル結果の異常値を特定しません。サマリビューで、異常値を実行し、異常値結果を表示できます。パターンの異常値フィルタを使用すると、カラムのパターンに基づいて異常値が表示されます。値頻度の異常値フィルタを使用すると、カラムの値または頻度に基づいて異常値が表示されます。異常値の検出はバックグラウンドで実行されるため、サマリビューで他のアクションを実行できます。

詳細ビューで、リストから **【ハイライトをオン】** オプションを選択すると、[値] ペインに異常値を表示できます。異常値は、オレンジ色の下線が付いた垂直バーとして表示されます。異常値のみを表示するには、リストから **【フィルタ】** オプションを選択する必要があります。



1. 異常値。異常値は、オレンジ色の下線が付いた垂直バーとして表示されます。

## 異常値の実行

異常値を実行すると、カラムの期待される値の範囲に収まらないパターン、値、または頻度を特定できます。

1. サマリビューで、**【アクション】** > **【異常値の検出】** をクリックします。  
**【フィルタ基準】** ペイン内のパターン異常値と値頻度の異常値が「N/A」から、検出された異常値の数に変わります。
2. **【フィルタ基準】** ペインで、**【パターン異常値】** をクリックします。  
サマリビューに、パターン異常値が出現するカラムが表示されます。
3. **【フィルタ基準】** ペインで、**【値頻度の異常値】** をクリックします。  
サマリビューに、値または頻度の異常値が出現するカラムが表示されます。
4. 詳細ビューで、**【異常値】** ドロップダウンリストから **【ハイライトをオン】** を選択します。  
**【値】** ペインで、異常値は、オレンジ色の下線が付いた垂直バーとして表示されます。
5. **【異常値】** ドロップダウンリストで **【フィルタ】** をクリックすると、異常値のみが表示されます。

## パターン

カラム値のパターンおよびパターンが出現する頻度を、サマリビューおよび詳細ビューで確認できます。

サマリビューでは、カラムの複数のパターンを水平棒グラフで表示できます。マウスポインタを棒グラフの上に移動すると、パターンの特性やカラムに出現する類似パターンの数をパーセンテージとして確認できます。詳細ビューでは、カラムにパターンが出現する頻度を、水平棒グラフとパーセンテージで表示できます。パターンをドリルダウンして参照テーブルにパターンを追加したり、選択したパターンを持つデータドメインを作成できます。

プロファイリングウェアハウスには、一意の最高頻度値が 16,000 個まで格納されます。これには、プロファイル結果の NULL 値もデフォルトで含まれます。プロファイル結果に NULL 値が 1 つ以上存在する場合、Analyst ツールでは NULL 値をパターンとして表示することができます。

**注:** Analyst ツールでは、精度が 38 より大きい数値カラムのパターンを導出することはできません。また、精度が 255 より大きい文字列カラムのパターンを導出することはできません。

以下の表に、カラムパターンのプロパティを示します。

プロパティ	説明
パターン	プロファイルのカラムのパターンが表示されます。
頻度	カラムでパターンが出現する回数（数値）が表示されます。
割合	カラムに特定のパターンが出現するパーセンテージが表示されます。

以下の表に、パターン文字と、各パターン文字が何を表わすかを示します。

Character	説明
「B」または「b」 または「」	空白（ブランクスペース）を表します。
「C」または「c」	任意の文字を表します。
「L」または「l」	任意の小文字の英字を表します。
「T」または「t」	タブを表します。
「U」または「u」	任意の大文字の英字を表します。
9	任意の数字を表します。Informatica Analyst は、最高 3 文字を「9」形式で別々に表示します。4 文字以上の場合、括弧内に値として示されます。例えば、「9 (8)」というフォーマットは 8 桁の数値を表します。
「X」または「x」	任意の英字を表します。Informatica Analyst は、最高 3 文字を「X」形式で別々に表示します。4 文字以上の場合、括弧内に値として示されます。例えば、「X (6)」というフォーマットは、値「Boston」に相当する場合があります。 <b>注:</b> パターン文字 X は、大文字と小文字を区別しません。このため、ソースデータ内の大文字または小文字に相当する可能性があります。
「P」または「p」	「(」(開き括弧)を表します。
「Q」または「q」	「)」(閉じ括弧)を表します。

**注:** カラムパターンには特殊文字を含めることができます。例えば、~、[,、=、-、?、=、{、\*、-、>、<、\$などの特殊文字を使用できます。

## 値

カラムの値とその値がカラムに出現する頻度を表示できます。

サマリビューで、特定のカラムの最小値と最大値を確認できます。詳細ビューでは、カラムの値プロパティを確認できます。

### サマリビューの値

サマリビューでは、最新のプロファイル実行、履歴プロファイル実行、統合済みプロファイル実行について、すべてのカラムとルール の最小値と最大値を確認できます。

## 例

小売店データベースの Employee テーブルに Employee という名前のカラムがあり、100～250 の従業員 ID、および名前 (Bob や Robert など) が格納されています。Employee テーブルに対してカラムプロファイルを実行すると、サマリビューに従業員 ID の値カラムとして、100 --> Robert が表示されます。

## 詳細ビューの値

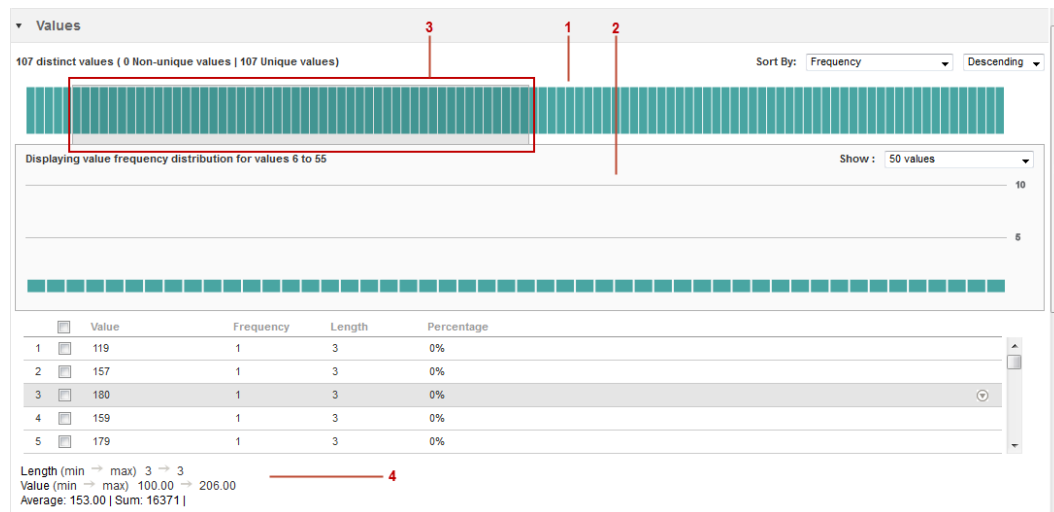
詳細ビューのカラム値には、カラムの値とその値がカラムに出現する頻度が表示されます。

**【値】** ペインに、カラム値が視覚的に表示されます。各値の頻度、長さ、パーセンテージを確認できます。値は、値または頻度に基づいてソートできます。データのドリルダウン、値の参照テーブルへの追加、値の頻度ルール作成、データドメインの作成を実行できます。NULL 値は赤い縦棒、値の頻度は黒い縦棒、異常値はオレンジ色に強調表示された縦棒として表示されます。異常値の強調表示、異常値の無効化、またはカラム内の異常値だけを表示するための結果のフィルタを行えます。

**【値】** ペインは、グラフィカルなレイアウトと値セクションで構成されます。

グラフィカルレイアウトは 2 つのパネルに分かれています。

次の画像は、詳細ビューの **【値】** ペインを示しています。



1. 上部パネル。値を垂直棒グラフとして表示できます。頻度と値を基準にして値をソートできます。昇順または降順で値をソートできます。異常値は、オレンジ色で強調表示された縦棒として表示できます。
2. 下部パネル。下部パネルのスライダに含まれる値を表示できます。ここでは、各値が縦棒で表されています。値のドリルダウン、値の参照テーブルへの追加、値の頻度ルール作成、値に対するデータドメインの作成を実行できます。一度に 50 個、75 個、または 100 個の値を表示するように選択できます。
3. スライダ。上部パネルの値の上でスライダをスライドさせることができます。下部パネルには、スライダ内の値が表示されます。
4. 値のプロパティ。値のプロパティセクションには、値とプロパティが表示されます。

次の表に、グラフィカルレイアウトの各パネルを示します。

パネル	説明
上部パネル	すべての値が垂直棒グラフで表示されます。上部パネルには、最大 16,000 個まで値を表示できます。スライダを使用して一連の値を表示することができます。
下部パネル	上部パネルで選択したバッチの各値が表示されます。デフォルトでは、Analyst ツールによって、50 個の値が表示されます。1 度に 75 個または 100 個の値を表示するよう選択できます。

次の表に、値セクションに表示されるカラム値のプロパティを示します。

プロパティ	説明
値	上部パネルで選択したバッチの値のリストが表示されます。 注: カラム値の CLOB、BLOB、Raw、およびバイナリの各データ型は、Analyst ツールによって除外されます。
頻度	カラムで値が出現する回数（数値）が表示されます。
長さ	カラム値の長さが表示されます。
割合	カラムに特定の値が出現するパーセンテージが表示されます。

次の表で、選択したカラムの統計について説明します。

統計	説明
長さ（最小～最大）	カラムの最短値と最長値の長さが表示されます。
値（最小～最大）	カラムの最小値と最大値が表示されます。
平均	カラムの値の平均が表示されます。
合計	カラム内のすべての値の合計が表示されます。

## 詳細ビューの値によるプロファイル結果の比較

詳細ビューの「値」ペインには、プロファイル結果を比較できるように、値のプロパティとして、重複しない値の数、値の最小値、最大値、最小長および最大長、平均、標準偏差、および合計が表示されます。

カラムの詳細ビューには、プロファイル結果を比較できるように、値のプロパティ、値、値の頻度と水平棒グラフが表示されます。

次の表に、2 つのプロファイル実行の結果を比較するとき、詳細ビューに表示されるカラム値のプロパティを示します。:

プロパティ	説明
重複しない値の数	カラムの重複しない値の数が表示されます。
最小値	カラムの最小値が表示されます。
最大値	カラムの最大値が表示されます。

プロパティ	説明
長さ（最小～最大）	カラムの最短値と最長値の長さが表示されます。
平均	カラムの値の平均が表示されます。
標準偏差	カラムのすべての値について、カラム値の標準偏差（カラム値のばらつき）が表示されます。
合計	カラム内のすべての値の合計が表示されます。

## プロファイル実行のタイプ

最新のプロファイル実行、履歴プロファイル実行、統合済みプロファイル実行のプロファイル結果を確認できます。プロファイル実行結果は、サマリビューで確認できます。

### 最新のプロファイル実行

サマリビューで、特定のプロファイルの最新のプロファイル実行の結果を確認できます。

次の操作を行うと、最新のプロファイル実行のプロファイル結果をサマリビューに表示できます。

- プロファイルを作成、保存して、実行する。
- 以前実行したプロファイルを【ライブラリ】ワークスペースから開く。
- 統合済みプロファイル実行のサマリビューまたは詳細ビューで【**最後のプロファイルの実行**に戻る】リンクをクリックする。
- 履歴プロファイル実行のサマリビューまたは詳細ビューで【**最後のプロファイルの実行**に戻る】リンクをクリックする。
- 【**プロファイル実行の選択**】ダイアログボックスで最新のプロファイル実行を選択し、【OK】をクリックする。

### 履歴プロファイル実行

サマリビューで、以前のプロファイル実行の結果を確認できます。

プロファイリングウェアハウスには、プロファイルのすべてのプロファイル実行の結果が保存されています。以前実行したプロファイル実行の結果を表示するには、【プロファイル実行の選択】ダイアログボックスでプロファイル実行を選択します。

### 統合済みプロファイル実行

サマリビューで、プロファイルの各カラムについて最新のプロファイル結果を確認できます。

統合済みプロファイル実行では、プロファイルの各カラムについて最新の結果を確認できます。【**プロファイル実行の選択**】ダイアログボックスで統合済みプロファイル実行を選択すると、プロファイリングウェアハウスによって特定のプロファイルのすべてのプロファイル実行から最新のカラム結果が取得されます。サマリビューで結果を確認できます。サマリビューのヘッダーには差分プロファイル実行が表示されます。

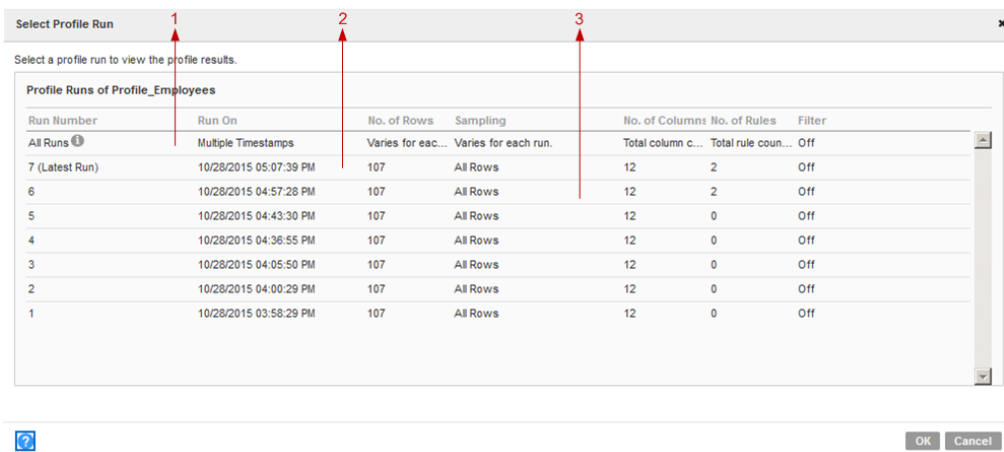
## 例

データアナリストは、プロファイルの各カラムの最新の結果を確認できます。例えば、プロファイル実行 A にはカラム 1、2、および 3 を選択し、プロファイル実行 B にはカラム 3、4、および 5 を選択します。このとき、すべてのカラムの最新の結果を表示するには、[プロファイル実行の選択] ダイアログボックスで統合済みプロファイル実行を選択します。サマリビューには、実行 A のカラム 1 と 2 の結果、および実行 B のカラム 3、4、5 の結果が表示されます。

## プロファイル実行の選択

プロファイル結果を表示する際には、履歴プロファイル実行、最新のプロファイル実行、または統合済みのプロファイル実行を選択できます。プロファイル結果はサマリビューで、カラム結果は詳細ビューで確認できます。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトまたはフォルダを選択するか、**【アセット】** ペインでプロファイルを選択します。
2. **【アクション】** > **【開く】** をクリックしてプロファイルを開きます。  
サマリビューが **【検出】** ワークスペースに表示されます。
3. サマリビューで、**【アクション】** > **【プロファイル実行の選択】** をクリックします。  
**【プロファイル実行の選択】** ダイアログボックスが表示されます。  
次の画像は、**【プロファイル実行の選択】** ダイアログボックスを示しています。



1. 統合済みプロファイル実行このプロファイル実行を選択すると、サマリビューで、各カラムの最新のプロファイル結果を確認できます。
2. 最新のプロファイル実行このプロファイル実行を選択すると、サマリビューで、特定のプロファイルの最新のプロファイル結果を確認できます。
3. 履歴プロファイル実行このプロファイル実行を選択すると、サマリビューで、以前のプロファイルの履歴プロファイル結果を確認できます。
4. **【プロファイル実行の選択】** ダイアログボックスで、いずれかのプロファイル実行を選択してそのプロファイル結果を表示します。
  - 最新のプロファイル実行のプロファイル結果を表示するには、最新のプロファイル実行を選択して、**【OK】** をクリックします。
  - 履歴プロファイル実行のプロファイル結果を表示するには、最新以外のプロファイル実行を選択して、**【OK】** をクリックします。
  - 統合済みプロファイル実行のプロファイル結果を表示するには、**【すべての実行】** を選択し、**【OK】** をクリックします。サマリビューに、各カラムの最新のプロファイル結果が表示されます。

Analyst ツールによってプロファイルが実行され、サマリビューにプロファイル結果が表示されます。

5. サマリビューでカラムをクリックすると、カラム結果が表示されます。  
詳細ビューが表示されます。

## 複数のプロファイル結果の比較の概要

2つのプロファイル実行のプロファイル結果を比較できます。サマリビューで比較結果を、詳細ビューでカラム結果を確認できます。

サマリビューでは、両方のプロファイル実行のすべてのカラムの比較結果を確認できます。

### 複数のプロファイル結果の比較

2つのプロファイル実行を比較した場合は、サマリビューで、プロファイル結果の比較を表示できます。

1. サマリビューで、**[アクション]** > **[プロファイル実行の比較]** をクリックします。

次の画像は、**[プロファイル実行の比較]** ダイアログボックスを示しています。

Compare Profile Runs

Select the profile runs that you want to compare.

Select a profile run

Run Number	Run On	No. of Rows	Pr. Sampling	No. of Columns	No. of Rules	Pr. Filter
3 (Latest Run)	09/09/2015 02:47:10 PM	27	No Sampling	4	0	Off
2	09/09/2015 02:45:24 PM	27	No Sampling	4	0	Off
1	09/07/2015 01:45:48 PM	27	No Sampling	4	0	Off

Select another profile run

Run Number	Run On	No. of Rows	Pr. Sampling	No. of Columns	No. of Rules	Pr. Filter
3 (Latest Run)	09/09/2015 02:47:10 PM	27	No Sampling	4	0	Off
2	09/09/2015 02:45:24 PM	27	No Sampling	4	0	Off
1	09/07/2015 01:45:48 PM	27	No Sampling	4	0	Off

OK Cancel

1. 実行 A。プロファイル実行を実行 A として選択します。
2. 実行 B。プロファイル実行を実行 B として選択します。  
**[プロファイル実行の比較]** ダイアログボックスが表示されます。
2. **[実行 A]** ペインでプロファイルを選択し、**[実行 B]** ペインで別のプロファイルを選択します。
3. **[OK]** をクリックします。  
サマリビューにプロファイル結果の統合ビューが表示されます。



## プロファイル結果の比較のサマリビュー

2つのプロファイル実行の比較時に、サマリビューに結果をグリッド形式で表示できます。サマリビューでデフォルトのフィルタを使用して、特定の統計を表示できます。

次の画像は、サマリビューに表示された、2つのプロファイル実行のプロファイルの比較結果を示しています。

Filter By:	All Columns & rules (4)
All Columns & rules	4
All Columns	4
Common Columns	3
Added Columns	0
Removed Columns	1
All Rules	0
Common Rules	0
Added Rules	0
Removed Rules	0

No	Name	Null	Unique	Non Unique	Pattern	Data Type	Data Domain
1	DEPARTMENT_ID	—	—	—	—	—	—
2	DEPARTMENT_NAME	—	—	—	—	—	▲ 1
3	LOCATION_ID	—	—	—	—	—	—
4	MANAGER_ID (Removed)	—	—	—	—	—	—

1. デフォルトフィルタ。サマリビューでは、デフォルトフィルタを使用してプロファイル比較結果を表示できます。
2. プロファイルヘッダー。ヘッダーでプロファイル名を表示できます。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。プロファイル実行の比較、プロファイル実行間の行の増減、プロファイルの行数、およびプロファイル実行の日時を表示できます。
4. サマリビュー。2つのプロファイル実行のカラム間の比較を表示できます。

## プロファイル結果の比較のためのサマリビューのプロパティ

プロファイル結果の比較のためのサマリビューのプロパティとして、重複なしの値、重複ありの値、NULL 値、パターン、推測されたデータ型、推測されたデータドメイン、リンク済みビジネス用語のそれぞれの数とパーセンテージがあります。サマリビューには、プロパティを視覚的に表したものが表示されます。各サマリのプロパティをクリックすると、そのプロパティの値に基づいてソートできます。

サマリビューでは、データ統合サービスによって、すべてのカラムおよびルールに昇順で数値が割り当てられます。

**注:** 数値カウントに付いている上矢印は、2つのプロファイル実行を比較したとき一方のプロパティの値が増加したことを示します。数値カウントに付いている下矢印は、2つのプロファイル実行を比較したとき一方のプロパティの値が減少したことを示します。

次の表に、プロファイル結果の比較のためのサマリプロパティを示します。

プロパティ	説明
×	カラムまたはルールが表示されます。
名前	プロファイルのカラムまたはルールの名前が表示されます。

プロパティ	説明
Null	NULL 値の増減が表示されます。
重複なし	重複なしの値の増減が表示されます。
重複あり	重複ありの値の増減が表示されます。
パターン	複数のプロファイル実行間のパターンのバリエーションが表示されます。
データ型	2つのプロファイル実行のカラムまたはルールについて、推測されたデータ型のバリエーションが表示されます。
データドメイン	2つのプロファイル実行のカラムまたはルールに関連付けられた推測データドメインのバリエーションが表示されます。

## サマリビューのプロファイル結果の比較で使用するデフォルトフィルタ

サマリビューでは、デフォルトフィルタを使用してプロファイル結果を表示できます。

サマリビューでソースカラムと仮想カラムを表示できます。ルールの出力は、サマリビューに仮想カラムとして表示されます。ルールの出力ポートを変更し、プロファイル実行と履歴実行を比較すると、履歴ルールの出力カラムは【**削除されたルール**】フィルタに表示され、新規ルールの出力カラムは【**追加されたルール**】フィルタに表示されます。1つの出力ルールのルールロジックを変更した場合、またはプロファイル実行で複数のルール出力の入力を変更して履歴実行と比較した場合、【**追加されたルール**】および【**削除されたルール**】のフィルタ出力は変更されません。フィルタはカラムの名前変更のみをフィルタに対する有効な入力とみなすため、フィルタ出力は変更されません。

次のデフォルトのフィルタオプションを使用し、特定の条件を満たすプロファイル結果を表示できます。

デフォルトフィルタオプション	説明
すべてのカラムとルール	ソースカラム、仮想カラム、およびルールカラムのプロファイル結果を表示します。ソースカラムとルールカラムを展開および縮小して結果を表示できます。
すべてのカラム	ソースカラムと仮想カラムのプロファイル結果を表示します。
共通カラム	両方のプロファイル実行結果で利用できるカラムを表示します。
追加されたカラム	最新のプロファイル実行で使用可能なカラムを表示します。例えば、実行5と実行3を比較する場合、【 <b>追加されたカラム</b> 】には、実行3ではなく実行5で使用可能なカラムが表示されます。
削除されたカラム	履歴プロファイル実行で使用可能なカラムを表示します。例えば、実行5と実行3を比較する場合、【 <b>削除されたカラム</b> 】には、実行5ではなく実行3で使用可能なカラムが表示されます。
すべてのルール	すべてのルールカラムのプロファイル結果を表示します。

デフォルトフィルタオプション	説明
追加されたルール	最新のプロファイル実行で使用可能なルールを表示します。例えば、実行 5 と実行 3 を比較する場合、[追加されたルール] には、実行 3 ではなく実行 5 で使用可能なルールが表示されます。
削除されたルール	履歴プロファイル実行で使用できるルールを表示します。例えば、実行 5 と実行 3 を比較する場合、[削除されたルール] には、実行 5 ではなく実行 3 で使用可能なルールが表示されます。

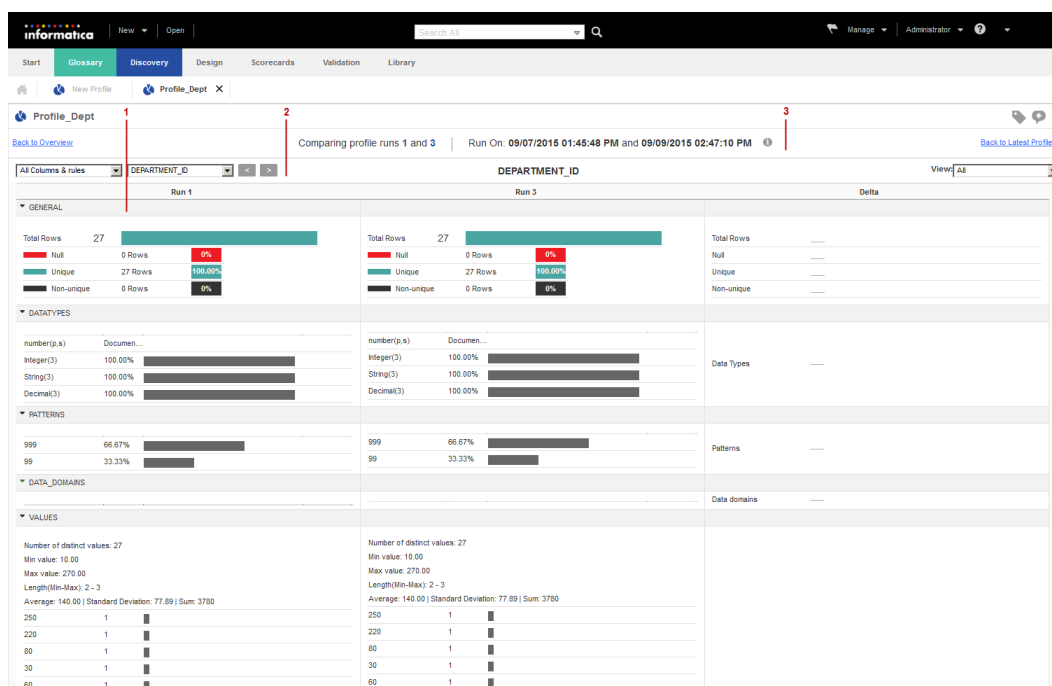
サマリビューには、すべてのソースカラムおよび仮想カラムのプロファイル結果がデフォルトで表示されます。

## プロファイル結果の比較の詳細ビュー

カラム結果は、詳細ビューにグリッド形式で表示されます。カラムの詳細には、重複なしの値、重複ありの値、NULL 値、パターン、データ型、データドメイン、ビジネス用語、値、データプレビューなどの全般的な情報が含まれます。

カラムの詳細ビューは、カラム名をクリックすると表示されます。実行 A と実行 B のカラム結果は別のカラムとして表示できます。データの比較は差分カラムに表示されます。

次の画像は、詳細ビューに表示された特定のカラムのプロファイル結果の比較を示しています。



1. ペイン。2つのプロファイル実行のカラムのプロファイル結果と統計、および2つのプロファイル実行のカラムの差分情報を各ペインに表示できます。
2. プロファイルヘッダー。ドロップダウンリストでカラムを選択するか、ナビゲーションボタンを使用して、カラム結果を表示できます。カラム名を表示できます。また、[表示] ドロップダウンリストのオプションを使用して特定の結果を表示することもできます。
3. サマリビューのヘッダー。サマリビューのヘッダーにはプロファイル固有の情報を表示できます。比較しているプロファイル実行と、プロファイル実行の日時を表示できます。

## プロファイル結果の比較のための [詳細ビュー] ペイン

詳細ビューには、2つのプロファイル実行について、特定のカラムの詳細なプロファイル結果と比較結果が表示されます。

詳細ビューでは、実行 A と実行 B のカラム結果が表示され、差分カラムにデータの比較が表示されます。他のカラム結果を表示するには、[フィルタ] ドロップダウンリストからフィルタを選択するか、[カラム] ドロップダウンリストからカラムを選択します。

## カラムプロファイルのドリルダウン

カラムプロファイルのドリルダウンオプションを使用すると、カラム値に基づいてデータソース内の特定の行にドリルダウンできます。ドリルダウンの対象として、データソース内の現在のデータを読み取るか、プロファイルウェアハウスにステージングされたプロファイルデータを読み取るかを選択できます。ステージングされたプロファイルデータの特定の行にドリルダウンする場合は、Analyst ツールによって、一致するカラムの値に絞り込むためのドリルダウンフィルタが作成されます。ドリルダウンの実行後、ドリルダウンフィルタの編集、再呼び出し、リセット、および保存ができます。

プロファイリング対象として選択していないカラムも、ドリルダウン対象のカラムとして選択できます。ドリルダウンの対象として、データソース内の現在のデータを読み取るか、プロファイリングウェアハウスにステージングされたプロファイルデータを読み取るかを選択できます。カラム値のドリルダウンを実行した後に、選択した値またはパターンのドリルダウンデータを、任意の場所の CSV ファイルにエクスポートすることができます。Informatica Analyst にはドリルダウンデータの最初の 200 個しか値が表示されませんが、CSV ファイルにはすべての値がエクスポートされます。

## 行データのドリルダウン

プロファイルを実行した後に、カラム値、データ型、またはパターンに一致する特定の行にドリルダウンすることができます。

1. プロファイルを実行します。  
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラム名をクリックします。  
詳細ビューにカラム結果が表示されます。
3. 詳細ビューで [値] ペインの値を右クリックし、[ドリルダウン] を選択します。  
[データプレビュー] ペインにドリルダウンデータが表示されます。

## ドリルダウンデータへのフィルタの適用

プロファイル結果のサブセットのデータの不整合を分析できるように、ドリルダウンデータを繰り返しフィルタすることができます。

1. [値] タブでカラム値を選択します。
2. 右クリックして [ドリルダウン] を選択します。  
[データプレビュー] ペインにドリルダウン結果が表示されます。
3. フィルタ条件を追加するには、[データプレビュー] ペインでカラム値を右クリックして、[フィルタに追加] を選択します。  
フィルタ条件を示す [ドリルダウンフィルタ] ダイアログボックスが表示されます。

4. 必要なフィルタ条件を追加して、**[OK]** をクリックします。  
推測されるデータ型にドリルダウンフィルタを適用できません。
5. フィルタを保存するには、**[アクション]** > **[フィルタの保存]** をクリックします。
6. ドリルダウンフィルタをクリアするには、**[アクション]** > **[更新]** をクリックします。
7. ドリルダウンデータを Microsoft Excel スプレッドシートにエクスポートするには、**[アクション]** > **[データのエクスポート]** をクリックします。

## Analyst ツールでのキュレーション

キュレーションとは、データソースで検出されたメタデータを検証し管理することで、使用とレポートに適したメタデータにするプロセスです。Analyst ツールでメタデータをキュレーションする場合、プロファイル結果の中の推測されたデータ型またはデータドメインを承認、拒否、リセットすることができます。

1つのカラムに対して1つのデータ型およびデータドメインを承認できます。カラムの拒否されたデータ型またはデータドメインは、非表示にできます。推測されたデータ型またはデータドメインを承認または拒否した後も、データ型またはデータドメインをリセットすると、ステータスを推測済みに戻すことができます。

### データ型とデータドメインの承認

プロファイル結果には、データソースの各カラムについて、推測されたデータ型とデータドメインが含まれています。Analyst ツールで、各カラムについて、1つのデータ型と1つのデータドメインを選択および承認できます。

1. プロファイルを実行します。  
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラム名をクリックします。  
詳細ビューにカラム結果が表示されます。
3. 詳細ビューの **[データ型]** ペインでデータ型を選択するか、**[データドメイン]** ペインでデータドメインを選択します。
4. **[アクション]** > **[承認]** をクリックします。
5. データ型またはデータドメインのステータスを推測済みに戻すには、データ型またはデータドメインを選択して、**[アクション]** > **[リセット]** をクリックします。

### データ型とデータドメインの拒否

詳細ビューで、データ型またはデータドメインを拒否できます。拒否されたデータ型およびデータドメインの表示/非表示を切り替えることができます。

1. プロファイルを実行します。  
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラム名をクリックします。  
詳細ビューにカラム結果が表示されます。
3. 詳細ビューの **[データ型]** ペインでデータ型を選択するか、**[データドメイン]** ペインでデータドメインを選択します。

4. **【アクション】** > **【拒否】** をクリックします。  
Analyst ツールによって、拒否されたデータ型がデータ型リストから削除されます。
5. 拒否されたデータドメインを表示するには、**【アクション】** > **【拒否済みを表示】** をクリックします。

## Informatica Analyst のカラムプロファイルのエクスポートファイル

プロファイル結果の一部と、完全な結果サマリのどちらを選択するかに基づいて、カラムプロファイル結果を CSV ファイルまたは Microsoft Excel ファイルにエクスポートすることができます。

選択した値およびパターンについて、値の頻度、パターンの頻度、データ型、またはドリルダウンデータを CSV ファイルにエクスポートすることができます。すべてのカラムのプロファイリング結果のサマ리를 Microsoft Excel ファイルにエクスポートすることができます。データ統合サービスの特権「**結果のドリルダウンおよびエクスポート**」を使用して、プロファイル結果のエクスポートを誰に許可するかをユーザーまたはグループ単位で決定します。

### CSV ファイルのプロファイルエクスポート結果

値の頻度、パターンの頻度、データ型、ドリルダウンデータをエクスポートして、1 つのファイルでデータを確認することができます。Analyst ツールは CSV ファイル内の情報を保存します。

推測されるカラムパターンをエクスポートすると、Analyst ツールは異なる形式のカラムパターンをエクスポートします。例えば、推測されるカラムパターン X (5) をエクスポートすると、形式 XXXXX のカラムパターンが CSV ファイルに表示されます。

### Microsoft Excel のプロファイルエクスポート結果

完全なプロファイル結果のサマ리를エクスポートすると、Analyst ツールは Microsoft Excel ファイルの複数のワークシートに情報を保存します。Analyst ツールにより、ファイルは「xlsx」形式で保存されます。

次の表に、エクスポートファイルの各ワークシートに表示される情報を示します。

タブ	説明
カラムプロファイル	プロファイルの実行後にサマリビューからエクスポートされるサマリ情報。これには、カラム名、ルール名、重複なしの値の数、NULL 値の数、推測されたデータ型、プロファイルの最終実行日時などが含まれます。
値	カラムとルールの値、および各カラムの値が出現する頻度。
パターン	プロファイルを実行したカラムとルールの値のパターン、およびそのパターンが出現する頻度。
データ型	カラムに対するすべてのデータ型、各データ型の頻度、パーセンテージ値、および [推測済み]、[承認済み]、[拒否済み] などのデータ型のステータス。

タブ	説明
統計	それぞれのカラムとルールに関する統計。これには、平均、長さ、上位の値、下位の値、標準偏差などが含まれます。
プロパティ	【プロパティ】ビューの情報。これには、プロファイルの名前、タイプ、サンプリングポリシー、行数が含まれます。

## Informatica Analyst からのプロファイル結果のエクスポート

プロファイルの結果を「.csv」または「.xlsx」ファイルにエクスポートして、ファイルでデータを表示することができます。

1. **【ライブラリ】** ワークスペースで、プロファイルを含むプロジェクトまたはフォルダを選択します。
2. プロファイルをクリックして開きます。  
サマリビューにプロファイル結果が表示されます。
3. サマリビューで、**【アクション】 > 【データのエクスポート】** をクリックします。  
**【ファイルへのデータのエクスポート】** ダイアログボックスが表示されます。
4. **【ファイルへのデータのエクスポート】** ダイアログボックスで、ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
5. **【すべて（サマリ、値、パターン、統計、プロパティ）】** または **【データドメイン検出結果】** を選択し、**【コードページ】** を選択します。 **【OK】** をクリックします。  
データが Microsoft Excel スプレッドシートにエクスポートされます。
6. サマリビューでカラムをクリックします。  
詳細ビューにカラム結果が表示されます。
7. 詳細ビューで、**【アクション】 > 【データのエクスポート】** をクリックします。  
**【ファイルへのデータのエクスポート】** ダイアログボックスが表示されます。
8. **【ファイルへのデータのエクスポート】** ダイアログボックスで、ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
9. 次のいずれかのオプションを選択します。
  - **【すべて】（サマリ、値、パターン、統計、プロパティ）**
  - **【選択されたカラムの値の頻度】**
  - **【選択されたカラムのパターンの頻度】**
  - **選択されたカラムのデータ型**
  - **選択された値でデータをドリルダウン**
  - **選択されたパターンでデータをドリルダウン**
  - **選択したデータ型のデータをドリルダウン**
10. ファイル形式を入力します。形式は、**【すべて】** オプションの場合は **【Excel】**、その他のオプションの場合は **【CSV】** です。フィールド名をファイルの先頭行としてエクスポートするように選択できます。
11. ファイルのコードページを選択します。
12. **【OK】** をクリックします。  
データはファイルにエクスポートされます。

## 第 10 章

# Informatica Analyst のビジネス用語、コメント、タグ

この章では、以下の項目について説明します。

- [Informatica Analyst でのビジネス用語、コメント、タグの概要, 80 ページ](#)
- [ビジネス用語, 80 ページ](#)
- [コメント, 81 ページ](#)
- [タグ, 82 ページ](#)

## Informatica Analyst でのビジネス用語、コメント、タグの概要

プロファイルまたはプロファイルのカラムに、ビジネス用語、コメント、およびタグを追加できます。サマリビューおよび詳細ビューで、ビジネス用語、コメント、およびタグの割り当て、表示、および編集を実行できます。

### ビジネス用語

Analyst ツールで、プロファイルのカラムにビジネス用語を割り当てることができます。アセットリンクの編集やカラムのビジネス用語の削除を実行できます。ビジネス用語集とは、ビジネスユーザー向けに、ビジネス言語を使用して概念を定義した用語の集まりです。ビジネス用語は、概念のビジネス上の定義および使用方法を説明したものです。

サマリビューと詳細ビューで、カラム内のビジネス用語の割り当て、表示、または削除を実行できます。**【用語集】** ワークスペースのビジネス用語を表示するには、詳細ビューでビジネス用語をクリックします。

ビジネス用語のアセットリンクのプロパティを編集できます。ビジネス用語は、仮想カラムとしてプロファイル結果と共に CSV ファイルにエクスポートできます。



## ビジネス用語のカラムへの割り当て

サマリビューおよび詳細ビューで、ビジネス用語をプロファイルのカラムに割り当てます。カラムのビジネス用語を削除するには、**【ビジネス用語】** パネルを使用します。ビジネス用語のアセットリンクのプロパティを編集するには、**【アセットリンクの編集】** ダイアログボックスを使用します。

1. サマリビューで、カラム名を右クリックし、**【ビジネス用語の管理】** を選択します。詳細ビューで、**【ビジネス用語】** タブの **【アクション】** メニューから **【ビジネス用語の管理】** を選択します。  
**【ビジネス用語】** パネルが表示されます。
2. プラス (+) アイコンをクリックするか、**【ビジネス用語の割り当て】** リンクをクリックしてビジネス用語を追加します。  
**【ビジネス用語の割り当て】** パネルが表示されます。
3. **【ビジネス用語の割り当て】** パネルに表示されたビジネス用語のリストからビジネス用語を選択します。アセット名を入力します。必要に応じて、アセットのコンテキストと説明を追加します。**【OK】** をクリックします。  
ビジネス用語が、**【ビジネス用語】** パネルに表示されます。

## コメント

プロファイルまたはプロファイルのカラムにコメントを追加することで、さらに綿密な関係と分析を行うための追加の情報を提供できます。

プロファイルレベルでは、プロファイル、プロファイル定義、またはプロファイルメタデータに関するコメントを追加できます。サマリビューでプロファイルのコメントを表示できます。

サマリビューおよび詳細ビューで、カラムのコメントを追加および表示できます。

コメントを使用して、次のタスクを実行できます。

- コメントを仮想カラムとしてプロファイル結果と共に CSV ファイルにエクスポートします。CSV ファイルには、プロファイルおよびプロファイル内カラムのすべてのコメントが出力されます。
- コメントカラムのキーワードを使用してプロファイル結果を検索します。
- プロファイルのソースカラムおよび仮想カラムの両方にコメントを追加します。

**注:** カラムを選択しない場合、またはカラムコメントを追加しない場合は、サマリビューの **【コメント】** パネルにプロファイルコメントが表示されます。

## プロファイルまたはカラムへのコメントの追加

コメントを追加または表示するには、**【コメント】** パネルを使用します。

1. サマリビューまたは詳細ビューでコメントを追加できます。
  - 詳細ビューでプロファイルコメントを追加するには、**【アクション】** > **【コメントの表示】** をクリックします。
  - サマリビューでカラムコメントを追加するには、カラムを右クリックして **【コメントの表示】** を選択します。
  - 詳細ビューで、**【全般】** ペインの **【コメントの追加】** をクリックします。  
**【コメント】** パネルが表示されます。

2. **【コメントの追加】** をクリックします。  
**【コメント】** パネルにテキストボックスが表示されます。
3. 理解しやすいコメントテキストを追加して、**【保存】** をクリックします。  
**【コメント】** パネルに、コメントと現在のユーザー名および作成日時が表示されます。

## タグ

プロファイルまたはプロファイルのカラムにタグを割り当てることによって、ビジネス用途に応じてオブジェクトをグループ化できます。

サマリビューでは、プロファイルタグの表示または割り当てができます。サマリビューおよび詳細ビューで、カラムタグの表示または割り当てができます。

カラムタグに対して次のタスクを実行できます。

- タグを仮想カラムとしてプロファイル結果と共に CSV ファイルにエクスポートします。CSV ファイルには、プロファイルおよびプロファイル内カラムのすべてのタグが出力されます。
- プロファイルのソースカラムおよび仮想カラムにタグを割り当てます。

**注:** カラムを選択しない場合、またはカラムタグを追加しない場合は、サマリビューの **【タグ】** パネルにプロファイルタグが表示されます。

## プロファイルまたはカラムへのタグの追加

サマリビューでプロファイルにタグを追加します。サマリビューまたは詳細ビューでカラムにタグを追加します。

1. サマリビューまたは詳細ビューでタグを追加できます。
  - サマリビューでプロファイルにタグを割り当てるには、**【アクション】** > **【タグの表示】** をクリックします。
  - サマリビューでカラムにタグを追加するには、カラムを右クリックして **【タグの表示】** をクリックします。
  - 詳細ビューで、**【全般】** ペインの **【タグの追加】** をクリックします。  
**【タグ】** パネルが表示されます。
2. プラス (+) アイコンをクリックするか、**【タグの割り当て】** リンクをクリックしてタグを割り当てます。  
**【タグの割り当て】** ダイアログボックスが表示されます。
3. プロファイルまたはカラムに割り当てる 1 つ以上のタグを選択します。**【OK】** をクリックして **【タグ】** パネルを開きます。  
**注:** タグを作成するには、**【タグの割り当て】** パネルで **【新しいタグの追加】** をクリックします。

## 第 11 章

# Informatica Analyst のスコアカード

この章では、以下の項目について説明します。

- [Informatica Analyst のスコアカードの概要, 83 ページ](#)
- [Informatica Analyst のスコアカードのプロセス, 84 ページ](#)
- [Informatica Analyst でのスコアカードの作成, 85 ページ](#)
- [既存のスコアカードへのカラムの追加, 86 ページ](#)
- [既存のスコアカードへカラムの追加, 86 ページ](#)
- [スコアカードの実行, 87 ページ](#)
- [スコアカードの表示, 88 ページ](#)
- [スコアカードの編集, 88 ページ](#)
- [メトリック, 89 ページ](#)
- [\[メトリックグループ\], 90 ページ](#)
- [カラムのドリルダウン, 92 ページ](#)
- [傾向グラフ, 93 ページ](#)
- [Informatica Analyst のスコアカードダッシュボード, 96 ページ](#)
- [Informatica Analyst のスコアカードエクスポートファイル, 102 ページ](#)
- [スコアカード通知, 103 ページ](#)
- [スコアカードリネージュ, 106 ページ](#)

## Informatica Analyst のスコアカードの概要

スコアカードは、プロファイル内のカラムの有効値を視覚的に表したものです。スコアカードを作成して、実データまたはステージングされたデータをドリルダウンすることができます。

データ品質の進捗を測定するには、スコアカードを使用します。例えば、データ品質ルールを適用する前に、スコアカードを作成してデータ品質を測定することができます。データ品質ルールを適用した後に、別のスコアカードを作成して、データ品質に対するルールの影響を比較することができます。

スコアカードには、カラムの値の頻度がスコアとして表示されます。スコアにはカラムの有効値の割合が反映されます。プロファイルの実行後に、プロファイルからスコアカードにカラムをメトリックとして追加できます。関連し合うメトリックをグループ化する単一のエンティティとして、メトリックグループを作成できます。レコード内のカラムで許容できる不良データの範囲を指定するしきい値を定義し、メトリックごとにメトリッ

クウェイトを割り当てます。スコアカードを実行すると、Analyst ツールは、各メトリックグループの加重平均を生成します。データ品質を詳細に評価するために、各メトリックに固定費または変動費を割り当てることもできます。スコアカードを実行すると、Analyst ツールでメトリックごとに不良データのコストの合計を計算して合計コストを表示します。

スコアカードを作成または編集すると、ソースデータに基づくスコアカードフィルタを作成できます。スコアカードフィルタにより、フィルタ条件に基づくメトリックスコアを再計算できます。有効なデータレコードと無効なレコードを識別するには、各メトリックをドリルダウンします。傾向グラフを使用して、メトリックスコアおよびメトリックの不良データのコストが、特定の期間にわたって変化する様子を追跡できます。スコアカードでプロファイルフィルタを再利用できます。

Analyst ツールでバージョン管理システムが有効になっていると、複数のバージョンのスコアカードを作成し、スコアカードのバージョン履歴を表示できます。デフォルトでは、スコアカードは作成後にチェックアウトされます。他のユーザーがスコアカードを編集できるようにするには、スコアカードをチェックインする必要があります。

**【スコアカード】** ワークスペースでスコアカードダッシュボードを表示できます。スコアカードダッシュボードでは、スコアカードのあるデータオブジェクト、プロジェクトのスコアカード、過去 6 か月間のスコアカード実行傾向、1 か月間に実行されたすべてのスコアカードに対する、正常、許容可能、および許容不可のメトリックの集計を表示できます。

Informatica Analyst で、スコアカードの電子メール通知を設定および管理できます。電子メールサービスを使用して、電子メール通知を管理します。電子メールサービスは、Informatica Administrator で設定できるシステムサービスです。

## Informatica Analyst のスコアカードのプロセス

Developer tool および Analyst ツールでは、スコアカードを作成、編集できます。Analyst ツールではスコアカードを実行できます。スコアカードは、データオブジェクトの現在のデータに対して実行するか、プロファイルリングウェアハウスにステージングされているデータに対して実行することができます。

**【スコアカード】** ワークスペースでスコアカードを表示できます。スコアカードを実行したら、**【スコアカード】** パネルでスコアを表示できます。スコアカード内のスコアからデータオブジェクトを選択し、そのデータオブジェクトに移動できます。データオブジェクトは別のタブで開きます。

スコアカードに関する以下の作業を実行できます。

1. Developer tool または Analyst ツールでスコアカードを作成し、プロファイルからカラムを追加します。
2. Analyst ツールでスコアカードを開きます。
3. プロファイルの実行後に、プロファイルカラムをメトリックとしてスコアカードに追加します。
4. 必要に応じて、ソースデータに基づくスコアカードフィルタを作成します。
5. オプションで、各メトリックの無効なデータのコストを設定します。
6. スコアカードを実行してカラムのスコアを生成します。
7. スコアカードを表示して、レコード内の各カラムのスコアを確認します。
8. スコアのカラムをドリルダウンします。
9. スコアカードを編集します。
10. スコアカードの各メトリックに対してしきい値を設定します。
11. グループを作成して、スコアカードにおいて関連し合うメトリックを追加または移動します。
12. 必要に応じて、グループを編集または削除します。
13. 各スコアのスコア傾向グラフを表示して、スコアの変化を時系列で監視します。

14. 必要に応じて、各メトリックのコストの傾向グラフを表示して、データ品質の値を監視します。
15. メトリックまたはメトリックグループごとにスコアカードリネージュを表示します。
16. 読み取りアクセス権限のあるスコアカードに関する統合情報を表示します。

## Informatica Analyst でのスコアカードの作成

スコアカードを作成し、プロファイルからスコアカードにカラムを追加します。スコアカードにカラムを追加する前に、プロファイルを実行する必要があります。

1. **【ライブラリ】** ワークスペースで、プロファイルを含むプロジェクトまたはフォルダを選択します。
2. プロファイルをクリックして開きます。  
プロファイル結果が **【検出】** ワークスペースのサマリビューに表示されます。
3. **【アクション】** > **【スコアカードに追加】** をクリックします。  
**【スコアカードに追加】** ウィザードが表示されます。
4. **【スコアカードに追加】** 画面で、新しいスコアカードを作成するように選択したり、既存のスコアカードを編集してカラムを定義済みのスコアカードに追加したりできます。デフォルトでは、**【新しいスコアカード】** オプションが選択されています。**【次へ】** をクリックします。
5. **【手順 2/8】** 画面で、スコアカードの名前を入力します。必要に応じて、スコアカードの説明を入力できます。スコアカードを保存するプロジェクトおよびフォルダを選択します。**【次へ】** をクリックします。  
デフォルトでは、プロファイルで定義されているカラムとルールが選択されます。プロファイルに含まれていないカラムは追加できません。
6. **【手順 3/8】** 画面で、スコアカードにメトリックとして追加するカラムとルールを選択します。必要に応じて、左側のカラムヘッダのチェックボックスをクリックしてすべてのカラムを選択します。必要に応じて、**【カラム名】** を選択してカラム名をソートします。**【次へ】** をクリックします。
7. **【手順 4/8】** 画面で、メトリックにフィルタを追加できます。  
プロファイルに対して作成したフィルタをメトリックに適用したり、新しいフィルタを作成したりできます。**【メトリックフィルタ】** ペインでメトリックを選択し、**【フィルタの管理】** アイコンをクリックして、**【フィルタの編集: カラム名ダイアログボックス】**を開きます。**【フィルタの編集: カラム名ダイアログボックス】**で、以下のいずれかのタスクを実行するように選択できます。
  - プロファイル用に作成したフィルタを選択します。**【次へ】** をクリックします。
  - 既存のフィルタを選択します。**【編集】** アイコンをクリックし、**【フィルタの編集】** ダイアログボックスでフィルタを編集します。**【次へ】** をクリックします。
  - プラス (+) アイコンをクリックし、**【新規フィルタ】** ダイアログボックスでフィルタを作成します。**【次へ】** をクリックします。必要に応じて、選択したフィルタをスコアカードのすべてのメトリックに適用することもできます。  
フィルタが **【メトリックフィルタ】** ペインに表示されます。
8. **【手順 4/8】** 画面で **【次へ】** をクリックします。
9. **【手順 5/8】** 画面で、**【メトリック】** ペインの各メトリックを選択し、次のタスクを実行します。
  - 有効な値を設定します。**【スコアで使用: 値】** ペインで、**【利用可能な値】** ペインの 1 つ以上の値を選択し、右矢印ボタンをクリックしてそれらの値を **【有効な値】** ペインに移動します。メトリックの有効な値の総数が **【利用可能な値】** ペインの上部に表示されます。

- メトリックのしきい値を設定します。**【メトリックのしきい値】** ペインで、**【正常】**、**【許容可能】**、および **【許容不可】** スコアに対してしきい値を設定します。
  - 無効なデータのコストを設定します。メトリックのコストに定数値を割り当てるには、**【固定費】** を選択します。数値カラムを変動費としてメトリックに添付するには、**【変動費】** を選択し、**【カラムの選択】** をクリックして数値カラムを選択します。必要に応じて、**【コスト単位の変更】** をクリックしてコストの単位を変更します。メトリックに対して無効なデータのコストを設定しない場合は、**【なし】** を選択します。
10. **【次へ】** をクリックします。
  11. **【手順 6/8】** 画面で、メトリックを追加できるメトリックグループを選択したり、新しいメトリックグループを作成したりできます。新しいメトリックグループを作成するには、グループアイコンをクリックします。**【次へ】** をクリックします。
  12. **【手順 7/8】** 画面で、グループ内のメトリックのウェイトと、グループのしきい値を指定します。
  13. **【手順 8/8】** 画面で、スコアカードを実行するためのランタイム環境として **【ネイティブ】**、**【Hadoop】**、または **【Databricks】** を選択します。Hadoop ランタイム環境では **【Blaze】** または **【Spark】** エンジンを選択できます。**【Databricks】** を選択した場合は、**【参照】** をクリックして、プロファイルを実行する Databricks 接続を選択します。**【Blaze】** または **【Spark】** エンジンを選択した場合は、**【参照】** をクリックして、プロファイルを実行する Hadoop 接続を選択します。
  14. **【保存】** をクリックしてスコアカードを保存するか、**【保存して実行】** をクリックしてスコアカードを保存してから実行します。
- スコアカードが **【スコアカード】** ワークスペースに表示されます。

## 既存のスコアカードへのカラムの追加

プロファイルの実行後に、既存のスコアカードにプロファイル結果のカラムを追加できます。メトリックまたはメトリックグループを追加し、カラムに対する有効な値を設定し、各メトリックに対する無効なデータのコストを追加できます。**【すべての行】** 以外のサンプリングのオプションが設定されているプロファイルからカラムをスコアカードに追加する場合、スコアカードにプロファイル結果が反映されないことがあります。

既存のスコアカードにカラムを追加する場合、**【スコアカードに追加】** ウィザードでスコアカードの既存のメトリックまたはメトリックグループを編集することはできません。スコアカードの既存のメトリックを変更するには、**【スコアカード】** ワークスペースに移動して、スコアカードを編集し、必要に応じてメトリックまたはメトリックグループを更新します。

## 既存のスコアカードへカラムの追加

プロファイルの実行後に、既存のスコアカードにカラムを追加できます。

1. プロファイルをクリックして開きます。  
サマリビューにプロファイル結果が表示されます。
2. カラムを選択します。**【アクション】** > **【スコアカードに追加】** をクリックします。  
**【スコアカードに追加】** ウィザードが表示されます。

**注:** スコアカードにカラムを追加するときは、以下のルールとガイドラインに従います。

- カラム名とスコアカード名が同じ場合、スコアカードにカラムを追加することはできません。
  - カラム名を変更しても、同じカラムをスコアカードに2回追加することはできません。
3. **【既存のスコアカード】** を選択して定義済みのスコアカードにカラムを追加します。**【次へ】** をクリックします。
  4. **【手順 2/7】** 画面で、カラムを追加するスコアカードを選択します。**【次へ】** をクリックします。  
スコアカードに関連付けられている既存のメトリックとメトリックグループを確認できます。
  5. **【手順 3/7】** 画面で、スコアカードにメトリックとして追加するカラムとルールを選択します。必要に応じて、左側のカラムヘッダのチェックボックスをクリックしてすべてのカラムを選択します。**【カラム名】** をクリックして、カラム名をソートします。**【次へ】** をクリックします。
  6. **【手順 4/7】** 画面で、メトリック用のフィルタを作成できます。プロファイルに対して作成したフィルタをメトリックに適用することもできます。
  7. **【手順 5/7】** 画面で、次のタスクを実行できます。
    - **【メトリック】** ペインで、他のペインの各メトリックを選択し、メトリック値を設定します。
    - **【スコアで使用: 値】** ペインで、**【利用可能な値】** ペインの複数の値を選択し、右矢印ボタンをクリックして値を**【有効な値】** ペインに移動します。  
メトリックの有効な値の総数が**【利用可能な値】** ペインの上部に表示されます。
    - **【メトリックのしきい値:]** ペインで、**【正常】**、**【許容可能】**、および**【許容不可】** スコアに対してしきい値を設定できます。
    - **【無効なデータのコスト】** では以下の操作が可能です。
      - 各メトリックを選択して、メトリックの無効なデータのコストを設定する。
      - **【固定費】** オプションを選択して、メトリックのコストに定数値を割り当てる。コストの単位を変更するには**【コスト単位の変更】** をクリックします。
      - **【変動費】** オプションを選択して、数値カラムを変動費としてメトリックに添付する。数値カラムを選択するには**【カラムの選択】** をクリックします。
  8. **【次へ】** をクリックします。
  9. **【手順 6/7】** 画面で、次のタスクを実行できます。
    - メトリックを追加するメトリックグループを選択します。
    - **【デフォルト - メトリック】** ペインで、デフォルトのメトリックウェイト「0」をダブルクリックして値を変更できます。
    - **【メトリックのしきい値:]** ペインで、**【正常】**、**【許容可能】**、および**【許容不可】** スコアに対してしきい値を設定できます。
  10. **【次へ】** をクリックします。
  11. **【手順 7/7】** 画面で、ランタイム環境を選択します。
  12. **【保存】** をクリックしてスコアカードを保存するか、**【保存して実行】** をクリックしてスコアカードを保存してから実行します。

## スコアカードの実行

スコアカードを実行してカラムのスコアを生成します。

1. **【アセット】** パネルで、実行するスコアカードを選択します。



2. スコアカードをクリックして開きます。  
スコアカードが【スコアカード】ワークスペースに表示されます。
3. 【アクション】 > 【スコアカードの実行】をクリックします。
4. 【メトリック】 ペインからスコアを選択し、【カラム】 ペインからドリルダウンするカラムを選択します。
5. 【ドリルダウン】 オプションで、実データをドリルダウンするかステージングされたデータをドリルダウンするかを選択します。  
最適なパフォーマンスを実現するには、実データをドリルダウンします。
6. 【実行】 をクリックします。

## スコアカードの表示

スコアカードを実行して、各メトリックのスコアを表示します。スコアカードには、スコアが円グラフおよび棒グラフとして表示されます。有効なデータまたは無効なデータを表示します。また、データオブジェクトの名前、スコアカードの情報、メトリックグループスコア、およびメトリックウェイトを表示することもできます。

1. スコアカードを実行してスコアを表示します。
2. 表示するスコアが含まれているメトリックを選択します。
3. 【アクション】 > 【ドリルダウン】 をクリックして、カラムで有効なデータの行および無効なデータの行を表示します。  
Analyst ツールにより、無効なデータの行が【ドリルダウン】 セクションにデフォルト表示されます。

## スコアカードの編集

スコアカードのメトリックの有効な値を編集します。スコアカードを編集する前に、そのスコアカードを実行する必要があります。

1. 【ライブラリ】 ワークスペースの【アセット】 ペインで、編集するスコアカードをクリックします。  
スコアカードが【スコアカード】 ワークスペースに表示されます。
2. バージョン管理システムを有効にしている場合は、【アクション】 > 【チェックアウト】 をクリックします。
3. 【アクション】 > 【編集】 > 【全般】 をクリックします。  
【スコアカードの編集】 ダイアログボックスが表示されます。
4. 【全般】 タブで、必要に応じてスコアカードの名前と説明を編集できます。
5. 【メトリック】 タブをクリックします。
6. 【メトリック】 ペインでスコアを選択し、【スコアで使用: 値】 ペイン内のすべての値のリストから有効な値を設定します。
7. 【メトリックしきい値】 ペインで、必要に応じてスコアしきい値を変更できます。
8. 各メトリックの無効なデータのコストを確認し、必要に応じて変更を加えます。
9. 【スコアカードフィルタ】 タブをクリックします。



10. フィルタを追加、編集、または削除できます。
11. **【メトリックグループ】** タブをクリックします。
12. メトリックグループを作成、編集、または削除できます。  
また、**【メトリックグループ】** タブで、メトリックウェイトおよびメトリックしきい値を編集することもできます。
13. **【通知】** タブをクリックします。
14. 必要に応じて、スコアカード通知設定を変更できます。  
メトリックおよびメトリックグループにグローバル設定およびカスタム設定を行うことができます。
15. ランタイム環境として、**【ネイティブ】**、**【Blaze】**、**【Spark】**、または **【Databricks】** を選択します。ランタイム環境に **【Blaze】** または **【Spark】** を選択した場合、**【参照】** をクリックして Hadoop 接続を選択します。**【Databricks】** を選択した場合は、**【選択】** をクリックして、Databricks 接続を選択します。
16. **【保存】** をクリックしてスコアカードを保存するか、**【保存して実行】** をクリックして変更を保存してからスコアカードを実行します。
17. **【チェックイン】** をクリックします。

## メトリック

メトリックは、スコアカードに含まれるルールの出力またはデータソースのカラムです。スコアカードを作成する際に、各メトリックにウェイトを割り当てることができます。スコアカード内の関連し合うメトリックを1つのまとまりとして分類するには、メトリックグループを作成します。

## メトリックウェイト

スコアカードを作成する際に、各メトリックにウェイトを割り当てることができます。ウェイトのデフォルト値は0です。

スコアカードを実行すると、Analyst ツールは、ユーザーが各メトリックに割り当てたメトリックスコアに基づいて、各メトリックグループの加重平均を計算します。

例えば、メトリック M1 にウェイト W1、メトリック M2 にウェイト W2 を割り当てるとします。Analyst ツールは、次の式を使用して加重平均を計算します。

$$(M1 \times W1 + M2 \times W2) / (W1 + W2)$$

## データ品質の値

ソースデータのデータ品質の測定値は、組織のデータアセット管理において重要な情報です。スコアカードに示されたメトリックの無効なデータのコストは、組織でソースデータのデータ品質を監視する際に、値を導出するのに役立ちます。データアナリストとして、値（通貨単位やカスタム単位など）をメトリックやメトリックグループに関連付けることができます。このような場合、スコアカードを実行して、ソースデータ内の無効なデータの合計コストを表示できます。

ビジネスニーズに合わせて、メトリックのコスト単位を定義できます。スコアカードの作成時または編集時に、各メトリックの変動費または固定費を設定することもできます。

## 固定費

固定費とは、スコアカード内のメトリックに割り当てる定数値です。事前定義されたコスト単位を選択するか、ビジネスニーズに合わせてカスタムのコスト単位を作成できます。

## 変動費

変動費は、データソースの数値カラムの値に基づいてメトリックに割り当てる値です。データ統合サービスでは、コストに割り当てたカラムまたは仮想カラムに基づいてメトリックの変動費を計算します。

### 例

ローン担当者として、顧客がローン返済額を提示できるように、顧客に支払記入帳を用意する必要があります。スコアカードを使用して顧客アドレスの精度を測定し、確実に支払記入帳を配送します。アドレス精度メトリックの月次支払額カラムに変動費を設定できます。スコアカードを実行して、顧客が月次額を期限内に支払わなかった場合にローン部門が失う合計コストを計算します。

## しきい値の定義

スコアカードの各スコアに対してしきい値を設定できます。しきい値では、レコード内のカラムで許容できる不良データの範囲を割合で指定します。正常、許容可能、または許容不可の範囲のデータに対してしきい値を設定できます。スコアカードにカラムを追加するとき、またはスコアカードを編集するときに、各カラムに対してしきい値を定義できます。

スコアカードのカラムに対してしきい値を定義する前に、次の前提条件タスクのいずれかを完了します。

- **【スコアカードに追加】** ダイアログボックスで、プロファイルを開いて、プロファイルからのカラムをのスコアカードに追加します。
  - 必要に応じて、**【ライブラリ】** ワークスペース内のスコアカードをクリックし、**【アクション】** > **【編集】** を選択して、**【スコアカードの編集】** ダイアログボックスでスコアカードを編集します。
1. **【スコアカードに追加】** ダイアログボックスまたは **【スコアカードの編集】** ダイアログボックスの **【メトリック】** ペインで、各メトリックを選択します。
  2. **【メトリックのしきい値】** ペインで、許容不可範囲の上限および許容可能範囲の下限を示すしきい値を入力します。  
しきい値は小数点以下 2 桁まで設定できます。
  3. **【次へ】** または **【保存】** をクリックします。

## 【メトリックグループ】

スコアカード内の関連し合うスコアを 1 つのまとまりとして分類するには、メトリックグループを作成します。デフォルトでは、すべてのスコアがデフォルトメトリックグループに分類されます。

メトリックグループを作成したら、デフォルトメトリックグループから別のメトリックグループにスコアを移動できます。デフォルトメトリックグループをはじめとするメトリックグループは、名前と説明を変更して、編集することができます。不要になったメトリックグループは削除できますが、デフォルトメトリックグループは削除できません。

## メトリックグループの作成

メトリックグループを作成して、そのグループにスコアカード内の関連し合うスコアを追加します。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。  
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。  
**【スコアカードの編集】** ウィンドウが表示されます。
3. **【メトリックグループ】** タブをクリックします。  
デフォルトグループが **【メトリックグループ】** パネルに表示され、デフォルトグループのスコアが **【メトリック】** パネルに表示されます。
4. **【新規グループ】** アイコンをクリックし、メトリックグループを作成します。  
**【メトリックグループ】** ダイアログボックスが表示されます。
5. 名前を入力し、必要に応じて説明を入力します。
6. **【OK】** をクリックします。
7. **【保存】** をクリックして、変更内容をスコアカードに保存します。

## メトリックグループへのスコアの移動

メトリックグループを作成した後で、関連し合うスコアをそのメトリックグループに移動できます。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。  
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。  
**【スコアカードの編集】** ウィンドウが表示されます。
3. **【メトリックグループ】** タブをクリックします。  
デフォルトグループが **【メトリックグループ】** パネルに表示され、デフォルトグループのスコアが **【メトリック】** パネルに表示されます。
4. **【メトリック】** パネルでメトリックを選択し、**【メトリックの移動】** アイコンをクリックします。  
**【メトリックの移動】** ダイアログボックスが表示されます。  
**注:** 複数のスコアを選択するには、Shift キーを押しながら選択します。
5. スコアの移動先のメトリックグループを選択します。
6. **【OK】** をクリックします。

## メトリックグループの編集

メトリックグループを編集して、名前と説明を変更します。デフォルトメトリックグループの名前も変更できます。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、編集するスコアカードをクリックします。  
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【アクション】** > **【編集】** をクリックします。  
**【スコアカードの編集】** ウィンドウが表示されます。
3. **【メトリックグループ】** タブをクリックします。

デフォルトメトリックグループが【**メトリックグループ**】パネルに表示され、デフォルトグループのメトリックが【**メトリック**】パネルに表示されます。

4. 【**メトリックグループ**】パネルでメトリックグループを選択し、【**グループの編集**】アイコンをクリックします。  
【**編集**】ダイアログボックスが表示されます。
5. 名前を入力し、必要に応じて説明を入力します。
6. 【**OK**】をクリックします。

## メトリックグループの削除

無効になったメトリックグループを削除できます。メトリックグループを削除するときに、メトリックグループ内のスコアをデフォルトメトリックグループに移動するように選択できます。デフォルトメトリックグループは削除できません。

1. 【**ライブラリ**】ワークスペースの【**アセット**】ペインで、編集するスコアカードをクリックします。  
スコアカードが【**スコアカード**】ワークスペースに表示されます。
2. 【**アクション**】 > 【**編集**】をクリックします。  
【**スコアカードの編集**】ウィンドウが表示されます。
3. 【**メトリックグループ**】タブをクリックします。  
デフォルトメトリックグループが【**メトリックグループ**】パネルに表示され、デフォルトグループのメトリックが【**メトリック**】パネルに表示されます。
4. 【**メトリックグループ**】パネルでメトリックグループを選択し、【**グループの削除**】アイコンをクリックします。  
【**グループの削除**】ダイアログボックスが表示されます。
5. メトリックグループを削除する前に、メトリックグループ内のメトリックを削除するオプション、またはメトリックをデフォルトメトリックグループに移動するオプションを選択します。
6. 【**OK**】をクリックします。

## カラムのドリルダウン

スコアに対応するカラムをドリルダウンして、有効なデータ行または無効なデータ行を表示したときに示されるカラムを選択します。ドリルダウン対象として選択したカラムは**ドリルダウン**パネルに表示されます。

1. スコアを表示するためにスコアカードを実行します。
2. 表示するスコアが含まれているカラムを選択します。
3. 【**アクション**】 > 【**ドリルダウン**】をクリックして、カラムの有効なデータ行または無効なデータ行を表示します。
4. 【**アクション**】 > 【**カラムのドリルダウン**】をクリックします。

選択したスコアに対応するカラムが【**ドリルダウン**】パネルに表示されます。デフォルトでは、Analyst ツールにはカラムの有効なデータ行が表示されます。無効なデータ行を表示する場合は、【**無効**】をクリックします。

# 傾向グラフ

傾向グラフを使用して、メトリックスコアおよびメトリックの無効なデータのコストが、特定期間にわたって変化する様子を監視します。

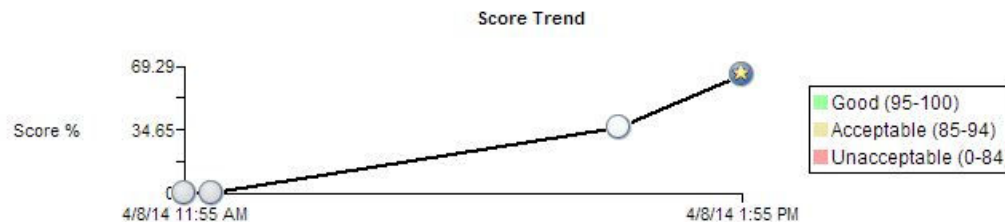
傾向グラフには、縦軸にスコアまたはコストの値を配置し、横軸にすべてのスコアカード実行を配置した、スコアグラフとコストグラフの両方が含まれます。デフォルトでは、傾向グラフには、最新の 10 のスコアカード実行で生成されたデータが表示されます。傾向グラフには、メトリックの合計行数と無効な行数を表示できます。傾向グラフには、スコアおよびコストの傾向が一定であるか、最後のスコアカード実行に基づいて上昇または下降しているのかも表示されます。

Analyst ツールでは、日付ごとの履歴スコアカード実行データと最新の有効なスコアの値を使用してスコアが計算されます。また、グラフの最新のしきい値設定を使用してスコアポイントの色が表示されます。スコアに対して、[正常]、[許容可能]、および [許容不可] のしきい値を表示できます。しきい値は、スコアカードでのスコアの値の編集後、スコアカードを実行するたびに更新されます。スコアカードをエクスポートすると、エクスポートファイル内のスコア情報とコスト情報を含め、Analyst ツールに傾向グラフの情報が含まれます。

## スコアの傾向グラフ

スコアの傾向グラフは、複数のプロファイル実行にわたって、メトリックのスコアの変化の様子を視覚的に表したものです。スコアの傾向グラフでは、メトリックのスコアの値が縦軸に配置され、すべてのスコアカード実行が横軸に配置されます。

次の画像は、スコアの傾向グラフの例を示しています。



### 例

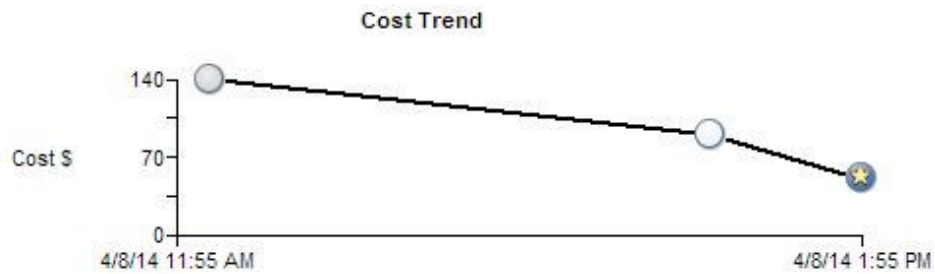
データアナリストはデータ品質を監視して、マッピングおよびその他のプロセスの変更によってデータ品質のスコアが上昇しているかどうかを分析できます。データ品質の変化を測定した後は、組織がデータ品質の変化を分析および使用できるようレポートできます。例えば、複数のスコアカード実行の終了で、社会保障番号カラム内の有効な値の割合が 84 から 90 に動いたとします。素早く分析するには、このデータ品質の変化を視覚的なグラフとしてレポートします。

## コストの傾向グラフ

コストの傾向グラフは、複数のプロファイル実行にわたって、メトリックの無効なデータの変化の様子を視覚的に表したものです。コストの傾向グラフでは、組織内のデータ品質の影響を測定することができます。コストの傾向グラフでは、コストの値が縦軸に配置され、すべてのスコアカードの実行が横軸に配置されます。また、メトリックの無効なデータと有効な値の合計コストが、コストの傾向グラフの下グリッドに表示されます。

コスト傾向グラフは、high-value レコードでの無効なデータの影響を追跡するのに役立ちます。固定費を使用して無効なデータを計算した場合、high-value レコードでの無効なデータの影響を見落とす恐れがあります。傾向グラフに、複数のスコアカードの実行にわたるスコアの改善とコスト全体の減少が示されることが原因で、この問題が発生します。ただし、high-value レコードには、スコアカードに現れる少数のデータ品質の問題が存在する可能性があります。

次の画像は、コストの傾向グラフの例を示しています。



#### 例

ある金融機関には、多額の銀行預金額と投資額（\$1,000 万など）を持つ残高の大きい顧客が複数存在します。残高の小さい顧客も多数存在します。スコア傾向グラフに、一定期間にわたるスコアの改善が示されることがあります。ただし、残高の大きい少数の顧客口座の住所または性別の間違いが、組織の最も価値の高い顧客との関係に影響を及ぼす恐れがあります。預金残高カラムを、無効なデータを計算するための変動費のカラムとして設定できます。カラムが原因で無効なデータのコストが高くなっている場合、合計値にリスクがあるとみなして、即座に対処できます。

## 傾向グラフの表示

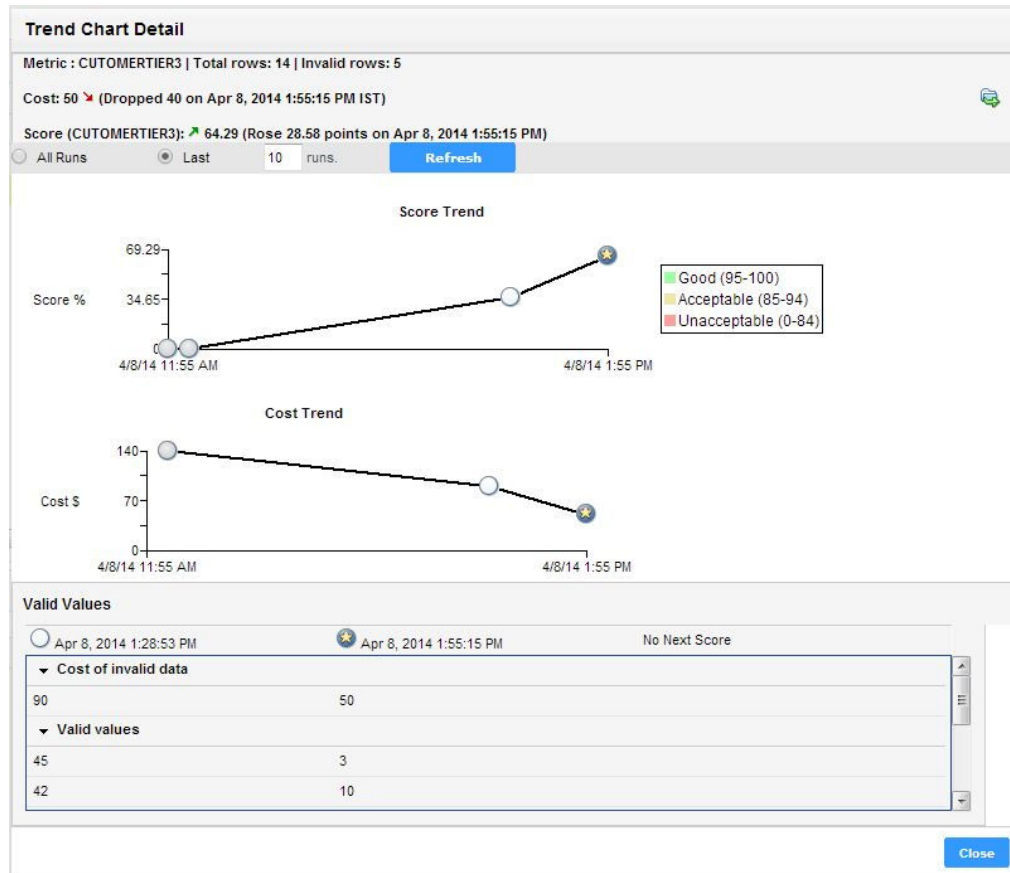
メトリックごとに傾向グラフを表示して、スコアまたは無効なデータのコストの変化を時系列で監視できます。

1. **【ライブラリ】** ワークスペースでスコアカードを含むプロジェクトまたはフォルダを選択します。
2. スコアカードをクリックして開きます。  
スコアカードが**【スコアカード】** ワークスペースに表示されます。
3. **【スコアカード】** ビューでメトリックを選択します。

4. [アクション] > [傾向グラフの表示] をクリックします。

[傾向グラフの詳細] ダイアログボックスが表示されます。

次の画像は、[傾向グラフの詳細] ダイアログボックスを示しています。



スコアとコストの値の変化を時系列で確認できます。ダイアログボックスの上部には、合計行数と無効な行の数が表示されます。Analyst ツールでは、日付ごとの履歴スコアカード実行データと最新の有効なスコアの値を使用してスコアが計算されます。スコアの傾向グラフとコストの傾向グラフの下には、メトリックの有効な値と、無効なデータのコストが表示されます。

## 傾向グラフのエクスポート

スコアの傾向グラフおよびコストの傾向グラフを.xlsx ファイルにエクスポートして、ファイル内のデータを表示できます。

1. スコアカードを開きます。
2. メトリックを選択して、[アクション] > [傾向グラフの表示] をクリックします。

[傾向グラフの詳細] ダイアログボックスが表示されます。

3. [データのエクスポート] アイコンをクリックします。

[ファイルへのデータのエクスポート] ダイアログボックスが表示されます。

4. [ファイル名] フィールドにファイル名を入力します。必要に応じて、デフォルトのファイル名を使用できます。

デフォルトのファイル形式は Microsoft Excel です。

5. **【無効な行】** フィールドに、エクスポートする無効な行の数を入力します。フィールドには最大 100,000 までの値を入力できます。
6. **【コードページ】** フィールドで、ファイルのコードページを選択できます。
7. **【OK】** をクリックします。

## Informatica Analyst のスコアカードダッシュボード

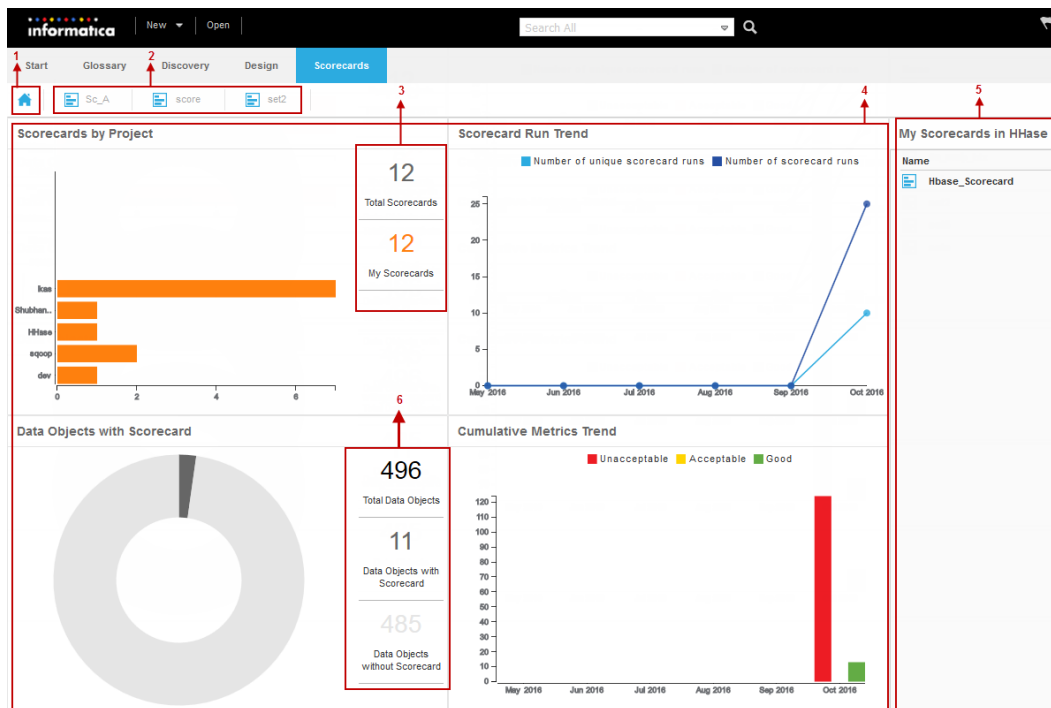
Informatica Analyst の **【スコアカード】** ワークスペースには、スコアカードダッシュボードが表示されます。スコアカードダッシュボードで、スコアカード付きデータオブジェクト、過去 6 か月間のスコアカード実行傾向、プロジェクトのスコアカード、1 か月間に実行されたすべてのスコアカードに対する、正常、許容可能、および許容不可のメトリックの集計、アセットリストペインを表示できます。

スコアカードが他のユーザーによって変更された場合、マシンのスコアカードダッシュボードは自動更新されません。スコアカードダッシュボードを更新するには、ファンクションキーの F5 を使用するか、ワークスペースまたはスコアカード結果のタブを切り替えます。

データをデータ系列またはデータポイントとしてペインに表示できます。各グラフでは、データポイントは不透明な小さい円として表示され、データ系列は水平棒、縦棒、またはスライスとして表示されます。

次の画像は、**【スコアカード】** ワークスペースのスコアカードダッシュボードおよびアセットペインを示しています。





1. スコアカードダッシュボードアイコン。スコアカードダッシュボードを表示します。
2. スコアカード結果のタブ。開いたスコアカードのスコアカード結果を表示します。
3. [プロジェクト別のスコアカード] ペインの凡例。すべてのプロジェクト内のスコアカードの総数と、すべてのプロジェクト内で自分の読み取りアクセス権限があるスコアカードの総数を表示します。
4. スコアカードダッシュボード。[プロジェクト別のスコアカード]、[スコアカード実行傾向]、[スコアカード付きデータオブジェクト]、および [累積メトリック傾向] ペインをダッシュボードに表示します。
5. アセットリストペイン。凡例、データ系列、またはデータポイントに関連付けられたスコアカードまたはデータオブジェクトのリストをグラフに表示します。
6. [スコアカード付きデータオブジェクト] ペインの凡例。データオブジェクトの総数、スコアカード付きデータオブジェクトの数、スコアカードなしのデータオブジェクトの数を表示します。

スコアカードダッシュボードでデータポイントまたはデータ系列をクリックすると、データポイントまたはデータ系列にマップされたスコアカードがアセットリストペインに表示されます。アセットリストペインでスコアカードをクリックすると、スコアカード結果が [スコアカード] ワークスペース内のタブに表示されます。アセットリストペインには、読み取りアクセス権限のあるスコアカードが表示されます。

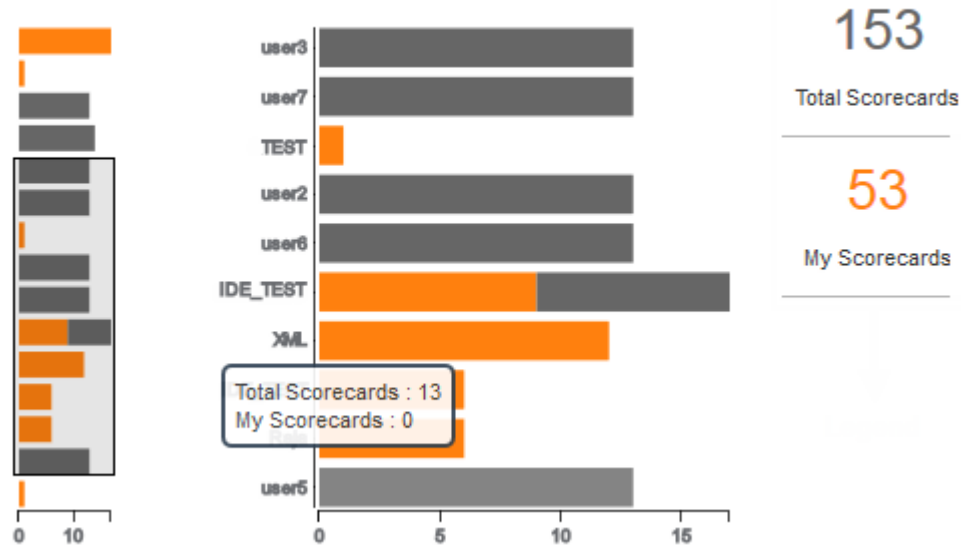
## プロジェクト別のスコアカード

[プロジェクト別のスコアカード] ペインには、モデルリポジトリ内のプロジェクトが、各プロジェクトに含まれるスコアカードの数と一緒に棒グラフとして表示されます。棒グラフの棒はプロジェクトを表します。グラフの X 軸にスコアカードの数が表示され、Y 軸にスコアカードがあるプロジェクトが表示されます。

プロジェクトのスコアカードは棒グラフで灰色に表示され、読み取りアクセス権限のあるスコアカードは棒グラフでオレンジ色に表示されます。凡例の [合計スコアカード] セクションに、モデルリポジトリ内のスコアカードの総数が表示されます。凡例の [マイスコアカード] セクションに、モデルリポジトリ内で読み取りアクセス権限のあるスコアカードの数が表示されます。

次の画像は、スコアカードダッシュボードの [プロジェクト別のスコアカード] ペインを示します。

## Scorecards by Project



このペインでは、次のグラフを表示できます。

- 詳細グラフ。モデルリポジトリ内にあるスコアカード付きのすべてのプロジェクトと、各プロジェクトに含まれるスコアカードの数を表示します。プロジェクトの数が 10 より大きい場合、[プロジェクト別のスコアカード] ペインにスライダが表示されます。
- ミニグラフ。スライダ内のすべてのプロジェクトと、各プロジェクトに含まれるスコアカード数が詳細グラフに表示されます。

ミニグラフにポインタを移動すると、スコアカードの総数と、プロジェクト内でアクセス権限を持つスコアカードの数がデータラベルに表示されます。

読み取りアクセス権限のあるプロジェクトのスコアカードを表示するには、水平棒のオレンジ色の部分をクリックします。モデルリポジトリ内で読み取りアクセス権限のあるスコアカードをすべて表示するには、棒グラフの【マイスコアカード】をクリックします。スコアカードがアセットリストペインに表示されます。スコアカード結果を表示するには、アセットペインでスコアカードをクリックします。

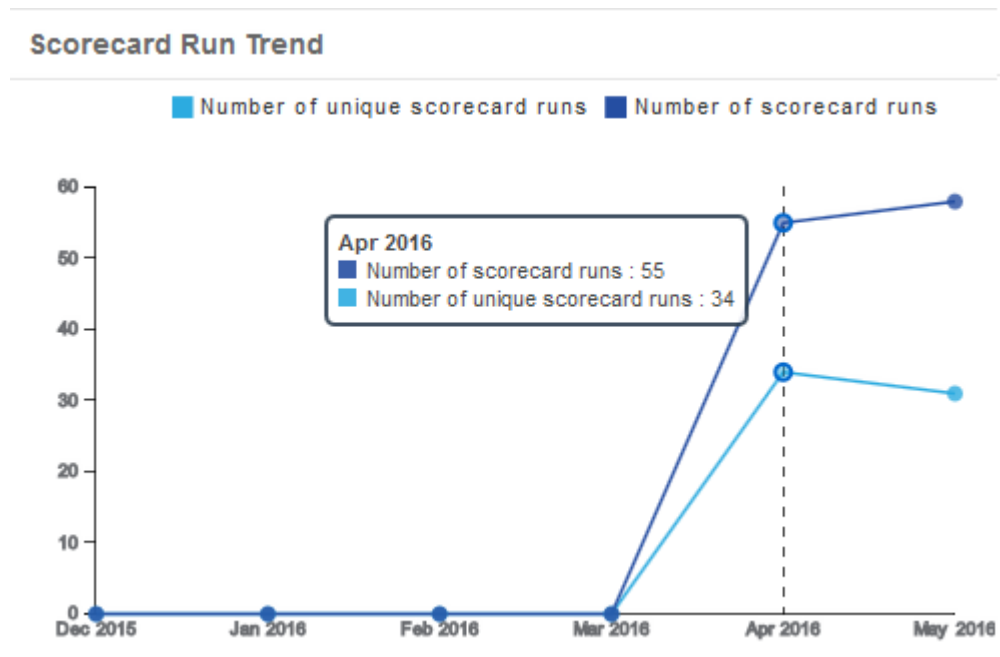
## スコアカード実行傾向

【スコアカード実行傾向】 ペインには、現在の月と過去 5 か月間のスコアカード実行傾向が、マーカー付きの折れ線グラフとして表示されます。グラフの X 軸に現在の月と過去 5 か月間が表示され、Y 軸にスコアカードの数が表示されます。折れ線グラフのマーカーはデータポイントです。グラフのマーカーにポインタを移動すると、月のスコアカード実行のサマリがデータラベルに表示されます。

このペインでは、次のマーカーを表示できます。

- **スコアカードの実行回数。** 1 か月間のスコアカードの実行総数がマーカーに表示されます。
- **一意のスコアカードの実行回数。** 1 か月間の一意のスコアカードの実行総数がマーカーに表示されます。

次の画像は、スコアカードダッシュボードの【スコアカード実行傾向】 ペインを示します。



ペインのマーカーをクリックすると、マーカーにマップされたスコアカードがアセットリストペインに表示されます。読み取りアクセス権限のあるスコアカードを表示できます。スコアカード結果を【スコアカード】ワークスペースに表示するには、アセットリストペインのスコアカードをクリックします。

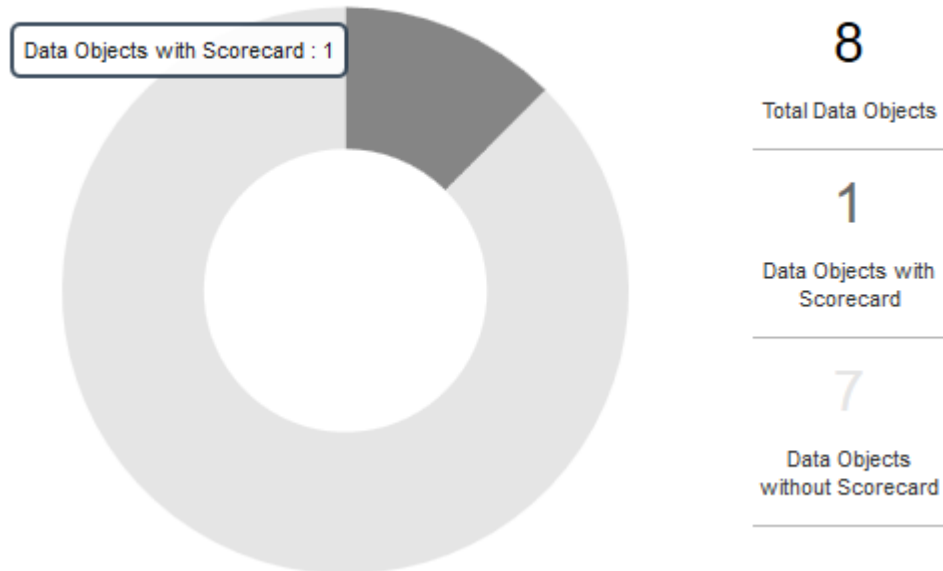
## スコアカード付きデータオブジェクト

**【スコアカード付きデータオブジェクト】** ペインには、ドーナツグラフが表示されます。このグラフでは、スコアカード付きデータオブジェクトの数と、スコアカードなしのデータオブジェクトの数をスライスとして表示できます。

ドーナツグラフにポインタを移動すると、スライスにマップされるデータがデータラベルに表示されます。

次の画像は、スコアカードダッシュボードの【スコアカード付きデータオブジェクト】ペインを示しています。

## Data Objects with Scorecard



【スコアカード付きデータオブジェクト】 ペインの凡例に、次のデータ統計が表示されます。

- 合計データオブジェクト。【ライブラリ】ワークスペースの【アセット】ペインで、【データオブジェクト】フォルダ内のデータオブジェクトの総数を表示します。データオブジェクトには、論理データオブジェクトやカスタマイズされたデータオブジェクトが含まれます。
- スコアカード付きデータオブジェクト。スコアカード付きデータオブジェクトの数を表示します。
- スコアカードなしのデータオブジェクト。スコアカードなしのデータオブジェクトの数を表示します。

ドーナツグラフのスライスか、【スコアカード付きデータオブジェクト】や【スコアカードなしのデータオブジェクト】の凡例をクリックすると、ドーナツグラフのスライスまたは凡例にマップされたスコアカードがアセットリストペインに表示されます。

## 累積メトリック傾向

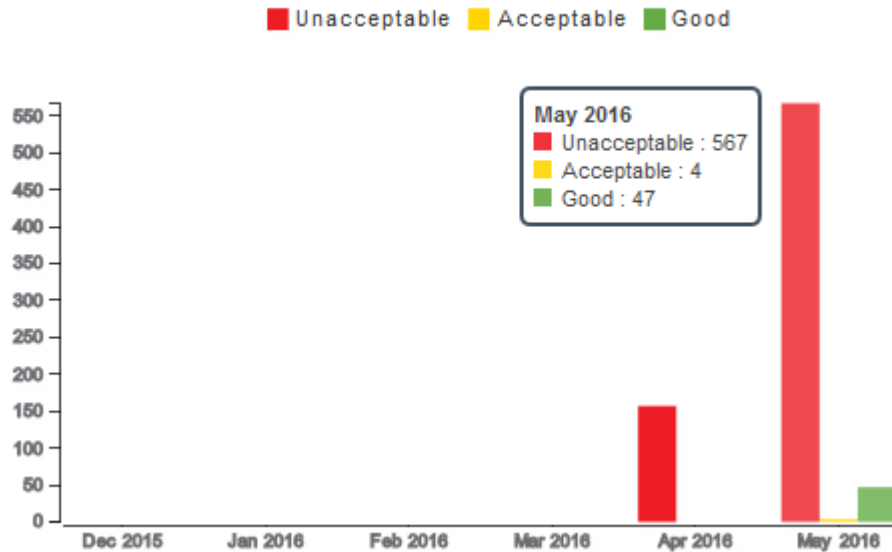
【累積メトリック傾向】 ペインには、縦棒グラフが表示されます。このグラフでは、1 か月間に実行されたすべてのスコアカードに対する、正常、許容可能、および許容不可のメトリックの集計を縦棒として表示できます。

【累積メトリック傾向】 ペインを使用し、現在の月と過去 5 か月間のメトリック傾向を表示して分析できます。

縦棒にポインタを移動すると、月のメトリックのサマリがデータラベルに表示されます。ペインの縦棒をクリックすると、関連するスコアカードがアセットリストペインに表示されます。読み取りアクセス権限のあるスコアカードを表示できます。スコアカード結果を表示するには、アセットリストペインでスコアカードをクリックします。

次の画像は、スコアカードダッシュボードの【累積メトリック傾向】 ペインを示します。

## Cumulative Metrics Trend



スコアカードの傾向が1か月間に時間とともに変化すると、アセットリストペインに許容不可メトリックリストおよび正常メトリックリストのスコアカードがいくつか表示されることがあります。メトリックを分析するには、スコアカードを開いてスコアカード結果を表示します。

### 例

カリフォルニア州にある一連の小売店の地域マネージャだとします。Sales テーブルの Sales\_SC スコアカードを作成します。

Sales\_SC スコアカードの Sales\_amt メトリックに対して次のしきい値を設定します。

- 許容不可= 0%~40%
- 許容可能= 41%~89%
- 正常= 90%~100%

販売データを毎日キャプチャするには、スケジューラサービスを使用して、Sales テーブルのスコアカードを毎晩実行します。5月を管理する月次レポートを作成し、スコアカードダッシュボードを使用してレポートを検証します。5月の場合、グラフの許容不可メトリックおよび正常メトリックの縦棒をクリックすると、アセットリストペインに Sales\_SC スコアカードが表示されます。

5月全体の販売を分析すると、次の傾向が見つかります。

1. 5月1日から5月25日まで、Sales\_amt メトリックは計算上 40%未満であり、許容不可メトリックとマークされます。
2. 5月の最終週、売上げの上昇により、Sales\_amt メトリックは計算上 98%を超え、正常メトリックとマークされます。

# Informatica Analyst のスコアカードエクスポートファイル

スコアカードの結果は Microsoft Excel ファイルにエクスポートできます。Analyst ツールにより、ファイルは XLSX 形式で保存されます。スコアカードをエクスポートすると、Microsoft Excel ファイルにはスコアカードのサマリ、傾向グラフ、無効な行、およびスコアカードのプロパティが複数のワークシートに表示されます。

スコアカードをエクスポートするときに、次のオプションを設定できます。

## 【データ】

データをエクスポートする際は次のオプションのいずれかを選択します。

- [すべて]。スコアカードのサマリ、傾向グラフ、無効な行、およびスコアカードのプロパティを Microsoft Excel ファイルにエクスポートします。
- [サマリビュー]。スコアカードのサマリ、傾向グラフ、およびスコアカードのプロパティを Microsoft Excel ファイルにエクスポートします。

## 【無効な行】

エクスポートする無効な行の数を入力します。フィールドには最大 100,000 までの値を入力できます。デフォルトでは、フィールドに 100 が表示されます。Analyst ツールは、各メトリックに対して最大で 100 の無効な行をワークシートにエクスポートします。

1 つのメトリックに対して 100 を超える無効な行をエクスポートすることを選択した場合は、データ統合サービスは次の手順を実行して残りのデータをエクスポートします。

1. スコアカード用のフォルダを<INFA\_HOME>/tomcat/bin/reject の場所に作成します。データ統合サービスは、スコアカード名とファイル作成時刻に基づいて、<scorecard\_name>\_HH\_MM\_SS 形式でフォルダに名前を付けます。  
例えば、スコアカード名が SD1 でファイル作成時刻が 2:23:15 の場合、データ統合サービスはフォルダに *SD1\_02\_23\_15* という名前を付けます。
2. スコアカードの各メトリックに対してサブフォルダを作成します。  
例えば、スコアカード SD1 に M1、M2、および M3 という名前のメトリックがある場合、データ統合サービスは M1、M2、および M3 という名前の 3 つのサブフォルダを作成します。
3. メトリックサブフォルダに Microsoft Excel ファイルを作成します。ファイルには、メトリック名と増分数に基づいて、<metric\_name>\_<IncrementalNumberStartingWith0>形式で名前が付けられます。最後に作成されたファイルは<metric\_name>\_Remaining 形式を使用します。各 Excel ファイルには、最大 10K の無効な行を含めることができます。  
例えば、メトリック M1 に 30K の無効な行がある場合、データ統合サービスは M1\_0、M1\_1、および M1\_Remaining という名前の 3 つの Microsoft Excel ファイルを作成し、各ファイルに 10K の無効な行を保存します。
4. データ統合サービスは、スコアカード内の他のすべてのメトリックに対して手順 3 を繰り返します。

## コードページ

コードページを選択します。

## Informatica Analyst からのスコアカード結果のエクスポート

スコアカード結果を「.xlsx」ファイルにエクスポートして、データをファイルで表示することができます。

1. スコアカードを開きます。
2. [アクション] > [データのエクスポート] をクリックします。  
[ファイルへのデータのエクスポート] ダイアログボックスが表示されます。

3. ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。  
デフォルトのファイル形式は Microsoft Excel です。
4. ファイルのコードページを選択します。
5. **[OK]** をクリックします。

## Microsoft Excel のスコアカードエクスポート結果

スコアカードの結果をエクスポートすると、Analyst ツールによって Microsoft Excel ファイルの複数のワークシートにその情報が保存されます。スコアカードのサマリ、傾向グラフ、無効な行、およびスコアカードのプロパティは、ファイル内のワークシートとして表示されます。Analyst ツールにより、ファイルは「xlsx」形式で保存されます。

次の表に、エクスポートファイルの各ワークシートに表示される情報を示します。

タブ	説明
スコアカードサマリ	エクスポートされたスコアカード結果のサマリ。この情報には、スコアカード名、各カラムの行の合計数、有効ではない行数、スコア、およびメトリックウェイトが含まれます。
傾向グラフ	スコアの傾向グラフです。
無効な行	各カラムの無効な行の詳細です。Analyst ツールは、最大で 100 行をワークシートにエクスポートします。 <b>[ファイルへのデータのエクスポート]</b> ダイアログボックスで <b>[データ]</b> > <b>[すべて]</b> を選択すると、 <b>[無効な行]</b> ワークシートが表示されます。
プロパティ	名前、タイプ、説明、場所などのスコアカードのプロパティです。

## スコアカード通知

特定のメトリックのスコア、メトリックグループのスコア、またはメトリックコストがしきい値を超えた場合は Analyst ツールで電子メールを送信するように、スコアカード通知設定を構成します。メトリックのスコアまたはメトリックグループのスコアは、しきい値を超えるか特定のスコア範囲（許容不可、許容可能、正常など）内にとどまります。メトリックコストの値は、設定したコストしきい値の上限または下限を超えて移動することがあります。

個々のメトリックスコア、メトリックグループ、メトリックコストに電子メール通知を設定できます。スコアにグローバル設定を使用すると、特定のメトリックのスコアがしきい値を超えて、スコア範囲である「正常」から「許容可能」、および「許容可能」から「不良」に変化した場合に、Analyst ツールが通知電子メールを送信します。また、スコアカードを再度実行してもスコアが「許容不可」のスコア範囲にとどまっている場合は、スコアカードの実行ごとに通知電子メールが送信されます。メトリックコストにグローバル設定を使用すると、選択したメトリックでの無効なデータのコストがしきい値の上限と下限を超えたときに、Analyst ツールが通知電子メールを送信します。

スコアがスコア範囲である「許容不可」から「許容可能」、および「許容可能」から「正常」に変化した場合にスコアカードユーザーが電子メール通知を受信するよう、通知設定をカスタマイズすることもできます。スコアカードの実行ごとにメトリックスコアまたはメトリックコストが特定の範囲内にとどまっている場合は、電子メール通知を送信するよう選択できます。通知設定で各メトリックの無効なデータの現在のコストを表示し、この値に基づいてコストしきい値を設定できます。

スコアカードを設定して電子メール通知を送信する前に、管理者は Administrator ツールで電子メールサービスを設定する必要があります。

## 通知電子メールメッセージテンプレート

Analyst ツールがスコアカード通知の一部として受信者に送信する電子メールメッセージのメッセージテキストおよび構造を設定できます。電子メールテンプレートには、オプションの導入部テキストセクション、読み取り専用のメッセージ本文セクション、およびオプションの結びのテキストセクションがあります。

次の表で、電子メールテンプレートのタグについて説明します。

タグ	説明
ScorecardName	スコアカードの名前。
ObjectURL	スコアカードへのハイパーリンク。ユーザー名とパスワードを入力する必要があります。
MetricGroupName	メトリックが属しているメトリックグループの名前。
CurrentWeightedAverage	現在のスコアカード実行に含まれるメトリックグループの加重平均値。
CurrentRange	現在のスコアカード実行に含まれるメトリックグループのスコア範囲（許容不可、許容可能、および正常など）。
PreviousWeightedAverage	前回のスコアカード実行に含まれるメトリックグループの加重平均値。
PreviousRange	前回のスコアカード実行に含まれるメトリックグループのスコア範囲（許容不可、許容可能、および正常など）。
MetricName	メトリックの名前。
MetricGroupName	メトリックグループの名前。
CurrentScore	最終スコアカード実行に基づくスコア。
CurrentRange	最終スコアカード実行に基づく現在のスコアが残っているスコア範囲。
PreviousScore	前回のスコアカード実行に基づくスコア。
PreviousRange	前回のスコアカード実行に基づくスコア範囲。
CurrentCost	最終スコアカード実行に基づく、メトリック内の無効なデータのコスト。
PreviousCost	前回のスコアカード実行に基づく、メトリック内の無効なデータのコスト。
ColumnName	メトリックが割り当てられているソースカラムの名前。
ColumnType	ソースカラムのタイプ。
RuleName	ルールの名前。



タグ	説明
RuleType	ルールのタイプ。
DataObjectName	ソースデータオブジェクトの名前。

## スコアカード通知の設定

スコアカード通知は、メトリックレベルとメトリックグルーplevelのいずれに対しても設定できます。グローバル通知設定は、個別に通知設定を持っていないメトリックやメトリックグループに適用されます。

1. Analyst ツールでスコアカードを実行します。
2. **【アクション】** > **【編集】** をクリックします。
3. **【通知】** タブをクリックします。
4. **【通知を有効にする】** を選択して、スコアカード通知の設定を開始します。
5. メトリックまたはメトリックグループを選択します。
6. **【通知】** チェックボックスをクリックし、メトリックまたはメトリックグループのグローバル設定を有効にします。
7. **【カスタム設定の使用】** を選択し、メトリックまたはメトリックグループの設定を変更します。  
スコアが **【許容不可】**、**【許容可能】**、および **【正常】** の範囲内にあり、しきい値を超えた場合に通知電子メールを送信するよう選択できます。メトリックコストがしきい値の上限または下限を超えた場合にも、通知電子メールを送信できます。
8. スコアカード通知のグローバル設定を編集するには、**【グローバル設定の編集】** アイコンをクリックします。  
**【グローバル設定の編集】** ダイアログボックスが表示されます。このダイアログボックスで、電子メールテンプレートなどの設定を編集できます。

## スコアカード通知のグローバル設定の実行

グローバルスコアカード通知設定を選択したときに、スコアが **【許容不可】** の範囲にある場合は、Analyst ツールがターゲットユーザに電子メールを送信します。通知設定を構成して、メトリックスコアまたはメトリックコストがしきい値を超えたときに電子メールを送信するようにもできます。電子メールアドレスとスコアカードのメッセージテキストを含む電子メールテンプレートを設定できます。

1. Analyst ツールでスコアカードを実行します。
2. **【アクション】** > **【編集】** > **【通知】** をクリックして、**【スコアカードの編集】** ダイアログボックスを開きます。
3. **【通知を有効にする】** を選択して、スコアカード通知の設定を開始します。
4. **【グローバル設定の編集】** アイコンをクリックします。  
**【グローバル設定の編集】** ダイアログボックスが表示されます。このダイアログボックスで、電子メールテンプレートなどの設定を編集できます。
5. **【許容可能スコアへの変化】** および **【許容不可スコアへの変化】** チェックボックスを使用して、メトリックスコアの電子メール通知を送信するタイミングを選択します。
6. **【コストの変化】** チェックボックスを使用して、メトリックコストの電子メール通知を送信するタイミングを選択します。

7. **【宛先】** フィールドに、受信者の電子メール ID を入力します。複数の電子メール ID を区切るには、セミコロンを使用します。  
送信者のデフォルトの電子メール ID は、ドメインの SMTP プロパティで設定された **【送信側電子メールアドレス】** です。
8. 電子メールの件名にテキストを入力します。
9. **【ボディ】** フィールドで、電子メールメッセージの導入部のテキストおよび結びのテキストを追加します。
10. グローバル設定を適用するには、**すべてのメトリックおよびメトリックグループに設定を適用する**を選択します。
11. **【OK】** をクリックします。

## スコアカードリネージュ

スコアカードリネージュは、データの発生元、パス、そしてメトリックまたはメトリックグループに対するデータフローを示します。スコアカードリネージュを使うと、メトリックまたはメトリックグループでスコアが許容範囲を超えてばらつくことの根本原因を分析することができます。Analyst ツールでスコアカードリネージュを表示します。

スコアカードリネージュを表示するには、次のタスクを完了します。

1. Informatica Administrator で Metadata Manager Service を Analyst Service に関連付けます。
2. プロジェクトを選択して、その中のスコアカードオブジェクトを、Developer tool 中の [メタデータマネージャのリソースファイルのエクスポート] オプションまたは `infacmd oie exportResources` コマンドを使用して XML ファイルにエクスポートします。
3. Metadata Manager で、エクスポートした XML ファイルを使用してリソースを作成してロードします。  
**注:** 作成して Metadata Manager にロードするリソースファイルに名前を付ける際は、`<MRS 名>_<プロジェクト名>` という形式にします。リソースファイルの作成とロードの方法に関する詳細は、『*Informatica PowerCenter Metadata Manager ユーザーガイド*』を参照してください。
4. Analyst ツールでスコアカードを開いてメトリックまたはメトリックグループを選択します。
5. スコアカードリネージュを表示します。

## Informatica Analyst でのスコアカードリネージュの表示

メトリックまたはメトリックグループのスコアカードリネージュを表示できます。Analyst ツールでスコアカードリネージュダイアグラムを表示できるようにするには、スコアカードリネージュと Metadata Manager のメタデータをロードする必要があります。

1. **【ライブラリ】** ワークスペースの **【アセット】** ペインで、表示するスコアカードをクリックします。  
スコアカードが **【スコアカード】** ワークスペースに表示されます。
2. **【スコアカード】** ビューで、メトリックまたはメトリックグループを選択します。
3. 右クリックして **【リネージュを表示する】** を選択します。  
スコアカードリネージュダイアグラムが別のウィンドウで表示されます。

**重要:** スコアカードオブジェクトのエクスポートされた XML 形式で Metadata Manager のリソースを作成またはロードしない場合、リソースはカタログで利用できないというエラーメッセージが出ることがあります。スコアカードリネージュへの XML ファイルのエクスポートの詳細については、[「スコアカードリネージュのリソースファイルのエクスポート」 \(ページ 177\)](#) を参照してください。

## 第 12 章

# Informatica Analyst でのデータドメイン検出

この章では、以下の項目について説明します。

- [Informatica Analyst でのデータドメイン検出の概要, 107 ページ](#)
- [Informatica Analyst のデータドメイングロッサリ, 107 ページ](#)
- [Informatica Analyst のデータドメイン検出オプション, 110 ページ](#)
- [Informatica Analyst でのデータドメイン検出実行カラムプロファイルの作成, 113 ページ](#)
- [Informatica Analyst でのカラムプロファイルとデータドメイン検出の編集, 114 ページ](#)
- [データドメイン検出実行プロファイルの実行, 115 ページ](#)
- [Informatica Analyst のデータドメイン検出結果, 115 ページ](#)
- [Informatica Analyst のデータドメイン検出エクスポートファイル, 116 ページ](#)

## Informatica Analyst でのデータドメイン検出の概要

プロファイルを作成してデータドメイン検出を実行するときに、ソースカラム、カラムデータおよびカラム名と照合するデータドメイン、ならびにサンプリングオプションを選択できます。データドメイン検出の適合条件を選択し、データドメイン検出の際に NULL 値を除外できます。

サンプリングオプションやフィルタを設定したプロファイルを作成して、データドメイン検出を実行できます。プロファイルを実行すると、データソースにサンプリングオプションとフィルタが適用されて、データセットが生成されます。データドメイン検出プロセスは、このデータセットを使用してデータドメインを検出します。

## Informatica Analyst のデータドメイングロッサリ

データドメイングロッサリには、データドメインとデータドメイングループが一覧表示されます。リストは、データドメインまたはデータドメイングループのいずれかでソートできます。データドメイングロッサリを使用して、データドメインとデータドメイングループを作成、管理、および削除します。Developer ツールでは、データドメインと関連付けられたルールロジックを表示したり変更したりすることができます。

## Informatica Analyst でのデータドメイングループの作成

データドメイングループは、データドメインを Personal Health Information (PHI)、Personally Identifiable Information (PII)、プロジェクトに関連する他のすべての概念グループなどの特定のグループに整理します。

1. **【管理】** > **【データドメイングロッサリ】** をクリックします。  
タブにデータドメイングロッサリが表示され、現在のデータドメインとデータドメイングループが一覧表示されます。
2. ナビゲータで、**【アクション】** > **【新規】** > **【データドメイングループ】** をクリックします。  
**【データドメイングループの作成】** ダイアログボックスが表示されます。
3. データドメイングループの名前と説明を入力します。
4. **【次へ】** をクリックします。
5. **【使用可能なデータドメイン】** ペインで、データドメイングループに追加するデータドメインを選択し、**【追加】** をクリックします。  
選択したデータドメインが **【選択したデータドメイン】** ペインに移動します。
6. **【完了】** をクリックします。  
データドメイングループがデータドメイングロッサリに追加されます。

## Informatica Analyst でのデータドメインの作成

データドメインを作成し、データドメイングロッサリに追加して、データドメインを 1 つまたは複数のデータドメイングループにグループ化できます。データドメインを作成するには、事前定義されたデータルールおよびカラム名ルールを使用するか、カラムプロファイル結果の値およびパターンを使用します。

データドメインを作成すると、Analyst ツールがデータドメインに関連付けられているルールや他の依存オブジェクトがデータドメイングロッサリにコピーします。データドメインに関連付けられているルールを編集するには、元のルールに移動してそこで変更を加える必要があります。その後、変更したルールをデータドメインに再び関連付けます。

1. **【管理】** > **【データドメイングロッサリ】** をクリックします。  
タブにデータドメイングロッサリが表示され、現在のデータドメインとデータドメイングループが一覧表示されます。
2. ナビゲータで、**【アクション】** > **【新規】** > **【データドメイン】** をクリックします。  
**【データドメインの作成】** ダイアログボックスが表示されます。
3. データドメインの名前と説明を入力します。
4. **【データルール】** チェックボックスをクリックして、カラムデータに基づいてデータドメインを検出します。**【カラム名ルール】** チェックボックスを選択して、データソース内のカラムタイトルに基づいてデータドメインを検出することもできます。  
**【選択】** ボタンが有効になります。
5. **【選択】** をクリックして、**【ルールの選択】** ダイアログボックスを開きます。
6. 適切なルールを選択し、**【OK】** をクリックします。  
選択したルールが、**【データルール】** フィールドと **【カラム名ルール】** フィールドに表示されます。
7. **【次へ】** をクリックします。
8. **【使用可能なデータドメイングループ】** ペインで、データドメインを含めるデータドメイングループを選択し、**【追加】** をクリックします。  
選択したデータドメイングループが **【選択したデータドメイングループ】** ペインに移動します。

9. **【完了】** をクリックします。

Analyst ツールがデータドメインをデータドメインングロッサリに追加されます。

## Informatica Analyst でのプロファイル結果からのデータドメインの作成

カラムプロファイルを実行して、ソースデータの値およびパターンを表示します。この後、プロファイル結果を確認し、そのプロファイル結果からデータドメインを作成できます。

1. カラムプロファイルを実行して、その結果を表示します。  
サマリビューにプロファイル結果が表示されます。
2. サマリビューでカラムをクリックすると、詳細ビューにカラム結果が表示されます。
3. **【値】** ペインまたは **【パターン】** ペインで、データドメインを作成するベースとなる値またはパターンを選択します。
4. 値またはパターンを右クリックし、**【データドメインの作成】** を選択します。  
**【データドメインの作成】** ダイアログボックスが表示されます。
5. データドメイン名と説明（省略可能）を入力します。
6. **【作成】** をクリックします。  
データドメインがデータドメインングロッサリに追加されます。

## Informatica Analyst でのデータドメインとデータドメイングループの検索

データドメインングロッサリで特定のデータドメインとデータドメイングループを検索することができます。**【データドメイン】** または **【データドメイングループ】** ビューを選択して、データドメインングロッサリ内のデータドメインのリストを表示することができます。

例えば、**PII** データドメイングループに追加したデータドメイン **Zipcode** があるとします。次の方法で、Zipcode とそのデータドメイングループ **PII** に関する詳細情報を検索することができます。

### データドメインを検索する。

ナビゲータの上部にあるテキストフィールドに、**zip** や **code** など、データドメイン名の一部を入力します。**【データドメイングループ】** ビューになっている場合は、**Zipcode** を含むデータドメイングループである **PII** が表示されます。**【データドメイン】** ビューが表示されている場合は、検索文字列の **Zip** または **code** を含むすべてのデータドメインが一覧表示されます。

注: 検索では大文字小文字は区別されません。

### すべてのデータドメイングループとその中のデータドメインを表示する。

ナビゲータで、**【データドメイングループビューの表示】** をクリックします。

### すべてのデータドメインを表示する。

ナビゲータで、**【データドメインビューの表示】** をクリックします。

### データドメインのプロパティを表示する。

**【データドメイン】** ビューになっていることを確認します。ナビゲータで、「**Zipcode**」をクリックすると、右側のペインにそのプロパティが表示されます。名前、タイプ、説明、関連付けられているルール、および属しているドメイングループ（この場合は **PII**）を表示できます。

データドメイングループのプロパティを表示する。

【データドメイングループ】ビューになっていることを確認します。ナビゲータで、「PII」をクリックすると、右側のペインにそのプロパティが表示されます。名前、タイプ、説明、および **PIIZipcode** を含む、データドメインのリストを表示できます。

データドメイングロッサリを更新する。

ナビゲータで、【アクション】 > 【更新】をクリックします。現在のビューに基づいて、データドメイングロッサリにデータドメインまたはデータドメイングループの現在のリストが表示されます。

## Informatica Analyst のデータドメイン検出オプション

データドメイン検出のカラム、データドメイン、および推測オプションを選択するには、データドメイン検出オプションを使用します。推測オプションでは、データドメイン検出の実行対象を、カラムデータ、カラム名、またはその両方から選択できます。

### Informatica Analyst でのデータドメインカラムの選択

【設定の指定】画面で【編集】をクリックすると、データドメイン検出の一部として実行するカラムを選択できます。プロファイルウィザードの【ソースの選択】画面に、データソース内のすべてのカラムを表示できます。カラムプロファイルおよびデータドメイン検出用にさまざまなカラムを選択できます。

以下の表で、データドメイン検出の【編集】ダイアログボックスの各プロパティについて説明します。

オプション	説明
名前	カラム名が表示されます。
タイプ	カラムの文書化されたデータ型が表示されます。
精度	カラムの最大精度が表示されます。
スケール	カラムのスケールが表示されます。
NULL 可能	NULL 値を指定できるカラムを示します。
キー	カラムがプライマリキーまたは外部キーのどちらとして文書化されているかを示します。

## Informatica Analyst でのデータドメインの選択

【設定の指定】画面の【データドメイン】ペインにデータドメイングループサリのすべてのデータドメインが一覧表示されます。データドメイン検出の実行対象とするデータドメインを選択できます。

次の表で、データドメイン検出の【データドメイン】プロパティについて説明します。

オプション	説明
名前	データドメイン名が表示されます。1 つ以上のデータドメインまたはデータドメイングループを選択できます。
説明	データドメインの説明が表示されます。
DomainGroups	データドメインが属しているデータドメイングループの名前が表示されます。

## Informatica Analyst のデータドメイン推測オプション

推測オプションでは、データドメイン検出の実行対象を、カラムデータ、カラム名、またはその両方から選択できます。プロファイルで分析できるソース行の最大数を指定できます。データドメイン検出の適合条件を選択できます。データドメイン検出から NULL 値を除外することが可能です。データドメインの推測オプションは、プロファイルウィザードの【設定の指定】画面で設定できます。

以下の表に、データドメイン検出の推測オプションを示します。

オプション	説明
データ	プロファイルをカラムデータに対して実行します。
カラム	プロファイルをカラムタイトルに対して実行します。
データおよびカラム	プロファイルをカラムデータとカラムタイトルの両方に対して実行します。
データドメインの最小一致率	データドメインの一致に必要なデータセット内の行の最小一致率。
行の最小数	データドメインの一致に必要なデータセット内の行の最小数。
データドメイン検出から NULL 値を除外	データドメイン検出用のデータセットから NULL 値を除外します。
編集	データドメイン検出のカラムを選択します。
すべての行	プロファイルをソースのすべての行に対して実行します。
サンプリング最初の	プロファイルを実行できる最大行数を選択します。Analyst ツールでは、ソースの先頭行から行が選択されます。行は最大で 2,147,483,647 行まで選択できます。
ランダムサンプリング	データソースから行のランダムサンプルを選択します。行は最大で 2,147,483,647 行まで選択できます。



オプション	説明
ランダムサンプリング（自動）	Analyst ツールでは、データソースのサイズに応じて行のランダムサンプルが選択されます。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次回のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

## 最小一致率

データドメイン検出の適合条件として、データセット内の行の最小割合を選択できます。

一致率は、一致する行の数を行の合計数で除算した割合です。

**注:** Analyst ツールでは、NULL 値は一致しない行とみなされます。NULL 値の数が多いカラムでは、最小一致率を低く指定しないと、結果がデータドメイン推測にならない場合があります。

### 例

行数が 10,000 のデータソースがあり、そのうちの 2,500 行のコメントカラムに社会保障番号が含まれている場合を考えます。カラムプロファイルとデータドメイン検出を作成し、適合条件として、行の最小割合を 30% に設定します。プロファイルを実行すると、結果には、推測されたデータドメインとして社会保障番号が表示されません。これは、最小の適合条件が、データソースに含まれる行の 30% または 3,000 行に設定されているためです。

## 適合する行の最小数

データドメイン検出の適合条件として、データセット内の行の最小数を選択できます。

### 例

行数が 10,000 のデータソースがあり、そのうちの 3 行のコメントカラムに電子メールアドレスが含まれている場合を考えます。カラムプロファイルとデータドメイン検出プロファイルを作成し、適合条件として、行の最小数を 1 に設定します。プロファイルを実行すると、結果には、他の推測されたデータドメインと一緒に、3 つの適合する行を持つ推測されたデータドメインとして、電子メールアドレスが表示されます。

## データドメイン検出からの NULL 値の除外

データソースに対してデータドメイン検出を実行する際に、NULL 値を除外することができます。NULL 値を除外するオプションを指定して行の最小割合を選択した場合、一致率は、合計行数からカラム内の NULL 値を除いた数で一致行数を除算した割合になります。

**【データドメイン検出から NULL 値を除外】** オプションと各種のサンプリングオプションやフィルタを選択すると、データドメイン検出プロセスの結果は異なるものになります。

次のシナリオでは、NULL 値を除外するオプションと、1 つのサンプリングオプションおよびフィルタを組み合わせで選択した場合のデータドメイン検出結果について説明しています。

- サンプリングオプションとして **【すべての行】** を選択し、フィルタは選択しない場合。データドメイン検出では、カラム内のすべての NULL 値が無視されます。
- サンプリングオプションを 1 つ選択し、フィルタは選択しない場合。サンプリングされたデータに含まれるすべての NULL 値が無視され、サンプリングされたデータの残りのデータに対してデータドメイン検出が実行されます。



- サンプルングオプションとして **【すべての行】** を選択し、フィルタを複数選択する場合。フィルタされたデータに含まれるすべての NULL 値が無視され、フィルタされたデータの残りのデータに対してデータドメイン検出が実行されます。
- サンプルングオプションを 1 つ選択し、フィルタを複数選択する場合。サンプル内のフィルタされたデータに含まれる NULL 値が無視され、フィルタされたデータの残りのデータに対してデータドメイン検出が実行されます。

## 例

行数が 10,000 のデータソースがあり、そのうちの 3,000 行のコメント列に社会保障番号が含まれている場合を考えます。カラムプロファイルとデータドメイン検出を作成し、次のオプションを選択します。

- **【データドメイン検出から NULL 値を除外】** オプションを選択します。
- サンプルングオプションとして **【すべての行】** を選択します。
- **【行の最小割合】** オプションを選択して、オプションを 12% に設定します。

プロファイルを実行すると、プロファイルはデータセットに対して実行され、データドメイン検出では NULL 値が無視されます。

# Informatica Analyst でのデータドメイン検出実行カラムプロファイルの作成

Analyst ツールでカラムプロファイルを作成してデータドメイン検出を実行する前に、データドメインを少なくとも 1 つ作成する必要があります。プロファイルでは、定義済みデータドメインに一致するカラム名とカラムデータの両方を検出できます。

1. **【検出】** ワークスペースで、**【プロファイル】** をクリックするか、Analyst ツール内の任意の場所で **【新規】** > **【プロファイル】** を選択します。  
**【新規プロファイル】** ウィザードが表示されます。
2. **【単一ソース】** オプションがデフォルトで選択されています。**【次へ】** をクリックします。
3. **【全般プロパティの指定】** 画面で、プロファイルの名前と説明（省略可能）を入力します。**【場所】** フィールドで、プロファイルの作成先のプロジェクトまたはフォルダを選択します。**【次へ】** をクリックします。
4. **【ソースの選択】** 画面で、**【選択】** をクリックしてデータオブジェクトを選択するか、**【新規】** をクリックしてデータオブジェクトをインポートします。**【次へ】** をクリックします。
5. **【設定の指定】** 画面で、カラムプロファイル、データドメイン検出、またはカラムプロファイルとデータドメイン検出のうちどれを実行するかを選択します。デフォルトでは、カラムプロファイルオプションが選択されています。
  - **【データドメイン検出を実行】** を選択して、データドメイン検出を実行します。**【データドメイン】** ペインでデータドメインオプションを選択します。
  - **【カラムプロファイルを実行】** と **【データドメイン検出を実行】** を選択して、カラムプロファイルとデータドメイン検出を実行します。**【データドメイン】** ペインでデータドメインオプションを選択します。

**注:** デフォルトでは、カラムプロファイルの対象として選択したカラムには、データドメイン検出も適用されます。**【編集】** をクリックすると、カラムプロファイルの対象として選択したカラムかどうかとは無関係に、データドメイン検出の実行対象カラムを選択または選択解除できます。

  - データドメイン検出の実行対象として、**【データ】**、**【カラム】**、**【データとカラム】** を選択します。
  - **【次に対してプロファイルを実行】** ペインでサンプルングオプションを選択します。

- **【ドリルダウン】** ペインでドリルダウンオプションを選択します。必要に応じて、**【カラムの選択】** をクリックして、ドリルダウンするカラムを選択します。承認済みのデータ型またはデータドメインを持つカラムの場合は、データ型およびデータドメインの推測を省略できます。
  - 適合条件を選択すると、**【データドメイン検出から NULL 値を除外】** オプションが選択できます。
  - ランタイム環境として、**【ネイティブ】** または **【Hadoop】** を選択します。Hadoop ランタイム環境では、**【Blaze】** または **【Spark】** オプションを選択できます。**【Blaze】** オプションを選択した場合は、**【選択】** をクリックして、**【Hadoop 接続の選択】** ダイアログボックスで Hadoop 接続を選択します。**【Spark】** オプションを選択した場合は、**【選択】** をクリックして、**【Hadoop 接続の選択】** ダイアログボックスで Hadoop 接続を選択します。
6. **【ルールとフィルタの指定】** 画面では、プロファイルのルールおよびフィルタを追加、編集、または削除できます。
  7. **【保存して完了】** をクリックしてプロファイルを作成するか、**【保存して実行】** をクリックしてプロファイルを作成および実行します。

## Informatica Analyst でのカラムプロファイルとデータドメイン検出の編集

プロファイルのプロパティはプロファイルの実行後に変更できます。カラムプロファイルをデータドメイン検出の一部として実行した場合は、カラムプロファイルの設定を変更できます。

1. **【ライブラリ】** ワークスペースでプロファイルを含むプロジェクトを選択するか、**【アセット】** ペインでプロファイルを選択します。
2. プロファイル名をクリックします。  
サマリビューが **【検出】** ワークスペースに表示されます。
3. バージョン管理システムが有効になっている場合、**【アクション】** > **【チェックアウト】** をクリックしてプロファイルをチェックアウトします。
4. **【アクション】** > **【プロファイルの編集】** をクリックします。  
**【プロファイル】** ウィザードが表示されます。
5. 変更する内容に基づいて、次のページオプションのいずれかを選択します。
  - **全般プロパティの指定。** 名前、説明、場所などの基本プロパティを変更します。
  - **ソースの選択。** プロファイルの実行対象として、別の一致するデータソースとカラムを選択します。
  - **設定の指定。** カラムプロファイルまたはカラムプロファイルとデータドメイン検出を実行するように選択します。データドメインオプション、サンプリングオプション、およびドリルダウンオプションを編集します。
  - **ルールとフィルタの指定。** ルールとフィルタを作成、編集、または削除します。
6. **【保存して完了】** をクリックしてプロファイルを編集するか、**【保存して実行】** をクリックしてプロファイルを編集および実行します。
7. バージョン管理システムが有効な場合、次のタスクを実行する必要があります。
  - **【保存して完了】** をクリックしてプロファイルの編集を完了します。
  - **【サマリ】** ビューで、**【チェックイン】** をクリックしてプロファイルをチェックインします。
  - **【アクション】** > **【プロファイルの実行】** をクリックしてプロファイルを実行します。

# データドメイン検出実行プロファイルの実行

データドメインルールのパターンに一致するカラムを表示するには、データドメイン検出時にプロファイルを実行します。

1. **【ライブラリナビゲータ】**に移動し、**【プロジェクト】** ペインで、目的のプロファイルが格納されているプロジェクトまたはフォルダを選択するか、**【アセット】** ペインで目的のプロファイルを選択します。
2. **【アクション】** > **【開く】** をクリックします。  
サマリビューが**【検出】** ワークスペースに表示されます。
3. **【アクション】** > **【プロファイルの実行】** をクリックします。

Analyst ツールによってプロファイルが実行され、サマリビューにプロファイル結果が表示されます。プロファイルサマリログファイルおよびマッピングログファイルを表示して、Analyst ツールによって実行されたタスクの詳細を取得できます。

**注:** オペレーティングシステムプロファイルを使用してプロファイルを実行した場合、サマリログはデータ統合サービス用に構成されたログディレクトリにあり、マッピングログはオペレーティングシステムのプロファイル用に構成されたログディレクトリにあります。

## Informatica Analyst のデータドメイン検出結果

サマリビューと詳細ビューにデータドメインの検出結果を表示できます。

データドメインフィールドには、データドメインと一致するカラムに関する統計が表示されます。サマリビューでは、推測されたデータドメインと、適合する行の割合および数を一緒に参照できます。

詳細ビューでは、以下のタスクを実行できます。

- 推測されたデータドメインと、適合する行の割合および数を一緒に水平棒グラフに表示する。
- 適合する行、適合しない行、および NULL 値に結果をドリルダウンする。
- データドメインの承認、拒否、またはリセットを行う。
- 拒否されたデータドメインを表示または非表示にする。
- データソース内のすべての行に対してデータドメイン検出を実行し、推測されたデータドメインを検出する。

### データドメインの承認

Analyst ツールでは複数のデータドメインを承認することができます。

1. **【ライブラリ】** ワークスペースで、プロファイルを含むプロジェクトまたはフォルダを選択します。
2. プロファイルをクリックして開きます。  
サマリビューにプロファイル結果が表示されます。
3. データドメインを承認するカラムをクリックします。  
詳細ビューにカラム結果が表示されます。
4. 詳細ビューの**【データドメイン】** ペインでデータドメインを選択します。**【アクション】** > **【承認】** をクリックします。  
カラムまたはデータドメインのステータスが**【承認済み】** に変わります。

5. カラムまたはデータドメインのステータスを推測済みに戻すには、そのデータドメインを選択してから【アクション】 > 【リセット】をクリックします。

## データドメインの却下

プロファイル結果を開くと、デフォルトで Analyst ツールに承認されたデータドメインが表示されます。却下されたデータドメインを表示または非表示にすることができます。

1. ライブラリナビゲータで、プロファイルを含むプロジェクトまたはフォルダを選択します。
2. プロファイルをクリックして開きます。  
サマリビューにプロファイル結果が表示されます。
3. データドメインを拒否するカラムをクリックします。  
詳細ビューにカラム結果が表示されます。
4. 推測されたデータドメインを却下するには、【アクション】 > 【却下】をクリックします。  
Analyst ツールでは、却下されたデータドメインをデータドメイン検出の結果から除外します。
5. 却下されたデータドメインを表示するには、【アクション】 > 【却下されたものを表示】をクリックします。
6. 却下されたデータドメインを非表示にするには、【アクション】 > 【却下されたものを非表示】をクリックします。

# Informatica Analyst のデータドメイン検出エクスポートファイル

Analyst ツールからデータドメイン検出結果をエクスポートする際、ファイル名およびコードページの値を指定できます。データドメイン検出結果は Microsoft Excel ファイルにエクスポートできます。

Microsoft Excel ファイルでは、検出結果が、カラム、データドメイン、およびデータドメイングループ別に表示されている複数のシートに格納されます。【プロパティ】シートには名前、説明、タイプ、場所、プロファイル最終変更日時、およびプロファイルへのリンクなど、プロファイルのプロパティが表示されます。

## Microsoft Excel でのデータドメイン検出結果

データドメイン検出結果を Microsoft Excel にエクスポートすると、Analyst ツールはカラム名、一致するデータドメインの名前、適合条件、および NULL 値を保存します。このほかに Excel ファイルに含まれるのは、各データドメインのデータドメイングループ名、およびカラムの文書化されたデータ型です。

次の表に、エクスポートファイルの各シートを示します。

タブ	説明
カラム別に表示	データソースカラム別にソートされたデータドメイン検出結果
データドメイン別に表示	データドメイン別にソートされたデータドメイン検出結果

タブ	説明
データドメイングループ別に表示	データドメイングループ別にソートされたデータドメイン検出結果
プロパティ	名前、説明、タイプ、場所、プロファイル最終変更日時、ならびにプロファイルへのリンクなど、基本的なプロファイルのプロパティです。

## Informatica Analyst からのデータドメイン検出結果のエクスポート

データドメイン検出結果を .xlsx ファイルにエクスポートして、データをファイルで表示し、エンタープライズ内で配布して今後使用することができます。

1. プロファイルを実行してデータドメイン検出を実行する。
2. サマリビューまたは詳細ビューで、**[アクション]** > **[データのエクスポート]** をクリックします。  
**[ファイルへのデータのエクスポート]** ダイアログボックスが表示されます。
3. ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
4. ファイルのコードページを選択します。
5. **[OK]** をクリックします。

## 第 13 章

# Informatica Analyst でのエンタープライズ検出

この章では、以下の項目について説明します。

- [Informatica Analyst でのエンタープライズ検出の概要, 118 ページ](#)
- [Informatica Analyst でのエンタープライズ検出プロセス, 119 ページ](#)
- [エンタープライズ検出の設定オプション, 119 ページ](#)
- [Informatica Analyst でのエンタープライズ検出プロファイルの作成, 121 ページ](#)
- [エンタープライズ検出のオプションの編集, 122 ページ](#)

## Informatica Analyst でのエンタープライズ検出の概要

エンタープライズ検出とは、多くのスキーマや外部リレーショナル接続からの複数のデータソースの中でカラムのメタデータやデータドメインを検出するプロセスです。エンタープライズ検出の実行は、モデルリポジトリにインポートしたデータソースと、外部リレーショナル接続からのデータソースの両方を対象に行うことができます。

データ解析担当の方なら、Analyst ツールでエンタープライズ検出を実行して、大量のデータソースを対象に特定のメタデータの特性を推測することができます。事前定義したデータドメインとソースデータを確認するのもよいでしょう。それから、推測されたエンタープライズ検出の結果に対してキュレーションを行って、データを検出検索やデータ品質のイニシアティブで使えるようにすることができます。Analyst ツールのエンタープライズ検出では、プロファイル結果を 1 つのまとめた結果サマリが生成されます。

エンタープライズ検出結果には、パターン、一意の値、データ型が競合するカラムなど、カラムプロファイルの統計が含まれます。データドメイン検出では、事前定義したデータドメインと一致するソースカラムを特定します。

Informatica Analyst でオペレーティングシステムプロファイルを選択できます。オペレーティングシステムプロファイルを選択すると、オペレーティングシステムプロファイルに定義したオペレーティングシステムユーザーの権限に基づいて、データ統合サービスによってエンタープライズ検出プロファイルが作成および実行されます。

# Informatica Analyst でのエンタープライズ検出プロセス

エンタープライズ検出プロファイルの作成、編集、削除を行うことができます。[検出] ワークスペースでエンタープライズ検出プロファイルを実行できます。エンタープライズ検出プロファイルを実行する前に、カラムプロファイルとデータドメイン検出の推測オプションを設定する必要があります。

Analyst ツールでエンタープライズ検出を行うには、次の手順に従ってください。

1. エンタープライズ検出プロファイルの全般プロパティを設定します。
2. エンタープライズ検出プロファイルに含める必要のあるデータオブジェクトをモデルリポジトリから選択します。
3. 外部データベース接続からリレーショナルデータソースをインポートします。
4. エンタープライズ検出プロファイルに対するデータ推測オプションと検出オプションを設定します。
5. 変更を保存して、エンタープライズ検出プロファイルを実行します。
6. プロファイルの実行を監視します。また、Analyst ツールが実行するプロファイルタスクのステータスを必要に応じて確認します。
7. エンタープライズ検出結果のサマリを確認します。[サマリ] パネルと[プロファイル] パネルに結果が表示されます。

## エンタープライズ検出の設定オプション

エンタープライズ検出の設定オプションには、データドメイン検出オプション、カラムプロファイルのサンプリングオプション、および一般的プロファイルプロパティ（名前と説明など）があります。

カラムプロファイルまたはプロファイルの実行を選択して、データドメイン検出を実行できます。また、カラムプロファイルとプロファイルの両方の実行を選択して、データドメイン検出を設定の一部として実行できます。

### データドメイン検出の設定

データドメイン検出の設定では、データドメイン検出の実行対象をカラムデータにするか、カラム名にするか、またはその両方にするかを選択することもできます。データドメインを選択して、データドメイン検出でデータソース内の行をすべて処理する必要があるかどうかを指定できます。データドメイン検出の適合条件を選択できます。データドメイン検出から NULL を除外することが可能です。

Analyst ツールでエンタープライズ検出に関して指定できるデータドメイン検出の設定値を、次の表に示します。

オプション	説明
データドメイン検出の有効化	エンタープライズ検出の一部としてデータドメイン検出を実行します。
データに対してデータドメイン検出を実行	カラムデータに対してデータドメイン検出を実行します。

オプション	説明
カラム名に対してデータドメイン検出を実行	各カラムの名前に対してデータドメイン検出を実行します。
最小一致率	データドメインの一致に必要なデータセット内の行の最小一致率。一致率は、一致行数を合計行数で除算した値です。 注: Analyst ツールでは、NULL 値は一致しない行とみなされます。
適合する行の最小数	データドメインの一致に必要なデータセット内の行の最小数。
データドメイン検出から NULL 値を除外	データドメイン検出用のデータセットから NULL 値を除外します。
承認されたデータドメインのあるカラムを除外	プロファイル実行時のデータドメイン推測から、承認されたデータドメインのあるカラムを除外します。
すべての行	すべてのソース行に対してデータドメイン検出を実行します。
最初	プロファイルを実行できる最大行数。Analyst ツールでは、ソースの先頭行から行が選択されます。行は最大で 2,147,483,647 行まで選択できます。

## カラムプロファイル設定

このサンプリングオプションでは、Analyst ツールでカラムプロファイルをデータソースのすべての行に対して実行するか、または一部の行に対して実行するかを指定します。

以下の表に、エンタープライズ検出プロファイルに設定できるカラムプロファイル設定を示します。

オプション	説明
カラムプロファイリングを有効にする	カラムプロファイルをエンタープライズ検出の一部分として実行します。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

以下の表に、エンタープライズ検出プロファイルに設定できるランタイム環境オプションを示します。

オプション	説明
Native	Analyst ツールが、プロファイリングサービスモジュールにプロファイルジョブを送信します。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスがこれらのマッピングを実行し、プロファイル結果をプロファイリングウェアハウスに書き込みます。
Blaze	データ統合サービスがプロファイルのロジックを Hadoop クラスタの Blaze エンジンにプッシュして、プロファイルを実行します。
Spark	データ統合サービスがプロファイルのロジックを Hadoop クラスタの Spark エンジンにプッシュして、プロファイルを実行します。



以下の表に、エンタープライズ検出プロファイルに設定できるサンプリングオプションを示します。

オプション	説明
すべての行	データソースのすべての行に対してカラムプロファイルを実行します。 ネイティブ、Blaze、および Spark ランタイム環境でサポートされます。
最初の<number>行	データオブジェクト内の行のうち、先頭からのサンプル行でプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
Limit n <number> 行	データオブジェクトの行数に基づいてプロファイルを実行します。Hadoop 検証環境でプロファイルを実行するように選択すると、Spark エンジンではデータオブジェクトの複数のパーティションからサンプルを収集し、これらのサンプルを単一のノードにブッシュしてサンプルサイズを計算します。Limit n のサンプリングオプションは、Oracle、SQL Server および DB2 データベースをサポートします。Limit n のサンプリングオプションには、詳細フィルタは適用できません。行は最大で 2,147,483,647 行まで選択できます。 Spark ランタイム環境でサポートされます。
ランダムパーセンテージ	データオブジェクト内の行の割合に対してプロファイルを実行します。 Spark ランタイム環境でサポートされます。

## Informatica Analyst でのエンタープライズ検出プロファイルの作成

Informatica Analyst のエンタープライズ検出の一部として、カラムプロファイルとデータドメイン検出を実行することができます。

1. **【検出】** ワークスペースで、**【新規】** > **【プロファイル】** を選択します。  
**【新規プロファイル】** ウィザードが表示されます。
2. **【エンタープライズ検出】** を選択します。**【次へ】** をクリックします。  
**【全般プロパティの指定】** タブが表示されます。
3. **【全般プロパティの指定】** タブで、エンタープライズ検出プロファイルの名前と説明（省略可能）を入力します。**【場所】** フィールドで、プロファイルの作成先のプロジェクトまたはフォルダを選択します。**【次へ】** をクリックします。  
**【データオブジェクトの選択】** タブが表示されます。
4. **【データオブジェクトの選択】** タブで **【選択】** をクリックします。  
**【データオブジェクトの選択】** ダイアログボックスが表示されます。
5. **【データオブジェクトの選択】** ダイアログボックスで、1 つ以上のデータオブジェクトを選択してプロファイルに追加します。**【保存】** をクリックします。  
**【データオブジェクト】** ペインにデータオブジェクトが表示されます。
6. **【次へ】** をクリックします。  
**【リソースの選択】** タブが表示されます。
7. **【リソースの選択】** タブで **【選択】** をクリックし、**【リソースの選択】** タブを開きます。  
複数のリレーショナルデータソースからデータをインポートできます。

8. **【リソースの選択】** タブで、プロファイルに含める接続、スキーマ、テーブル、およびビューを選択します。**【保存】** をクリックします。

ダイアログボックスの左ペインには、Informatica ドメインの内部接続、外部接続、スキーマ、テーブル、ビューのすべてが一覧表示されます。

**【リソース】** ペインにリソースが表示されます。

9. **【次へ】** をクリックします。

**【設定の指定】** タブが表示されます。

10. **【設定の指定】** タブで、カラムプロファイルのオプションとデータドメイン検出のオプションを指定できます。**【保存して完了】** をクリックしてエンタープライズ検出プロファイルを保存するか、**【保存して実行】** をクリックしてプロファイルを実行します。

**【設定の指定】** タブで次のタスクを実行できます。

- データドメイン検出の有効化。**【選択】** をクリックして、**【データドメインの選択】** ダイアログボックスから検出するデータドメインを選択します。**【データドメイン検出のデータドメイン】** ペインに、選択したデータドメインが表示されます。
- データ、カラム名、またはデータとカラム名の両方を対象に、データドメイン検出を実行します。
- データソースのすべての行を選択するか、ドメイン検出を実行する行の最大数を選択します。
- データドメイン検出の最小一致率を選択するか、適合する行の最小数を指定します。
- カラムプロファイル設定を有効にし、カラムプロファイルのデータソースのすべての行または最初の数行を選択します。カラムプロファイルでデータ型が承認されているカラムについてのデータ型の推測を除外できます。
- ランタイム環境として、**【ネイティブ】**、**【Blaze】**、**【Spark】**、または **【Databricks】** を選択します。**【Blaze】** または **【Spark】** を選択した場合は、プロファイルを実行する Hadoop 接続を選択します。**【Databricks】** を選択した場合は、Databricks 接続を選択します。

**【サマリ】** タブと **【プロファイル】** タブで、エンタープライズ検出結果を表示できます。

## エンタープライズ検出のオプションの編集

エンタープライズ検出を実行後にエンタープライズ検出のオプションを変更することができます。プロファイル名を変え、データオブジェクトの選択、データドメインの選択、および推測オプションを変更することができます。

1. エンタープライズ検出を行うために実行するプロファイルを開きます。  
プロファイル結果が **【検出】** ワークスペースに表示されます。
2. バージョン管理システムが有効になっている場合、**【アクション】** > **【チェックアウト】** をクリックしてプロファイルをチェックアウトします。
3. **【プロファイルの編集】** をクリックします。
4. **【全般プロパティの指定】** タブで、必要に応じてプロファイルのプロパティを更新します。
5. データオブジェクトの選択に変更を加えるには、**【データオブジェクトの選択】** タブをクリックします。
6. エンタープライズ検出の外部データソースに変更を加えるには、**【リソースの選択】** タブをクリックします。
7. データドメイン推測オプションとカラムプロファイル設定に変更を加えるには、**【設定の指定】** タブをクリックします。

8. エンタープライズ検出プロファイル内のすべてのデータメインプロファイルタスクとカラムプロファイルタスクに設定変更を適用するには、**【すべてのプロファイルにグローバル設定を使用します】**を選択します。このオプションを選択しないと、プロファイル設定に加える変更はプロファイルに新たに追加するデータオブジェクトまたはリソースにのみ適用されます。

デフォルトでは、変更内容は、エンタープライズ検出プロファイルに新たに追加されるデータオブジェクトに適用されます。

9. 変更を取り消して元に戻すには、**【キャンセル】** をクリックします。
10. **【保存して実行】** をクリックし、変更を保存してプロファイルを再度実行します。
11. バージョン管理システムが有効な場合、次のタスクを実行する必要があります。
  - **【保存して完了】** をクリックしてプロファイルの編集を完了します。
  - [サマリ] ビューで、**【チェックイン】** をクリックしてプロファイルをチェックインします。
  - **【アクション】** > **【プロファイルの実行】** をクリックしてプロファイルを実行します。

## 第 14 章

# Informatica Analyst でのエンタープライズ検出の結果

この章では、以下の項目について説明します。

- [Informatica Analyst でのエンタープライズ検出結果の概要, 124 ページ](#)
- [サマリビュー, 124 ページ](#)
- [データ型の競合, 127 ページ](#)
- [プロファイルビュー, 127 ページ](#)

## Informatica Analyst でのエンタープライズ検出結果の概要

【サマリ】ビューと【プロファイル】ビューでエンタープライズ検出の結果を確認することができます。

【サマリ】ビューには、カラムプロファイルの結果とデータドメイン検出の結果が表示されます。【データドメイン検出】セクションに、プロファイルの実行に含めたデータドメインと、データドメインの一致があるカラムの数が一覧表示されます。【カラムプロファイリング】セクションにソースカラムに関する統計が表示されます。プロファイル結果の各行をクリックすると、【サマリ】ビューの右ペインに表示される詳細を確認できます。

## サマリビュー

【サマリ】ビューには、カラムプロファイルの結果とデータドメイン検出の結果に関するサマリが表示されます。データドメイン一致のあるカラムの数とカラムに一致があるデータドメイン名を確認することができます。カラムの統計には、上位 10 のパターン一致のカラムの数、すべての一意の値、すべての NULL 値が含まれます。

す。カラムの統計には、推測されたデータ型と記述されているデータ型の間にデータ型の競合があるカラムの数も表示されます。

## サマリビューのプロファイル結果

サマリビューでは、エンタープライズ検出の件名がデータドメイン検出セクションとカラムプロファイリングセクションに表示されます。

### データドメイン検出

次の表に、データドメイン検出の結果のカラムを示します。

カラム名	説明
名前	データドメインの名前。
対象カラム数	データドメインの一致があるカラムの総数。
プロファイル	一致するカラムを含むプロファイルの名前。
カラム名	一致するカラムの名前。
データ一致率	データドメインの一致に必要な行の最小一致率。
接続名	リレーショナルデータベース接続の名前。
ソース名	データソースの名前。
推測ステータス	データドメイン推測ステータス。ステータスは、 <b>[承認済み]</b> 、 <b>[拒否済み]</b> 、および <b>[推測済み]</b> です。
NULL の行の%	カラムの NULL 値の割合。
合計行数	行の合計数です。
適合する行	データドメインの一致に必要な行の最小数。
カラム名の一致	カラム名がデータドメイン名に一致するかどうかを示します。
文書化されたデータ型	プロファイルオブジェクト内のカラムに対して宣言されたデータ型。
検証済み	データソースのすべての行についてのデータドメインの一致の検証を示します。
最終実行時刻	最後にプロファイルを実行した日時。

## カラムプロファイリング

以下の表に、カラムプロファイル結果のカラムを示します。

カラム名	説明
名前	プロファイル結果タイプの名前（パターン、100% NULL、100%一意など）。
対象カラム数	プロファイル結果タイプが一致するカラムの総数。
プロファイル	一致するカラムを含むプロファイルの名前。
接続	リレーショナルデータベース接続の名前。
データソース	プロファイルのデータソース。
カラム数	プロファイル内で一致するプロファイル結果タイプのあるカラムの数。

## データドメイン検出結果の表示

データドメイン名をクリックすると、データドメイン検出の結果を表示することができます。データドメイン検出の結果から特定のプロファイルを開くことができます。

1. エンタープライズ検出を行うためのプロファイルを実行します。
2. **【サマリ】** ビューが表示されていることを確認します。
3. **【データドメイン検出】** セクションでデータドメインをクリックして、その検出結果を確認します。  
データドメインを含むプロファイルの一覧が、右側のペインに表示されます。
4. 必要に応じて右ペインの中の行を 1 つ選択します。  
プロファイルへのハイパーリンクが青で表示されます。
5. プロファイル名リンクまたはカラム名リンクをクリックしてプロファイルを開きます。  
プロファイルが開き、データドメイン検出の結果が表示されます。Analyst ツールが、結果の中のデータドメインを持つ行を強調表示します。必要があれば、検出検索など、さらに効率的な使用のためにプロファイル結果をキュレーションすることができます。
6. **【サマリ】** ビューに戻るには、**【エンタープライズ検出に戻る】** をクリックします。

## カラムプロファイルの結果の表示

**【サマリ】** ビューでエンタープライズ検出のカラムプロファイル結果を確認することができます。データドメイン検出の結果から特定のプロファイルを開くことができます。

1. エンタープライズ検出を行うためのプロファイルを実行します。
2. **【サマリ】** ビューが表示されていることを確認します。
3. 推測されたパターンの詳細を確認するには、**【カラムプロファイリング】** セクションで上位 10 パターンのうちの 1 つをクリックします。  
推測パターン結果を含むプロファイルの一覧が、右側のペインに表示されます。
4. すべての NULL 値、すべての一意の値、データ型の競合などの情報を確認するには、**【100% NULL】**、**【100%一意】**、または **【推測されたデータ型と記述されているデータ型の競合】** をクリックします。  
プロファイルの一致リストが右ペインに表示されます。

5. プロファイル名リンクまたはカラム名リンクをクリックしてプロファイルを開きます。  
プロファイルが開き、カラムプロファイルの結果が表示されます。
6. **【サマリ】** ビューに戻るには、**【エンタープライズ検出に戻る】** をクリックします。

## データ型の競合

エンタープライズ検出は、カラム内のデータ型の競合を特定します。データ型の競合とは、エンタープライズ検出を実行した後に、カラムのデータ型が推測と仕様で一致していない状態です。推測されたデータ型とは、Analyst ツールがカラムデータに基づいて導き出すデータソースカラムのデータ型です。文書化されたデータ型とは、ソースデータベースの中のカラムに対して宣言されているデータ型です。

エンタープライズ検出はカラムのデータ型をカラムデータに基づいて推測するため、カラムのデータ型の仕様とは異なるデータ型に推測される場合があります。例えば、仕様では文字列データ型のカラムを日付データ型であるとエンタープライズ検出が推測する可能性があります。データ型競合を確認し、そのカラムに最適なデータ型である日付を選択して承認することができます。

### データ型の競合の表示

データ型の競合があるプロファイルを **【サマリ】** ビューから開くと、Analyst ツールがデータ型の競合を赤で強調表示します。

1. エンタープライズ検出を行うためのプロファイルを実行します。
2. **【サマリ】** ビューが表示されていることを確認します。
3. **【カラムプロファイリング】** セクションで、**【推測されているデータ型と記述されているデータ型の競合】** をクリックしカラムプロファイルの結果にみられるデータ型の競合を表示します。  
データ型の競合があるカラムを含むプロファイルのリストが右ペインに表示されます。
4. 必要に応じて右ペインの中の行を 1 つ選択します。  
プロファイルへのハイパーリンクが青で表示されます。
5. プロファイル名リンクまたはカラム名リンクをクリックしてプロファイルを開きます。  
プロファイルが開き、データ型の競合が赤で表示されます。データ型の競合を解決する場合は、推測されたデータ型をここでキュレーションすることができます。
6. データ型をキュレーションするには、競合しているデータ型の行を選択し、**【データ型】** ビューをクリックします。
7. **【アクション】** をクリックしてから **【承認】** または **【却下】** を選択します。
8. **【サマリ】** ビューに戻るには、**【エンタープライズ検出に戻る】** をクリックします。

## プロファイルビュー

**【プロファイル】** ビューには、Analyst ツールでエンタープライズ検出の一部として実行される単一データオブジェクトプロファイルのすべてが一覧表示されます。プロファイルリストは各プロファイルの実行ステータス

も表示します。各プロファイルを開いてカラムプロファイルの結果やデータドメイン検出の結果を確認することができます。

## プロファイルプロパティの確認

**【プロファイル】** ビューでエンタープライズ検出の一部であるプロファイルの一覧を確認することができます。各プロファイルを開いて必要に応じてプロファイル結果をキュレーションすることができます。

1. エンタープライズ検出を行うためのプロファイルを実行します。
2. **【プロファイル】** ビューが表示されていることを確認します。
3. プロファイルのプロファイルプロパティを確認するには、プロファイルの名前をクリックします。  
プロファイルプロパティが右ペインに表示されます。プロファイルプロパティは、ソースデータオブジェクトの名前、接続名、および行数などで構成されています。
4. プロファイル結果を確認するには、**【プロファイルを開く】** をクリックします。  
プロファイルにカラムプロファイル結果が表示されます。
5. **【プロファイル】** ビューに戻るには、**【検出】** ワークスペースの左上隅でフォルダーまたはプロジェクトの名前のリンクをクリックします。



## 第 15 章

# Informatica Analyst での検出検索

この章では、以下の項目について説明します。

- [Informatica Analyst の概要での検出検索, 129 ページ](#)
- [検出検索の前提条件, 130 ページ](#)
- [Informatica Analyst での検出検索プロセス, 130 ページ](#)
- [検出検索オプション, 130 ページ](#)
- [Informatica Analyst での検出検索結果, 132 ページ](#)
- [一致タイプ, 134 ページ](#)
- [関連するアセット, 135 ページ](#)
- [FAQ \(よくある質問\) , 136 ページ](#)

## Informatica Analyst の概要での検出検索

検出検索はアセットを検索し、エンタープライズ内のデータベースおよびスキーマにある他のアセットとのリレーションを識別します。エンタープライズユーザーは、エンタープライズ内でデータとメタデータの存在する場所を検索するのに検出検索を使用できます。データオブジェクト、ルール、プロファイルなどの特定のアセットを検索できます。

グローバル検索を実行する場合、Analyst ツールはデータオブジェクト、データソース、およびフォルダに対し、テキストベースの検索を実行します。検出検索を実行する場合はテキストの一致に加え、検索基準に一致するオブジェクトへの関係を持つオブジェクトが検索結果に含まれます。検出検索には、データ型やデータパターンなどのプロファイルメタデータに基づく一致も含まれます。例えば、特定のデータパターンが含まれ、特定のキーワードを含む名前を持つオブジェクトを検索できます。

検出検索の検索結果には、次のタイプの情報が含まれます。

### モデルリポジトリのオブジェクト

検出検索の検索基準に一致するオブジェクトに関連するプライマリオブジェクトを検索します。例えば、プロファイルを検索する場合、プロファイル結果にはプロファイルのデータオブジェクトが含まれます。

### プロファイルウェアハウスの結果

データドメインやデータパターンなど、プロファイルからの推測結果が含まれます。

### Business Glossary の用語

ライセンスに基づき、ルールに関連付けられたビジネス用語などのメタデータが検索に含まれます。

### 検出検索の例

社内のデータスチュワードは、機密企業データが適切にマスキングされていることを確認する担当者です。データスチュワードは自分やデータアーキテクトがエンタープライズ検出を実行したスキーマとデータベース全

体にわたって、Personally Identifiable Information (PII) を特定することがあります。データスチュワードは、データソース内でまだ検出されていない重要データを特定するためのデータドメインを作成した可能性があります。データスチュワードは、「SSN」文字列の検索を実行できます。Analyst ツールにより、社会保障データドメイン、およびデータソース内で全カラムが表示されます。また、検出検索文字列によって、説明または名前に「SSN」が含まれるその他のカラムやテーブルも検出される可能性があります。検索を絞り込むには、マッピング仕様をフィルタリングすると、データオブジェクトを参照するマッピング仕様を表示できます。追加のフィルタを適用して、その他のマッピング仕様をプロジェクトやユーザーを基にフィルタリングできます。次に、マッピング仕様の結果を開いて、マッピング仕様が社内のプライバシーポリシーを満たしていることを確認することがあります。

## 検出検索の前提条件

社内のデータベース全体で効果的な検出検索を実行するには、社内のデータベースとスキーマに対し、エンタープライズ検出をあらかじめ実行します。

エンタープライズ検出を実行したら、Analyst ツールによってすべてのプロファイル結果がプロファイルウェアハウスに表示されます。すべての必須データソースがモデルリポジトリ内にあることを確認します。必要に応じて、モデルリポジトリ内の対応するアセットに、自身に関連付けられたビジネス用語があることを確認します。検出検索を実行すると、モデルリポジトリアセットとプロファイリングウェアハウスの結果を基に検索インデックス情報が検索サービスによって取得されます。検索サービスは、インデックス付き情報を使用して、対応するオブジェクトのメタデータとリレーションに基づく検索結果を表示します。

## Informatica Analyst での検出検索プロセス

プロファイル結果のテキスト、パターン、データ型などの条件を基にアセットを検索できます。検索によって、検索文字列に関連するアセットのリストが返されます。

Analyst ツールで検出検索を実行するには、次の手順に従ってください。

1. エンタープライズ検出を実行し、社内のデータソースに対して必要な単一データオブジェクトプロファイルを実行します。検出検索を実行すると、Analyst ツールによってプロファイル結果とモデルリポジトリオブジェクト内の情報が検索されます。
2. 検索する情報のタイプを選択します。例えば、機密データまたは特定のデータパターンのデータドメイン定義に関連したすべてのアセットを検索するとします。
3. 検索を実行します。
4. 検索結果を分析して、アセットおよび他のアセットとのリレーションを特定します。
5. 必要に応じて、検出データがビジネス要件に準拠していることを確認します。

## 検出検索オプション

グローバル検索または検出検索を実行して、アセットを見つけて、他のアセットへのリレーションを特定することができます。グローバル検索では、モデルリポジトリ、および Business Glossary（任意）から結果を取

得します。検出検索では、プロファイリングウェアハウスのプロファイル結果に基づくプロファイルに加えて、モデルリポジトリと Business Glossary から結果を取得します。

データオブジェクト、プロファイル、マッピング仕様などの特定のアセットを検索できます。検索文字列を入力して、その検索文字列に一致し、関連を持つアセットを検索します。アセットの検索時、ワイルドカード文字を使用できます。

アセットの検索時、次のワイルドカード文字を使用できます。

**\* (アスタリスク)**

検索文字列の末尾に追加して、その検索文字列で始まるすべてのアセット名を見つけます。例えば、「emp」文字列で始まる全てのアセット名を検索するには、検索フィールドに「emp\*」と入力します。

**? (疑問符)**

英数文字を表す文字を検索文字列に含めます。

**注:** アセットを検索する際、検索文字列をワイルドカードから始めることはできません。検索は、大文字小文字を区別しません。

2 つ以上のワイルドカードを同時に使用するフレーズで検索するには、単語を二重引用符で囲みます。+文字は AND 演算子を意味し、検索結果に必ず含まれる語句を示します。例えば、検索文字列が+sensitive +data である場合、検索サービスは両方の語句を含むメタデータを検索します。空白スペースは OR 演算子を意味します。例えば、検索文字列が sensitive data である場合、検索サービスはこの語句のいずれかを含むメタデータを検索します。

検索文字列に、ハイフン (-)、下線 (\_)、またはキャメルケースが含まれている場合、検索サービスによって、語全体または区切り文字で区切られた語の一部が検出されます。例えば、Profile\_Customer を検索する場合、検索エンジンはリポジトリ内で Profile、Customer や Profile\_Customer を見つけます。検索文字列に\*や?などの特殊文字を含める場合、これらの特殊文字を二重引用符で囲んで検索文字列に含めます。

検出検索にキーワード検索や検出フィルタを含めて実行することができます。例えば、データセキュリティのリスクを識別するために、<FirstNameInitial><LastNameInitial>-<SSN>という形式を使用する従業員 ID カラムを検索する必要があるとします。従業員 ID カラムを検索するには、[ライブラリ] ワークスペースの **【検索】** パネルで Employee ID と入力し、パターンフィルタに XX-999999999 <= 100%と設定します。

## 検出検索基準

パターン、データ型、一意の値、NULL 値などの基準に基づいた情報を検索するには、検出検索基準を使用します。検索では、=、>=、または<=などの条件演算子を使用できます。

次の表では、検出検索で利用できる検出検索条件について説明します。

オプション	説明
検索	検索対象のテキスト表現。
クリア	検索文字列および以前選択した他のすべての検索条件をクリアします。
パターン	検索に含めるカラムのパターンと割合。 <b>注:</b> このオプションでは、パターンに制御文字を使用できません。
データ型	検索に含めるカラムのデータ型と割合。
一意の値	検索に含めるカラムの一意の値の割合。
NULL	検索に含めるカラムの NULL 値の割合。

## アセットの選択

【ライブラリ】ワークスペース内でアセットを検索できます。検索結果には、Developer tool と Analyst ツールの両方で作成されたアセットが含まれます。

1. 【ライブラリ】ワークスペースを開きます。
2. 【検出結果】セクションが表示されていることを確認します。
3. 【検索】フィールドで、検索する検索文字列を入力します。
4. 検索フィルタを設定して検索を絞り込みます。  
フィルタには、パターン、データ型、一意の値、および NULL 値が含まれます。
5. 【検索】アイコンをクリックします。

## Informatica Analyst での検出検索結果

検出検索は、モデルリポジトリやプロファイリングウェアハウスなど、検出検索のためのすべてのライセンスされたリポジトリでアセットを検索します。

検出検索結果には、一致の総数と一致のリストが含まれます。それぞれの一致項目を展開すると、存在する場合は一致プロパティ、直接一致情報、間接一致情報、および関連するアセットの合計数を確認できます。直接一致は、検索クエリにアセットの一部またはすべてのメタデータのある一致を意味します。間接一致は、検索クエリに直接アセットにリンクされたアセットの一致です。

検索結果の表示順序は、次の要因によって決まります。

- 検索基準に一致するオブジェクトプロパティ。オブジェクトの名前は、オブジェクトの説明に優先します。オブジェクトの説明は、他のオブジェクトプロパティに優先します。
- オブジェクトのタイプ。データドメインとデータドメイングループの優先順位は他のオブジェクトより低くなります。
- キュレーション。キュレーションされたプロファイル結果の優先順位は、キュレーションしていないプロファイル結果より高くなります。
- 検索基準がオブジェクトと一致する回数には、直接一致と間接一致が含まれます。
- キーワードの相対頻度。

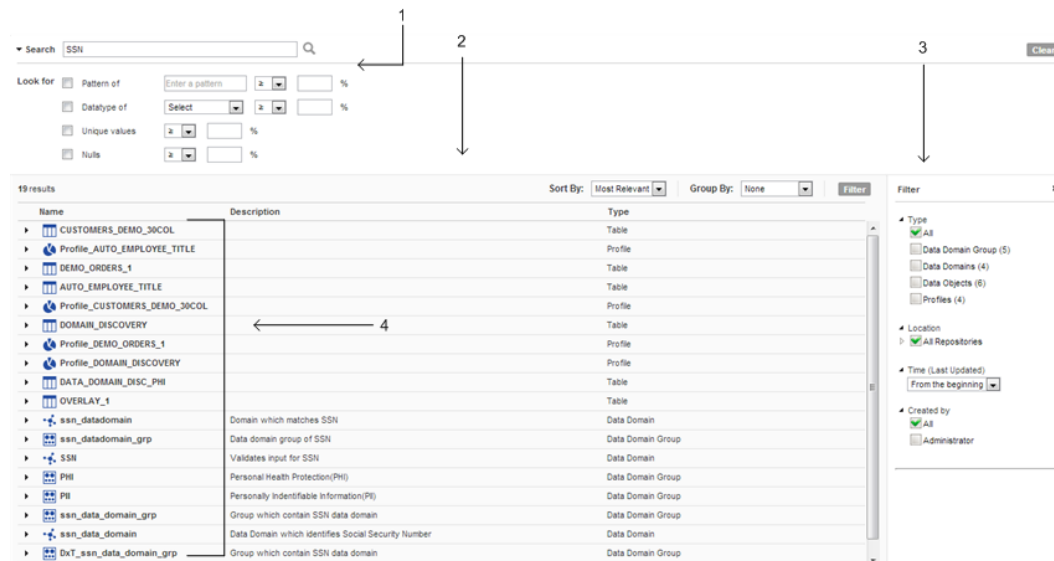
検索結果を表示するには、直接一致と間接一致を含むプロジェクトに対して、適切な権限を持つ必要があります。

### 【検出検索結果】 パネル

検出検索結果には、アセット名、アセットタイプ、およびアセットの説明が含まれます。フィルタを使用して検索結果を絞り込みます。

検索結果は、【ライブラリ】ワークスペースの結果グリッドに表示されます。関連性に基づいて結果をソートできます。また、アセットタイプ、リポジトリの場所、時間、アセットの作成者に基づいて結果をグループ化できます。

次の図に、検出検索結果のインターフェースを示します。



1. 検索基準
2. 結果グリッド
3. フィルタ
4. 検索結果

「検出検索結果」パネルには、デフォルトで次のセクションが表示されます。

#### 検索基準

プロファイルメタデータに基づく検索フィルタを含む検索フィールドを表示します。ここで検索を絞り込むための設定を行うことができます。検索フィールドは、Analyst ツール上部に表示されます。

#### 結果グリッド

検出基準フィールドで選択した基準に基づく一致の総数と一致のリストを表示します。結果グリッドには、オブジェクトの説明、オブジェクトタイプ、検索結果をソートおよびグループ化するドロップダウンも含まれています。

#### フィルタ

検索結果のフィルタリング用に設定できるフィルタが表示されます。【フィルタ】セクションは、Analyst ツールの右側の領域に表示されます。

#### 検索結果

検索文字列に基づく一致の検索結果（検索条件を満たすオブジェクトを含む）が表示されます。検索結果セクションには、一致プロパティである【直接一致】セクションと【間接一致】セクションが含まれ、これらは一致を展開すると表示されます。また、検索グリッドの一致に関連するアセットの合計数も表示できます。

## 検出検索結果のフィルタリング

また、アセットタイプ、リポジトリの場所、時間、アセットの作成者に基づいて検索結果をフィルタリングできます。Business Glossary をインストールしてある場合は、Business Glossary のビジネス用語、カテゴリ、ポリシーのアセット固有のフィルタも使用できます。

1. 【ライブラリ】ワークスペースの【検出検索】セクションで、グローバル検索または検出検索を実行します。

2. 結果グリッドの【フィルタ】をクリックして、【フィルタ】セクションを開きます。
3. 【フィルタ】セクションで、必要なフィルタと関連設定を選択します。
4. 選択したフィルタ設定に基づいて、変更された検索結果が結果グリッドに表示されます。
5. フィルタ設定を消去するには、結果グリッド上部の【すべてクリア】をクリックします。

## 一致タイプ

検出検索結果には、直接一致と間接一致が含まれます。直接一致は、検索クエリにアセットの一部またはすべてのメタデータのある一致を意味します。間接一致は、検索クエリに直接一致する別のアセットにリンクされたアセットの一致です。

検索クエリに複数の検索条件が含まれている場合、検索結果は検索条件を直接満たすか、間接的に満たすか、または直接と間接の両方で満たす可能性があります。検索結果から直接一致と間接一致の一部を読み取り専用または編集モードで開くことができます。

### 直接一致

直接一致は、検索クエリにアセットの一部またはすべてのメタデータのある一致を意味します。例えば、「Customer」という名前のすべてのアセットを検索すると、Analyst ツールによって、「Customer」という名前のデータオブジェクトとプロフィールが直接一致として表示されます。検出検索の実行後、表示される一致リストには、オブジェクトの一部へのリンクが含まれます。

検索結果のアセットを展開すると、アセットプロパティなど、直接一致について詳細を表示できます。

### 間接一致

間接一致とは、直接一致へリンクされた一致のことです。例えば、検索キーワードを含むルールを使用するスコアカードがあるとします。検出検索により、ルールが直接一致として、スコアカードが間接一致として返されます。スコアカードはルールを参照するため、間接一致となります。

間接一致情報を使用すると、オブジェクト間に隠された関係を識別し、オブジェクトの関係性をより良く理解することができます。また、間接一致の結果を使用すると検出検索がオブジェクトを返す理由を理解することもできます。

## 一致情報の表示

検出検索の実行後、直接一致と間接一致を含む一致情報を表示できます。また、アセットタイプ、説明、関連アセットなどのアセットプロパティも表示できます。検索結果からアセットの一部を開いて、必要に応じてこれらのアセットを変更できます。

1. 【ライブラリ】ワークスペースの【検索】セクションで、グローバル検索または検出検索を実行します。
2. 結果グリッドで、アセット名の先頭にある【展開】アイコンをクリックします。  
アセットプロパティと一致情報が、アセット名の下にセクションに表示されます。
3. 直接一致と間接一致の情報を確認します。  
アセットリレーション、および関連アセットの合計数といったその他の情報を表示できます。アセットリレーションには、直接一致と間接一致の両方が含まれます。
4. アセット情報にハイパーリンクが含まれている場合は、ハイパーリンクをクリックして、アセットを別のワークスペースで開きます。

5. [展開] アイコンを再びクリックして、一致情報セクションを閉じます。

## 検出検索結果からアセットを開く

検出検索結果からアセットを表示するには、必要なプロジェクト、アセット、およびライセンスの権限を持つ必要があります。

1. **【ライブラリ】** ワークスペースの **【検索】** セクションで、グローバル検索または検出検索を実行します。
2. 結果グリッドで、アセット名を右クリックします。  
ショートカットメニューが表示されます。
3. アセットをそのワークスペース内に、読み取り専用モードで表示するには、**【開く】** を選択します。
4. ワークスペース内でアセットに変更を加えるには、**【編集】** を選択します。
5. 検索結果からアセットを削除するには、**【削除】** を選択します。  
検索結果からアセットを削除すると、Analyst ツールによってそのアセットがモデルリポジトリから削除されます。
6. **【ライブラリ】** ワークスペースに戻るには、**【ライブラリ】** をクリックします。

## 関連するアセット

検索結果からアセットの関連アセットを表示できます。関連アセットは、検索結果で選択したアセットに関連付けられたモデルリポジトリまたは Business Glossary 内のアセットです。関連アセットは、一部のメタデータを検索結果のアセットと共有します。データソースは、プロファイル、推測されたデータドメイン、マッピングを関連アセットとして持つことができます。

例えば、プロファイルを検索結果の一部にすることができます。プロファイルのルールやデータソースなど、プロファイルの関連アセットを表示できます。関連するアセットは **【関連するアセット】** ワークスペースで表示できます。表示される関連アセットは、アセットタイプによって異なります。例えば、ルールの関連アセットを表示する場合、関連付けられたビジネス用語、マッピング仕様、プロファイルなどのアセットを表示できます。

## 各アセットタイプの関連アセット

Analyst ツールで表示されるアセットの関連アセットは、検索するアセットのタイプによって異なります。

次の表では、各アセットの関連アセットについて説明します。

アセットタイプ	関連するアセット
ビジネス用語	データドメイン、データオブジェクト、マプレット、およびルール
データドメイン	ビジネス用語、データドメイングループ、データオブジェクト、およびプロファイル。
データドメイングループ	データドメイン、データオブジェクト、およびプロファイル。
データオブジェクト	ビジネス用語、データドメイン、データドメイングループ、マッピング仕様、プロファイル、スコアカード、マッピング、およびマプレット。



アセットタイプ	関連するアセット
エンタープライズ検出プロファイル	データオブジェクトとプロファイル。
マッピング 注: このオブジェクトは Developer ツールで開きます。	データオブジェクト、マッピング仕様、マップレット、およびルール。
マッピング仕様	データオブジェクト、プロファイル、スコアカード、マッピング、マップレット、およびルール。
マップレット 注: このオブジェクトは Developer ツールで開きます。	ビジネス用語、データオブジェクト、マッピング仕様、マッピング、マップレット、およびルール。
プロファイル	データドメイン、データドメイングループ、データオブジェクト、ルール、およびマッピング仕様。 注: スコアカードは、プロファイルの関連アセットには含まれません。
ルール	データオブジェクト、ルール、ビジネス用語、マッピング仕様、プロファイル、スコアカード、およびマッピング。

## 関連アセットの表示

検索結果でアセット一致情報を表示するときに、関連アセットの合計数を表示できます。

1. **【ライブラリ】** ワークスペースで検出検索を実行します。
2. 結果グリッドで、**【展開】** アイコンをクリックし、関連アセット数のリンクをクリックするか、アセット名を右クリックして **【関連するアセットを表示】** を選択します。  
すべての関連アセットのリストが **【関連アセット】** ワークスペースに表示されます。
3. アセットの詳細を表示するには、アセット名をクリックするか、右クリックして **【開く】** を選択します。
4. 関連アセットの関連アセットを表示するには、アセット名を右クリックして、**【関連するアセットを表示】** を選択します。  
関連アセットの情報がワークスペースに表示されます。
5. 複数の関連アセットのワークスペース間を移動するには、**【関連アセット】** ワークスペースから、最近開いたアセットの1つを選択します。

## FAQ（よくある質問）

**表示されるはずの検索結果の一部が表示できないのはなぜですか。**

検索結果はさまざまな理由のために表示されないことがあります。検索条件が次のガイドラインを満たしていることを確認します。

- 検索結果に表示されるアセットは、プロジェクトの権限によって決まります。
- 検出検索結果には、カラムプロファイル結果からの値の頻度は含まれません。
- プロファイル結果のキュレーション時に拒否したプロファイル結果は、検索結果に含まれません。



- 表示される検索結果は、検索インデックスの抽出間隔および検索インデックスのアセットの可用性によって決まります。

**後で利用したり結果を他のユーザーと共有するために、検出検索結果を保存することはできますか？**

いいえ。検出検索結果を保存したり共有することはできません。

**検出検索結果の一部が検索結果の上部または下部に表示されるのはなぜですか。**

Analyst ツールに表示される検索結果の順序は、いくつかの要因に応じて異なります。この要因には、オブジェクトタイプ、キュレーションしたプロファイル結果、最初に検索基準に一致したオブジェクトプロパティ、各オブジェクトの内部検索ランクなどがあります。

**検出検索結果をエクスポートできますか？**

いいえ。検索結果をエクスポートすることはできません。

## 第 16 章

# Informatica Analyst での Business Glossary Desktop

この章では、以下の項目について説明します。

- [ビジネス用語, 138 ページ](#)
- [Metadata Manager ビジネス用語集でビジネス用語を管理, 139 ページ](#)
- [ビジネス用語集デスクトップでのビジネス用語のルックアップ, 139 ページ](#)

## ビジネス用語

ビジネス用語集デスクトップで、ビジネス用語をルックアップできます。Metadata Manager のライセンスに基づいて、ビジネス用語を表示したり、ビジネス用語タスクを実行したりできます。

ビジネス用語集とは、ビジネスユーザー向けに、ビジネス言語を使用して概念を定義した用語の集まりです。ビジネス用語には、概念のビジネス上の定義および使用方法を説明しています。

ビジネス用語集デスクトップは、ビジネス用語集をホストする Metadata Manager Service に接続するクライアントです。Analyst ツールオブジェクト名をルックアップできるようになる前に、あらかじめビジネス用語デスクトップを開いておく必要があります。ビジネス用語集デスクトップでビジネス用語としての Analyst ツールオブジェクト名の意味をルックアップすると、そのビジネス要件と現在の実装を理解できます。

Metadata Manager はビジネス用語集をホストしています。Metadata Manager ビジネス用語集を Analyst ツールから参照するには、Metadata Manager Service と Analyst Service と関連付ける必要があります。ビジネス用語集では、ビジネス用語や、カテゴリでグループ化したビジネス用語を表示できます。Metadata Manager のビジネス用語は編集が可能です。

Metadata Manager のビジネス用語で Metadata Manager リポジトリ内の Metadata Manager オブジェクトを検索できます。検索結果から Metadata Manager オブジェクトを選択し、Analyst ツールにデータオブジェクトとしてインポートすることができます。Metadata Manager のビジネス用語集に Metadata Manager のビジネス用語を追加することはできません。

# Metadata Manager ビジネス用語集でビジネス用語を管理

Analyst ツールから Metadata Manager ビジネス用語集にアクセスして、Metadata Manager のビジネス用語を管理することができます。

1. Analyst ツールのヘッダで、**[管理]** > **[用語の管理]** をクリックします。  
Metadata Manager と Metadata Manager ビジネス用語集が別のタブで開きます。Metadata Manager のビジネス用語は、Metadata Manager の **[用語集]** ビューに表示されます。
2. ビジネス用語集を選択するには、**[表示]** リストから用語集を選択します。
3. ビジネス用語をカテゴリ別にグループ化して表示するには、**[アクション]** > **[表示]** > **[カテゴリ]** をクリックします。
4. ビジネス用語集内のすべてのビジネス用語をアルファベット順に表示するには、**[アクション]** > **[表示]** > **[アルファベット]** をクリックします。
5. 特定の文字で始まるビジネス用語をすべて表示するには、その文字をクリックします。
6. ビジネス用語を編集するには、ビジネス用語を選択し、**[アクション]** > **[プロパティの編集]** をクリックします。

## ビジネス用語集デスクトップでのビジネス用語のルックアップ

そのビジネス要件と現在の実装を理解するには、ビジネス用語集デスクトップでビジネス用語としての Analyst ツールオブジェクト名をルックアップします。

ビジネス用語集デスクトップは、使用するマシンにあらかじめインストールしておく必要があります。

1. オブジェクトの名前をハイライト表示します。
2. ホットキーの組み合わせを使用し、ビジネス用語集デスクトップでビジネス用語としてのオブジェクト名をルックアップします。

デフォルトのホットキーの組み合わせは、SHIFT+ALT+Q です。

# パート III: Informatica Developer を使用したデータ検出

この部には、以下の章があります。

- [Informatica Developer のプロフィール, 141 ページ](#)
- [データオブジェクトプロフィール, 144 ページ](#)
- [半構造化データソースのカラムプロフィール, 159 ページ](#)
- [Informatica Developer のルール, 166 ページ](#)
- [マプレットとマッピングのプロファイリング, 168 ページ](#)
- [Informatica Developer のカラムプロフィール結果, 171 ページ](#)
- [Informatica Developer のスコアカード, 176 ページ](#)
- [Informatica Developer でのデータドメイン検出, 179 ページ](#)
- [Informatica Developer でのエンタープライズ検出, 191 ページ](#)
- [エンタープライズ検出結果, 208 ページ](#)
- [Informatica Developer での Business Glossary Desktop, 219 ページ](#)

## 第 17 章

# Informatica Developer のプロファイル

この章では、以下の項目について説明します。

- [Informatica Developer のプロファイルの概要, 141 ページ](#)
- [Informatica Developer のプロファイルのビュー, 142 ページ](#)
- [リポジトリオブジェクトのロックおよびバージョン管理されたオブジェクトを使用したチームベース開発, 143 ページ](#)

## Informatica Developer のプロファイルの概要

データセットのデータ品質に関する問題を検出したり、データセットのカラムのリレーションを把握したりするには、Informatica Developer でプロファイルを作成して実行します。

次のタイプのデータ分析のプロファイルを作成できます。

- カラムプロファイリング
- 半構造化データソースのカラムプロファイリング
- プライマリキー検出
- 機能依存性検出
- 外部キー検出
- 結合分析
- 重複検出
- データドメイン検出
- エンタープライズ検出

Developer tool ではウィザードを使用してプロファイルを作成します。プロファイル作成ウィザードでは、**[プロファイル]**、**[複数のプロファイル]**、および **[エンタープライズ検出プロファイル]** のオプションを使用してプロファイルを作成できます。

### プロファイル

単一のデータオブジェクトに対するプロファイルを作成します。単一のプロファイルでは、カラムプロファイリングのフィルタ、ルール、およびドリルダウンのオプションを定義します。カラムプロファイル、プライマリキープロファイル、機能依存性プロファイルを作成するため、およびデータドメイン検出用に詳細オプションを選択することもできます。結果には、カラムプロファイル、プライマリキー推測、機能依存性、およびデ

ータドメイン推測が表示されます。カラムプロファイルは、フラットファイルデータオブジェクト、リレーショナルデータオブジェクト、および半構造化データオブジェクトに対して作成できます。

### 複数のプロファイル

複数のオブジェクトに対する一連のプロファイルを作成します。Developer tool により、オブジェクトごとにプロファイルが作成されて同時に実行されます。複数のプロファイルを一度に作成した場合、複数のオブジェクトにまたがってデータを分析することはできません。

### エンタープライズ検出プロファイル

複数のデータオブジェクトからデータモデルを作成し、複数のオブジェクトにまたがってデータを分析するプロファイルを作成します。エンタープライズ検出プロファイルを作成して、一緒にプロファイリングする物理データオブジェクトをそのプロファイルに追加します。データオブジェクトプロファイル、外部キープロファイル、および結合プロファイルを作成できます。エンタープライズ検出プロファイルのデータオブジェクトごとに、全般的なプロパティ、プロファイリングするカラム、キー、および関係を設定することができます。1 つまたは複数のデータソース内の重複データを検出できます。

また、カラムプロファイル、データドメイン検出、プライマリキープロファイル、および外部キープロファイルなどのデータ検出タスクを作成および実行するエンタープライズ検出を実行することもできます。複数の接続にまたがる多数のデータソース上でエンタープライズ検出が実行されます。

以下の表に、各プロファイルタイプで実行できる操作を示します。

プロファイルオプション	プロファイルの操作
プロファイル	<ul style="list-style-type: none"><li>- カラムプロファイルを実行する。</li><li>- プライマリキーを検出する。</li><li>- 機能依存関係を検出する。</li><li>- データドメインを特定する</li></ul>
複数のプロファイル	複数のオブジェクトに対してカラムプロファイルを作成し、同時に実行する。
エンタープライズ検出プロファイル	<ul style="list-style-type: none"><li>- 単一のデータセットに対してカラムプロファイルを実行する</li><li>- プライマリキーを検出する。</li><li>- 外部キーを検出する。</li><li>- 機能依存関係を検出する。</li><li>- 結合分析を実行する。</li><li>- 2 つのカラム間の重複を検出する。</li><li>- エンタープライズ検出を実行する。</li></ul>

## Informatica Developer のプロファイルのビュー

Informatica Developer のプロファイルに関する情報は、**【概要】**、**【定義】**、**【コメント】**、および **【結果】** の各ビューを使用して表示したり追加したりできます。

**【Object Explorer】** ビューからプロファイルを開くと、右ペインのエディタの以下のビューにプロファイル情報が表示されます。

### 概要

名前、説明、場所など、プロファイルの全般的な情報が表示されます。

### 定義

プロファイル定義が表示され、設定することもできます。

この情報にはプロファイルに割り当てられるフィルタとルール、ドリルダウンオプション、プロファイルの実行中に有効になるプロファイル関数のリストが含まれます。

#### 結果

プロファイルの実行結果が表示されます。プロファイルの実行後、結果をエクスポートできます。

#### コメント

プロファイルに対するコメントが表示され、追加することもできます。

## リポジトリオブジェクトのロックおよびバージョン管理されたオブジェクトを使用したチームベース開発

モデルリポジトリでは、ユーザーが他のユーザーの作業を上書きしないように、プロファイルがロックされます。モデルリポジトリがバージョン管理システムと統合されている場合は、複数のバージョンのアセットが保存され、バージョンにバージョン番号が割り当てられます。プロファイルのチェックアウトとチェックイン、チェックアウトの取り消し、チェックアウトしたプロファイルの表示などが可能です。

Developer tool が予期せず停止した場合、モデルリポジトリはオブジェクトをロックした状態で保持します。モデルリポジトリに再接続すると、ロックしたオブジェクトを表示できます。オブジェクトの編集を継続するか、オブジェクトのロックを解除することができます。**【ロックされたオブジェクト】** ダイアログボックスで、ロックされたオブジェクトを表示してロックを解除できます。**【ロックされたオブジェクト】** ダイアログボックスを表示するには、**【表示】** > **【ロックされたオブジェクト】** をクリックします。

モデルリポジトリがバージョン管理システムと統合されている場合、Developer tool でバージョン管理されたオブジェクト管理を使用してオブジェクトのバージョンを管理できます。オブジェクトのチェックアウトとチェックイン、オブジェクトのバージョン履歴の表示と取得、チェックアウトの取り消しなどのアクションを実行できます。

モデルリポジトリでは、開発チームの他のメンバによって上書きされないようにオブジェクトが保護されます。別のユーザーによってすでにチェックアウトされているオブジェクトを開くと、そのオブジェクトをチェックアウトしたユーザーを特定する通知が送られてきます。チェックアウト済みのオブジェクトは、読み取り専用モードで開くことができるほか、別の名前でも保存することもできます。

オブジェクトが変更されると、モデルリポジトリはオブジェクトの新しいバージョンを作成します。

データ型、プライマリキー、外部キー、またはデータドメインを監督すると、モデルリポジトリはバージョン番号を 1 つ増やします。

バージョンをリストアすると、**【結果】** ビューにはリストアされたバージョンのプロファイル結果ではなく、最新のプロファイルの結果が表示されます。これは、バージョン管理システムがモデルリポジトリのすべてのバージョンのプロファイル定義を維持しており、プロファイル結果がプロファイリングウェアハウスから抽出されるためです。リポジトリオブジェクトのロックおよびバージョン管理されたオブジェクトの管理の詳細については、『Developer Tool ガイド』を参照してください。

## 第 18 章

# データオブジェクトプロファイル

この章では、以下の項目について説明します。

- [データオブジェクトプロファイルの概要, 144 ページ](#)
- [Informatica Developer のカラムプロファイル, 145 ページ](#)
- [ランタイム環境, 148 ページ](#)
- [プライマリキー検出, 149 ページ](#)
- [機能依存性検出, 151 ページ](#)
- [Informatica Developer のオペレーティングシステムプロファイル, 152 ページ](#)
- [Informatica Developer での単一のデータオブジェクトプロファイルの作成, 153 ページ](#)
- [Informatica Developer での複数のデータオブジェクトプロファイルの作成, 153 ページ](#)
- [プロファイルの編集, 154 ページ](#)
- [同期オプション, 155 ページ](#)
- [コメント, 158 ページ](#)

## データオブジェクトプロファイルの概要

データオブジェクトプロファイルでは、データソース内のカラムデータやメタデータに関する情報を検出します。Informatica Developer では、単一のデータオブジェクトおよび複数のデータオブジェクトに対してプロファイルを実行できます。単一のデータオブジェクトプロファイルでは 1 つのデータソースを分析し、複数のデータオブジェクトプロファイルでは複数のデータソースを分析します。複数のデータオブジェクトプロファイルを作成する場合、これらのプロファイルに対してカラムプロファイルを実行できます。

次の表に、単一のデータオブジェクトプロファイルに対して実行できるデータ検出タスクを示します。

タスク	説明
カラムプロファイリング	頻度、割合、パターンなどのデータの特性を検出します。フィルタを追加することで、プロファイルの実行時に条件を満たす行のみ読み取るようにすることができます。プロファイルでは、フィルタ条件を満たさない行は処理されません。
プライマリキー検出	データソース内で行を一意に識別できる値を含むカラムを検出します。



タスク	説明
機能依存性検出	データソース内のカラムのペア間の依存関係を検出します。
データドメイン検出	カラム値またはカラム名のいずれかに基づいて、カラムのすべてのデータドメインを特定します。

次の表に、[エンタープライズ検出プロファイル] オプションを使用してデータモデルを作成するときに、複数のデータオブジェクトに対して実行できるデータ検出タスクを示します。

タスク	説明
外部キー検出	別のデータソース内のプライマリキー値と一致する値を含むカラムを検出します。
結合分析	1 つのデータソース内、または 2 つのデータソース間における、2 つのカラム内のデータ間に見込まれる結合度を検出します。
重複検出	1 つまたは複数のデータソース内のカラムのペア間で発生している重複データの割合を検出します。
エンタープライズ検出	複数の接続またはスキーマにまたがる大量のデータソース内のカラムプロファイル統計、データドメイン、プライマリキー、および外部キーを検出します。

## Informatica Developer のカラムプロファイル

データソースのカラムの特性（値の割合やパターンなど）を分析するには、カラムプロファイルを使用します。フィルタを追加することで、プロファイルの実行時に条件を満たす行のみ読み取るようにすることができます。プロファイルでは、フィルタ条件を満たさない行は処理されません。

プロファイルのカラムに対して実行して検出できる情報は、次のタイプです。

- カラムの値が出現する回数。
- カラムの各値が出現する頻度（割合または行数で表される）。
- カラム内の値の文字パターン。
- 統計値（カラムの値の最大長と最小長、および最初の値と最後の値など）。
- 推測されるデータ型、頻度、データドメイン検出の適合条件、およびデータ型の推測ステータス。

カラムプロファイルは、マッピングまたはマップレット内のデータオブジェクトか、モデルリポジトリ内のオブジェクトについて定義できます。リポジトリ内のオブジェクトは、単一のデータオブジェクトプロファイル、複数のデータオブジェクトプロファイル、またはエンタープライズ検出プロファイルに含めることができます。

カラムプロファイルのサンプリングオプション、ドリルダウンオプション、およびランタイム環境を選択できます。カラムプロファイルにルールやフィルタを追加できます。

## フィルタリングのオプション

詳細フィルタまたは SQL フィルタを追加し、カラムプロファイルの実行時にカラムプロファイルで使用される行を特定できます。プロファイルでは、フィルタ条件を満たさない行は処理されません。

## 詳細フィルタの作成

AND、OR、NOTなどの式を使用して詳細フィルタを作成することで、元のデータソースのサブセットを作成できます。

1. 単一のデータオブジェクトプロファイルを作成するか開きます。
2. **【フィルタ】** ビューを選択します。
3. **【追加】** をクリックします。  
**【ウィザードの選択】** ダイアログボックスが表示されます。
4. **【ウィザードの選択】** ダイアログボックスで、**【詳細フィルタ】** をクリックします。  
**【フィルタ】** ダイアログボックスが表示されます。
5. 詳細フィルタの名前と説明（省略可能）を入力します。
6. **【アクティブとして設定】** を選択してフィルタをプロファイルに適用します。**【次へ】** をクリックします。
7. **【フィルタ定義】** を選択し、フィルタを定義します。
8. 詳細フィルタは、**【関数】** パネルまたは**【カラム】** パネルを使用して作成できます。
  - **【関数】** パネルでは、関数カテゴリを選択して、右矢印ボタン（>>）をクリックします。  
ダイアログボックスでパラメータを指定して、**【OK】** をクリックします。関数、およびカラムと値が、**【式】** パネルに表示されます。
  - **【カラム】** パネルでカラムを選択し、右矢印ボタン（>>）をクリックします。**【式】** パネルにカラムが表示されます。  
関数、式、値を追加して詳細フィルタを作成します。
9. 詳細フィルタを確認するには、**【検証】** をクリックします。
10. フィルタを作成または編集した後、**【データプレビュー】** を選択し、フィルタ済みのデータを表示します。  
**【プレビューする最大行数】** オプションを設定できます。
11. **【完了】** をクリックします。  
**【フィルタ】** ビューで、**【新しいプロファイル】** ウィザードにフィルタが表示されます。

## SQL フィルタの作成

SQL クエリを使用して SQL フィルタを作成できます。リレーショナルデータソースに対して SQL フィルタを作成できます。

1. 単一のデータオブジェクトプロファイルを作成するか開きます。
2. **【フィルタ】** ビューを選択します。
3. **【追加】** をクリックします。  
**【ウィザードの選択】** ダイアログボックスが表示されます。
4. **【ウィザードの選択】** ダイアログボックスで、**【SQL フィルタ】** をクリックします。  
**【フィルタ】** ダイアログボックスが表示されます。
5. 詳細フィルタの名前と説明（省略可能）を入力します。
6. **【アクティブとして設定】** を選択してフィルタをプロファイルに適用します。**【次へ】** をクリックします。
7. **【フィルタ定義】** を選択し、フィルタを定義します。
8. **【カラム】** パネルのカラムを使用し、SQL フィルタを作成します。
9. フィルタを確認するには、**【検証】** をクリックします。
10. フィルタを作成または編集した後、**【データプレビュー】** を選択し、フィルタ済みのデータを表示します。  
**【プレビューする最大行数】** オプションを設定できます。

11. **【完了】** をクリックします。

**【フィルタ】** ビューで、**【新しいプロファイル】** ウィザードにフィルタが表示されます。

## サンプリングのオプション

サンプリングオプションは、Developer tool がプロファイルを実行する行数を決定します。サンプリングのオプションは、プロファイルを定義または実行するときに設定することができます。

以下の表に、プロファイルのサンプリングのオプションを示します。

プロパティ	説明
すべての行	データオブジェクト内のすべての行でプロファイルを実行します。 ネイティブ、Blaze、Spark、および Databricks ランタイム環境でサポートされます。
最初の<number>行をサンプリング	データオブジェクト内の行のうち、先頭からのサンプル行でプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
<number>行のランダムサンプリング	データオブジェクトでランダムに選択された数の行に対してプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
ランダムサンプリング（自動）	データオブジェクト内の行数に基づいて計算されたサンプル行でプロファイルを実行します。 ネイティブおよび Blaze ランタイム環境でサポートされます。
Limit n <number>行	データオブジェクトの行数に基づいてプロファイルを実行します。 Hadoop 検証環境でプロファイルを実行するように選択すると、Spark エンジンではデータオブジェクトの複数のパーティションからサンプルを収集し、これらのサンプルを単一のノードにプッシュしてサンプルサイズを計算します。Limit n のサンプリングオプションは、Oracle、SQL Server および DB2 データベースをサポートします。Limit n のサンプリングオプションには、詳細フィルタは適用できません。 Spark ランタイム環境でサポートされます。
ランダムパーセンテージ	データオブジェクト内の行の割合に対してプロファイルを実行します。 Spark および Databricks ランタイム環境でサポートされます。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次回のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

ランダムな行のサンプルでプロファイルを実行するように選択すると、ランダムサンプルアルゴリズムでは、データオブジェクト内の行をランダムに選択して、プロファイルを実行します。カラムプロファイルのランダムサンプリングオプションを選択すると、Developer tool は、ステージングされたデータに対してドリルダウンを実行します。これは、ドリルダウンのパフォーマンスに影響を与える可能性があります。データドメイン検出プロファイルのランダムサンプリングオプションを選択すると、Developer tool はライブデータのドリルダウンを実行します。

# ランタイム環境

カラムプロファイルのランタイム環境として、[ネイティブ] または [Hadoop] を選択します。Hadoop ランタイム環境では、[Blaze] または [Spark] オプションを選択できます。Informatica Developer では、ランタイム環境の選択後に、プロファイル定義にランタイム環境が設定されます。

## ネイティブ環境

ネイティブのランタイム環境でプロファイルを実行すると、Developer tool がプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、データ統合サービスが実行する同じマッピング上でこれらのマッピングを実行し、プロファイル結果をプロファイリングウェアハウスに書き込みます。デフォルトでは、すべてのプロファイルがネイティブのランタイム環境で実行されます。

ネイティブソースを使用し、ネイティブ環境でプロファイルを作成して実行できます。ネイティブのデータソースは、フラットファイル、リレーショナルソース、メインフレームソースなど、Hadoop 以外のソースのことです。また、ネイティブ環境で Hive または HDFS データソースを使用してマッピング仕様または論理データソースに対してプロファイルを実行することもできます。

## Hadoop 環境

Hadoop ランタイム環境でプロファイルを実行する場合、[Blaze] または [Spark] オプションを選択できます。

[Blaze] オプションの選択後に、Hadoop 接続を選択できます。データ統合サービスがプロファイルのロジックを Hadoop クラスタの Blaze エンジンにプッシュして、プロファイルを実行します。

Hadoop 環境でプロファイルを実行すると、Developer tool がプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、これらのマッピングを Hadoop 接続で Blaze エンジンにプッシュします。Blaze エンジンはマッピングを処理し、データ統合サービスはプロファイル結果をプロファイリングウェアハウスに書き込みます。

[Spark] オプションの選択後に、Hadoop 接続を選択できます。データ統合サービスがプロファイルのロジックを Hadoop クラスタの Spark エンジンにプッシュして、プロファイルを実行します。Hadoop 環境でプロファイルを実行すると、Developer tool がプロファイルジョブをプロファイリングサービスモジュールにサブミットします。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスは、これらのマッピングを Hadoop 接続で Spark エンジンにプッシュします。Spark エンジンはマッピングを処理し、データ統合サービスはプロファイル結果をプロファイリングウェアハウスに書き込みます。

## Sqoop データソースのカラムプロファイル

Sqoop を使用するデータオブジェクトでカラムプロファイルを実行できます。Hadoop を検証環境として選択したら、Hadoop 接続で Blaze エンジンまたは Spark エンジンを選択して、カラムプロファイルを実行します。

論理データオブジェクトまたはカスタマイズデータオブジェクトのカラムプロファイルを実行すると、num-mappers 引数を設定して並行処理を実現しパフォーマンスを最適化することができます。また、Sqoop で作業単位を分割する際の基準のカラムを指定するように、split-by 引数も設定しなければなりません。

以下の構文を使用します。

```
--split-by <column_name>
```

プライマリキーの範囲の上限と下限の間で値の均等分布がない場合、データの均等分布がある別のカラムを指定して作業単位を分割するように、split-by 引数を設定できます。

split-by カラムを定義しない場合、Sqoop では次の条件に基づいて作業単位が分割されます。

- データオブジェクトに単一プライマリキーが含まれる場合、Sqoop はプライマリキーを split-by カラムとして使用します。
- データオブジェクトにコンポジットプライマリキーが含まれる場合、Sqoop のデフォルトは、split-by 引数を使用せずにコンポジットプライマリキーを処理する Sqoop の動作になります。詳細については、Sqoop のマニュアルを参照してください。
- データオブジェクトに同一のカラムを持つ 2 つのテーブルが含まれる場合、split-by カラムをテーブル修飾名で定義する必要があります。例えば、テーブル名が CUSTOMER であり、カラム名が FULL\_NAME の場合、split-by カラムを次のように定義します。  
--split-by CUSTOMER.FULL\_NAME
- データオブジェクトにプライマリキーがない場合、m 引数の値と num-mappers 引数は、デフォルトで 1 になります。

Cloudera Connector Powered by Teradata または Hortonworks Connector for Teradata を使用し、Teradata テーブルにプライマリキーがない場合、split-by 引数は必須です。

## プライマリキー検出

プライマリキー検出では、指定したカラムからプライマリキー候補が生成されます。

プライマリキーは、データソース内で行を一意に識別するカラムまたはカラムの組み合わせです。プライマリキー検出では、特定の信頼度を満たすカラムおよびカラムの組み合わせを特定します。信頼度、およびプライマリキーの識別のために組み合わせるカラムの最大数を編集できます。

プライマリキー検出では、プライマリキー候補における一意ではない行を特定することで、データ品質に関する潜在的な問題を強調表示することができます。これは特に、プライマリキー検出で多数のカラムを組み合わせている場合に役立ちます。不適合レコードには重複情報が含まれている可能性が高いためです。

## プライマリキー推測のプロパティ

単一のデータオブジェクトプロファイルを作成するときは、**【プライマリキーのプロファイリング】** ビューを使用して、プライマリキー推測のプロパティを設定できます。

以下の表に、**【プライマリキーのプロファイリング】** ビューのプライマリキー推測のプロパティを示します。

プロパティ	説明
デフォルトの推測オプションをオーバーライド	プライマリキー推測のカスタム設定を実行できる。
最大キーカラム数	プライマリキーを構成できる最大カラム数。
最大行数	プロファイリングする行数。
適合条件	プライマリキーを判別するときにプロファイルで許容できるキー違反の行の割合または数の最小値。

プロパティ	説明
文書化された、ユーザー定義のキーを持つデータオブジェクトの除外	文書化されたプライマリキー、またはユーザー定義のプライマリキーを持つデータオブジェクトを除外します。
承認されたキーを持つデータオブジェクトの除外	承認されたキーを持つデータオブジェクトを除外します。

## 推測されるプライマリキーのプロパティ

単一のデータオブジェクトプロファイルを実行した後、**【プライマリキーのプロファイリング】** ビューを使用して、データソース内の推測されるプライマリキーの詳細を確認できます。

以下の表に、**【プライマリキーのプロファイリング】** ビューの推測されるプライマリキーのプロパティを示します。

プロパティ	説明
カラム	プロファイル内のカラムの名前。
適合する行の%	カラム内の一意の値の割合。
重複する行の%	カラムの重複する値の割合。
NULL の行の%	カラムの NULL 値の割合。
検証済み	カラムがプライマリキーカラムかどうかを判別する。
推測ステータス	カラムの推測ステータス。
最終実行時刻	最後にプライマリキープロファイルを実行した日時。

## キー違反のプロパティ

単一のデータオブジェクトプロファイルを実行した後、**【プライマリキーのプロファイリング】** ビューを使用して、データソース内のプライマリキー違反の詳細を確認できます。

以下の表に、**【プライマリキーのプロファイリング】** ビューのキー違反のプロパティを示します。

プロパティ	説明
カラム	プロファイルでプライマリキー候補の推測に使用されるカラムの名前。
キー違反の数	プライマリキー候補におけるキー違反の数。

# 機能依存性検出

機能依存性検出を行うと、データソース内のカラムのペア間の依存性に関する情報が得られます。

一方のカラムの値が他方のカラムの値を確実に予測できる場合、そのカラムのペアは機能的に依存関係にあります。例えば、データセットに雇用者 ID カラムと生年月日カラムが含まれている場合、特定の雇用者 ID が含まれているすべての行で生年月日は同じになるはずです。

機能依存性検出では、カラムの機能依存性に適合しないレコードを特定することで、データ品質に関する潜在的な問題を強調表示することができます。例えば、データソース内の行の 99.8% が機能的に依存関係にある場合は、残りの行に正しくない情報が含まれている可能性が高くなります。

## 機能依存性推測のプロパティ

【機能依存性プロファイリング】ビューには、カラム間の機能依存性に関する情報が表示されます。

以下の表に、【機能依存性プロファイリング】ビューの機能依存性推測のプロパティを示します。

プロパティ	説明
デフォルトの推測オプションをオーバーライド	機能依存性推測のカスタム設定を実行できる。
決定側の最大カラム数	プロファイルで決定側を検出するために組み合わせることができるカラムの数。
最大行数	プロファイリングする行数。
返された依存関係数	プロファイルに表示される依存関係数。デフォルトは【最小範囲】で、各カラムが依存側または決定側に少なくとも 1 回出現する最小の依存性セットが表示される。
返された最大依存関係数	プロファイルに表示される最大依存関係数。
適合条件	機能依存側を判別するときにプロファイルで許容できる依存性違反の割合または数の最小値。

## 推測される機能依存性のプロパティ

単一のデータオブジェクトプロファイルを実行した後、【機能依存性推測】ビューを使用して、データソース内のカラム間の推測される機能依存性の詳細を確認できます。

以下の表に、【機能依存性推測】ビューの推測される機能依存性のプロパティを示します。

プロパティ	説明
決定側カラム	機能依存性があるかどうかを分析するカラムの名前。
依存カラム	決定側カラムに依存しているカラムの名前。
NULL の行の%	カラムの NULL 値の割合。
適合する行の%	機能依存性の一致の割合。



プロパティ	説明
検証済み	カラムが機能的に依存関係にあるかどうかを判別する。
最終実行時刻	最後に機能依存性プロファイルを実行した日時。

## 機能依存性違反のプロパティ

カラム間の機能依存性に関する情報がビューに表示されます。単一のデータオブジェクトプロファイルを実行した後、**【機能依存性推測】** ビューを使用して、データソース内の機能依存性違反の詳細を確認できます。

以下の表に、**【機能依存性推測】** ビューの機能依存性違反のプロパティを示します。

プロパティ	説明
決定側カラム	機能依存性があるかどうかを分析するカラムの名前。
別個の依存関係	一意の機能依存関係数。

# Informatica Developer のオペレーティングシステムプロファイル

Developer tool でオペレーティングシステムプロファイルを選択できます。オペレーティングシステムプロファイルを選択すると、データ統合サービスによってカラムプロファイルおよびエンタープライズ検出プロファイルが作成および実行され、スコアカードが作成されます。この処理は、オペレーティングシステムプロファイルユーザーの権限に基づいて行われます。

## オペレーティングシステムプロファイルの選択

Informatica Developer でオペレーティングシステムプロファイルを選択できます。データ統合サービスでは、プロファイリングジョブを実行するために、オペレーティングシステムプロファイルユーザーの権限を使用します。

1. Informatica Developer で **【Windows】** > **【設定】** をクリックします。  
**【設定】** ダイアログボックスが表示されます。
2. **【Informatica】** > **【実行設定】** > **【マッピング】** をクリックします。  
**【マッピング】** ダイアログボックスが表示されます。
3. **【マッピング】** ダイアログボックスで、**【デフォルトのデータ統合サービスを使用】** オプションの選択を解除します。
4. **【参照】** をクリックして、リストからオペレーティングシステムプロファイルを選択します。
5. **【OK】** をクリックします。



# Informatica Developer での単一のデータオブジェクトプロファイルの作成

データオブジェクトの 1 つ以上のカラムに対して単一のデータオブジェクトプロファイルを作成し、プロファイルオブジェクトをモデルリポジトリに格納することができます。

1. **【オブジェクトエクスプローラ】** ビューで、プロファイリングするデータオブジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【プロファイル】** をクリックしてプロファイルウィザードを開きます。
3. **【プロファイル】** を選択して **【次へ】** をクリックします。
4. プロファイルの名前を入力し、プロジェクトの場所を確認します。必要に応じて、新しい場所を参照します。
5. 必要に応じて、プロファイルの説明を入力します。
6. 選択したデータオブジェクトの名前が **【データオブジェクト】** セクション内に表示されることを確認します。
7. **【次へ】** をクリックします。
8. 実行するプロファイルの操作を P 設定します。以下の操作を設定できます。
  - カラムプロファイリング
  - プライマリキー検出
  - 機能依存性検出
  - データドメイン検出

**注:** プロファイルの操作を有効にするには、その操作に対して **["プロファイルの実行"アクションの一部として有効化]** を選択します。カラムプロファイリングはデフォルトで有効になっています。

9. プロファイルのオプションを確認します。

すべてのプロファイルタイプのカラム選択を編集できます。カラムプロファイルのフィルタおよびサンプリングのオプションを確認します。プライマリキー、機能依存性、およびデータドメイン検出の推測オプションを確認できます。また、データドメイン検出のデータドメイン選択も確認できます。
10. ドリルダウンのオプションを確認し、必要に応じて編集します。デフォルトでは、**【行のドリルダウンを有効にする】** オプションが選択されています。カラムプロファイルでは、ドリルダウンのオプションを編集できます。このオプションでは、ドリルダウン操作の読み取り元をデータソースにするか、またはスレージングされたデータにするか、および以前のプロファイルの実行による結果データを格納するかも決定します。
11. **【実行の設定】** セクションで、検証環境を選択します。検証環境として、**【ネイティブ】**、**【Hadoop】**、または **【Databricks】** を選択します。ランタイム環境として **【ネイティブ】**、**【Blaze】**、または **【Spark】** を選択できます。**【Blaze】** または **【Spark】** を選択すると、Hadoop 接続を選択できます。**【Databricks】** を選択した場合は、Databricks 接続を選択できます。
12. **【完了】** をクリックします。

# Informatica Developer での複数のデータオブジェクトプロファイルの作成

複数のデータオブジェクトに対して複数のデータオブジェクトプロファイルを実行する場合、Developer tool はデフォルトのカラムプロファイリングオプションを使用して、1 つ以上のデータオブジェクトのカラムプロ

ファイルを生成します。必要に応じて、複数のデータオブジェクトに対してプロファイルを実行するためのエンタープライズ検出プロファイルを作成することもできます。

1. **【オブジェクトエクスプローラ】**ビューで、プロファイリングするデータオブジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【プロファイル】** をクリックして **【新しいプロファイル】** ウィザードを開きます。
3. **【新規】** ウィザードで **【複数のプロファイル】** オプションを選択して、**【次へ】** をクリックします。
4. **【複数のプロファイル】** ウィンドウで、プロファイルを作成する場所を選択します。プロファイリングしたオブジェクトと同じ場所に各プロファイルを作成するか、またはプロファイルの共通の場所を指定することができます。
5. 選択したデータオブジェクトの名前が **【データオブジェクト】** セクション内に表示されることを確認します。

必要に応じて、**【追加】** をクリックして別のデータオブジェクトを追加します。

6. 必要に応じて、プロファイリングする行数を指定し、ウィザードの完了時にプロファイルを実行するかどうかを選択します。
7. **【次へ】** をクリックします。
8. **【検証環境】** セクションで、**【ネイティブ】** を選択します。

**注:** 複数のデータオブジェクトプロファイルを実行するには、必ず **【ネイティブ】** オプションを選択します。Hadoop ランタイム環境の Blaze または Spark エンジンで複数のデータオブジェクトを実行するには、エンタープライズデータ検出プロファイルを選択します。Databricks ランタイム環境の Databricks Spark で複数のデータオブジェクトを実行するには、エンタープライズデータ検出プロファイルを選択します。

9. **【完了】** をクリックします。
10. 必要に応じて、プロファイル名に追加するプレフィックスおよびサフィックス文字列を入力します。
11. **【OK】** をクリックします。

## プロファイルの編集

単一のデータオブジェクトプロファイル、または複数のデータオブジェクトプロファイルを編集できます。バージョン管理システムが有効な場合、プロファイルはデフォルトでチェックアウトされます。

1. **【オブジェクトエクスプローラ】** ビューでプロファイルを右クリックして、**【開く】** をクリックします。  
**【結果】** ビューが表示されます。
2. **【定義】** ビューで、必要に応じてプロパティを更新します。
3. **【チーム】** > **【チェックイン】** をクリックしてプロファイルをチェックインします。
4. プロファイルを右クリックし、**【プロファイルの実行】** をクリックしてプロファイルを実行します。  
**【結果】** ビューにプロファイルの結果が表示されます。

# 同期オプション

外部データソースのメタデータを変更する場合、モデルリポジトリ内のデータオブジェクトのメタデータはデフォルトでは更新されません。データオブジェクトのメタデータをデータソースのメタデータと同期させるには、[同期] オプションを使用します。

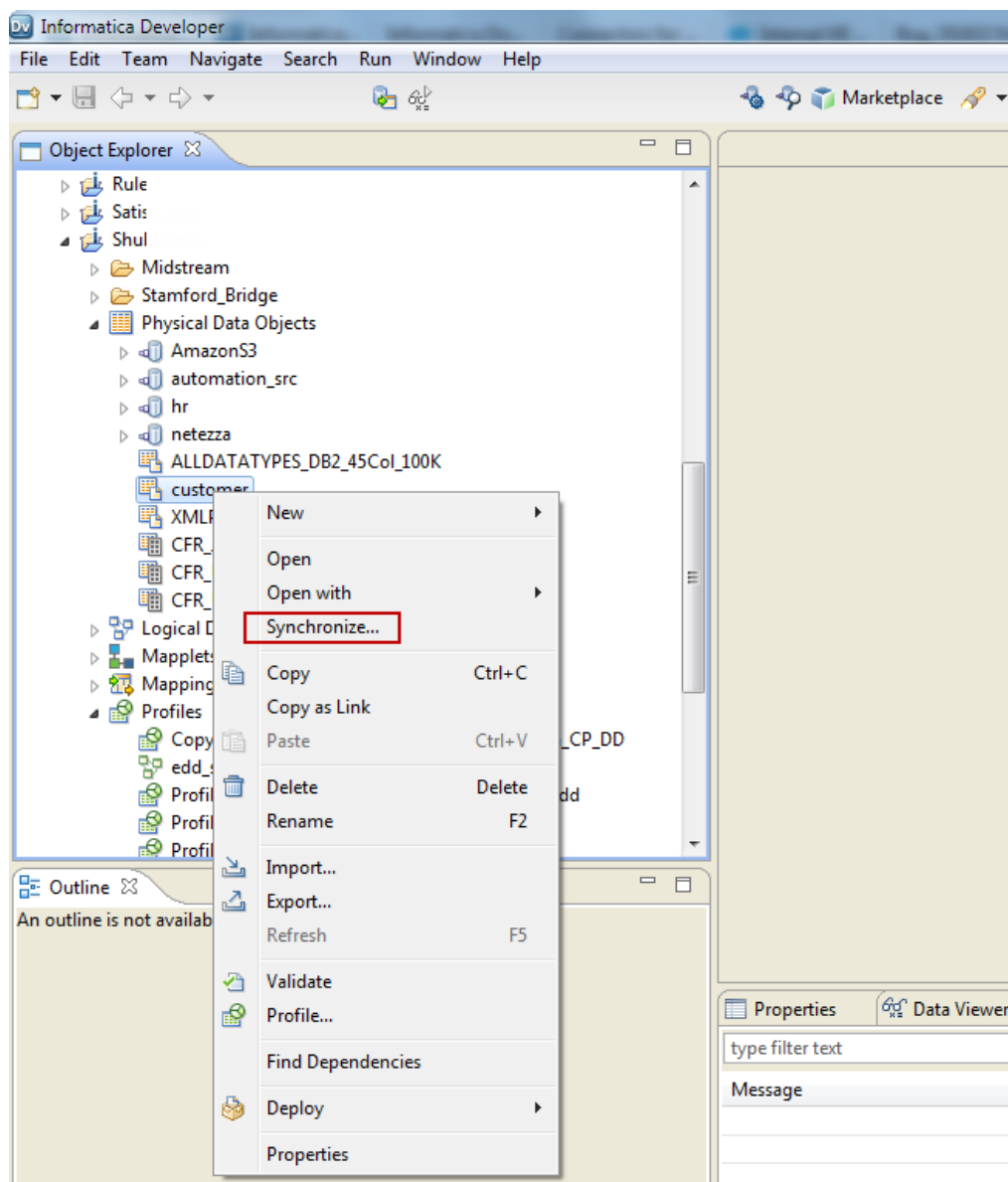
Developer tool で、[同期] オプションを使用した後、データオブジェクトを使用するプロファイルまたはスコアカードを開くと、エディタのプロファイル名またはスコアカード名の横にアスタリスクが表示されます。アスタリスクは、データオブジェクトのメタデータがプロファイルまたはスコアカードに対して変更されたことを示します。プロファイルまたはスコアカードを開いて保存し、モデルリポジトリのプロファイル定義を更新します。プロファイルまたはスコアカードのデータオブジェクトを同期した後に、プロファイルまたはスコアカードを開いたときに、Analyst ツールに目に見える変更内容が示されないことに注意してください。[同期] オプションは、カラムプロファイル、エンタープライズ検出プロファイル、およびスコアカードに使用できます。外部データソースは、リレーショナルデータソースまたはフラットファイルデータソースにすることができます。

## Informatica Developer でのフラットファイルデータオブジェクトの同期

Informatica Developer で、外部フラットファイルデータソースに加えられた変更をそのデータオブジェクトに同期させることができます。データオブジェクトを同期させるには、[フラットファイルの同期] ウィザードを使用します。

1. **【オブジェクトエクスプローラ】** ビューで、フラットファイルデータオブジェクトを選択します。
2. 右クリックし、**【同期】** を選択します。

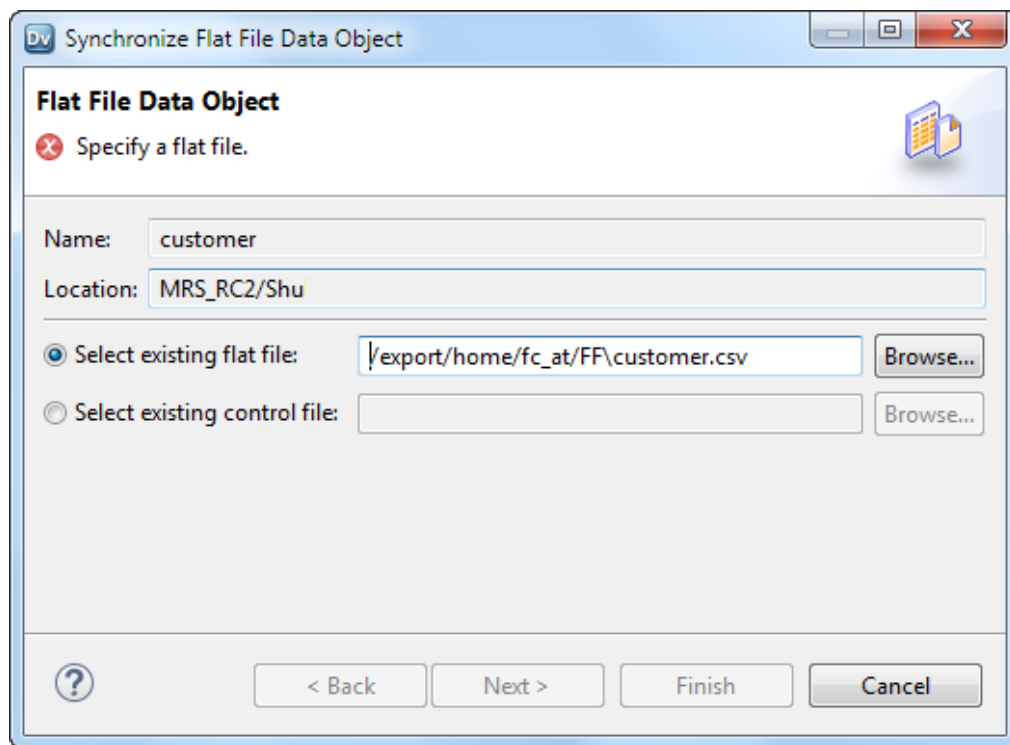
次の図は、データオブジェクトの同期オプションを示しています。



【フラットファイルデータオブジェクトの同期】ウィザードが表示されます。

3. 【フラットファイルデータオブジェクトの同期】ウィザードで、【既存のフラットファイルを選択】フィールドのフラットファイルパスを確認します。

次の図は、フラットファイルデータオブジェクトの同期ウィザードを示しています。



4. **【次へ】** をクリックします。
5. 必要に応じ、コードページ、形式、区切り形式プロパティ、およびカラムプロパティを選択します。
6. **【完了】** をクリックし、**【OK】** をクリックします。

## Informatica Developer でのリレーショナルデータオブジェクトの同期

Informatica Developer で、リレーショナルデータソースの外部データソース変更をそのデータオブジェクトに同期させることができます。外部データソース変更には、カラムの追加、変更、削除や、ルールの変更などがあります。

1. **【Object Explorer】** ビューで、リレーショナルデータオブジェクトを選択します。
2. 右クリックし、**【同期】** を選択します。  
操作の確認を求めるメッセージが表示されます。
3. 同期プロセスを実行するには、**【OK】** をクリックします。  
同期プロセスのステータスメッセージが表示されます。
4. 「**同期が完了しました**」というメッセージが表示された時点で、**【OK】** をクリックします。  
このメッセージには、データオブジェクトに加えられたメタデータ変更の概要が示されます。

# コメント

プロファイルにはコメントとして説明を追加できます。カラムプロファイル結果のカラムにコメントを追加することもできます。

カラムには複数のコメントを追加できます。Developer tool の【コメント】ビューを使用してコメントを追加および表示できます。

## Informatica Developer でのコメントの追加

カラムプロファイル結果のカラムにコメントを追加します。プロファイル結果をエクスポートする際には、Developer tool によってコメントが取り込まれます。

1. 【Object Explorer】ビューで、プロファイルを開きます。
2. 必要に応じて、プロファイルを実行してプロファイル結果を更新します。
3. 【コメント】ビューを選択します。
4. 【追加】をクリックして【コメントの追加】ダイアログボックスを開きます。
5. リストから、プロファイル名またはいずれかのカラムを選択します。  
以前追加したコメントは、ダイアログボックス内で確認できます。
6. 【コメント】フィールドに説明を入力します。
7. 【OK】をクリックします。  
Developer tool によって【コメント】ビューにコメントが表示されます。
8. コメントを削除するには、【コメント】ビューでコメントを選択し、【削除】をクリックします。

## 第 19 章

# 半構造化データソースのカラムプロファイル

この章では、以下の項目について説明します。

- [半構造化データソースのカラムプロファイルの概要, 159 ページ](#)
- [JSON および XML データオブジェクト, 160 ページ](#)
- [HDFS 内の半構造化データソースの複合型ファイルデータオブジェクト, 161 ページ](#)
- [HDFS 接続の作成, 162 ページ](#)
- [HDFS 内の JSON または XML ファイルからの複合型ファイルデータオブジェクトの作成, 162 ページ](#)
- [Avro または Parquet データソースからの複合型ファイルデータオブジェクトの作成, 163 ページ](#)
- [半構造化データソースでのカラムプロファイルの作成, 164 ページ](#)

## 半構造化データソースのカラムプロファイルの概要

Avro、JSON、Parquet、および XML データソースからデータオブジェクトを作成し、次に、そのデータオブジェクトのカラムプロファイルを作成できます。

Avro、JSON、Parquet、および XML の各形式は、半構造化データソースです。カラムプロファイルの作成に半構造化データソースを使用するために、次のタスクを実行できます。

1. 半構造化データソースに対して、物理データオブジェクトを作成する。
2. 物理データオブジェクトでカラムプロファイルを作成して実行する。

JSON または XML データソースの場合は、フラットファイルデータオブジェクトを作成できます。Hadoop 分散ファイルシステム（HDFS）内の Avro、JSON、Parquet、および XML データソースの場合は、複合型ファイルデータオブジェクトを作成できます。

# JSON および XML データオブジェクト

JSON または XML データソースから、フラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトを作成できます。データオブジェクトにカラムプロファイルを作成して実行できます。

JSON または XML データソースのパスが含まれるテキストファイルを作成して、そのファイルをデータソースとして使用して、フラットファイルデータオブジェクトを作成します。このテキストファイルに、複数の JSON または複数の XML データソースへのファイルパスを追加することもできます。

複合ファイルリーダーを使用すると、JSON または XML データソースから複合型ファイルデータオブジェクトを作成できます。複合ファイルリーダーは、データプロセッサトランスフォーメーションに入力を渡します。このトランスフォーメーションにより、ファイルが解析され、ソースデータがフラットなカンマ区切り値のレコードに変換されます。

**注:** Developer tool は、UTF-8 エンコードを使用した JSON データソースには対応していません。

## JSON または XML データソースからのデータオブジェクトの作成

JSON または XML データソースから、フラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトを作成できます。

1. Developer tool の **【オブジェクトエクスプローラ】** ビューで、データオブジェクトおよびカラムプロファイルを作成するプロジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【データオブジェクト】** をクリックします。  
**【新規】** ダイアログボックスが表示されます。
3. フラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトの作成を選択できます。
  - フラットファイルデータオブジェクトを作成するには、次のタスクを実行します。
    1. **【物理データオブジェクト】 > 【フラットファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。  
**【新しいフラットファイルデータオブジェクト】** ダイアログボックスが表示されます。
    2. **【既存のフラットファイルから作成】** を選択し、**【参照】** をクリックしてテキストファイルを選択します。**【次へ】** をクリックします。
    3. コードページが **【MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセット】** であり、形式が **【区切り】** であることを確認します。**【次へ】** をクリックします。
    4. 区切り文字が **【カンマ】** に設定されていることを確認します。**【完了】** をクリックします。
  - 複合型ファイルデータオブジェクトを作成するには、次のタスクを実行します。
    1. **【物理データオブジェクト】 > 【複合型ファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。  
**【新しい複合型ファイルデータオブジェクト】** ダイアログボックスが表示されます。
    2. データオブジェクトの名前を入力します。アクセスタイプとして **【ファイル】** を選択します。
    3. **【参照】** をクリックして JSON または XML ファイルを選択します。 **【完了】** をクリックします。  
Developer サーバーが Linux にある場合、データソースのファイルパスをそのサーバー上の場所に更新する必要があります。ファイルパスを更新するには、複合型ファイルデータオブジェクトを選択して **【データオブジェクト操作】** タブの **【読み取り】** をクリックし、**【データオブジェクト操作の詳細】** ペインの **【詳細】** タブでファイルパスを追加します。

データオブジェクトがプロジェクトフォルダに表示されます。



# HDFS 内の半構造化データソースの複合型ファイルデータオブジェクト

HDFS を使用する Avro、JSON、Parquet、または XML ファイルに対して、カラムプロファイルを作成して実行できます。HDFS 内の JSON または XML ファイルを読み取るには、JSON または XML 形式の入力をデータプロセッサトランスフォーメーションに渡すための複合型ファイルリーダーを使用します。

## HDFS 内の JSON または XML データソースからの複合型ファイルデータオブジェクト

JSON または XML ファイルから、複合型ファイルデータオブジェクトを作成できます。このデータオブジェクトに対して、カラムプロファイルを作成して実行できます。

HDFS 内の JSON または XML ファイルに対してデータオブジェクトを作成する前に、HDFS への接続を作成します。

次のいずれかの方法を使用して、HDFS 内の JSON または XML ファイルからデータオブジェクトを作成できます。

- 1 つの JSON または XML ファイルに対して、複合型ファイルデータオブジェクトを作成する。
- 複数の JSON または複数の XML ファイルが格納されたフォルダに対して、複合型ファイルデータオブジェクトを作成する。

データオブジェクトを作成したら、そのデータオブジェクトに対してカラムプロファイルを作成および実行できます。

## HDFS 内の Avro または Parquet データソースからの複合型ファイルデータオブジェクト

HDFS 内の Avro または Parquet データソースから、複合型ファイルデータオブジェクトを作成できます。このデータオブジェクトを使用して、カラムプロファイルを作成して実行できます。

複合型ファイルデータオブジェクトの作成は、1 つの Avro または Parquet ファイルから、または複数の Avro ファイルや複数の Parquet ファイルが含まれるフォルダに対して行うことができます。Avro および Parquet データソースからの複合型ファイルデータオブジェクトは、ファイルまたは接続のアクセスタイプで、Binary、Avro、または Parquet のリソース形式で作成できます。Avro および Parquet データソースから複合型ファイルデータオブジェクトを作成する前に、HDFS 接続を作成しておく必要があります。

**注:** リソース形式として **[Avro]** または **[Parquet]** を選択できるのは、フラットな構造の Avro および Parquet データソースの場合のみです。

HDFS 内の Avro および Parquet ファイルからデータオブジェクトを作成する場合は、次のオプションのいずれかを選択できます。

- アクセスタイプをファイルにし、リソース形式を Binary として選択する。
- アクセスタイプをファイルにし、リソース形式を Avro または Parquet として選択する。
- アクセスタイプを接続にし、リソース形式を Avro または Parquet として選択する。

## HDFS 接続の作成

Informatica Developer で HDFS 接続を設定して、HDFS 内の Avro、JSON、Parquet、および XML データソースに対してカラムプロファイルを作成します。HDFS 接続の作成後に、複合型ファイルデータオブジェクトを作成できます。

1. **【ウィンドウ】** > **【設定】** をクリックします。
2. **【Informatica】** > **【接続】** を選択します。
3. ドメインを展開します。
4. 接続タイプを **【ファイルシステム】** > **【Hadoop ファイルシステム】** で選択し、**【追加】** をクリックします。
5. 接続名を入力します。
6. 必要に応じて、接続の説明を入力します。
7. **【次へ】** をクリックします。
8. 接続のプロパティを設定します。
9. **【テスト接続】** をクリックし、HDFS への接続を確認します。
10. **【完了】** をクリックします。

## HDFS 内の JSON または XML ファイルからの複合型ファイルデータオブジェクトの作成

HDFS を使用した JSON または XML ソースファイルの複合型ファイルデータオブジェクトを作成し、そのデータオブジェクトに対するカラムプロファイルを作成できます。

1. Developer tool の **【オブジェクトエクスプローラ】** ビューで、物理データオブジェクトおよびカラムプロファイルを作成するプロジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【データオブジェクト】** をクリックします。  
**【新規】** ダイアログボックスが表示されます。
3. **【物理データオブジェクト】** > **【複合型ファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。  
**【新しい複合型ファイルデータオブジェクト】** ダイアログボックスが表示されます。
4. データオブジェクトの名前を入力します。アクセスタイプとして **【接続】** を選択します。
5. 1 つの JSON または XML ファイルから、または複数の JSON ファイルまたは複数の XML ファイルが含まれるフォルダに対して、データオブジェクトを作成できます。
  - JSON または XML ファイルから複合型ファイルデータオブジェクトを作成するには、次の手順を実行します。
    1. **【参照】** をクリックして接続を選択します。
    2. **【リソースの追加】** ダイアログボックスで、**【追加】** をクリックして JSON ファイルまたは XML ファイルを選択します。
    3. **【完了】** をクリックします。  
データオブジェクトがプロジェクトフォルダに表示されます。

- 複数の JSON ファイルまたは複数の XML ファイルが含まれるフォルダに対して、複合型ファイルデータオブジェクトを作成するには、次の手順を実行します。
  1. **【参照】** をクリックして接続を選択します。
  2. **【リソースの追加】** ダイアログボックスで、**【追加】** をクリックしてフォルダ内の JSON または XML ファイルを選択します。
  3. **【完了】** をクリックします。  
データオブジェクトがプロジェクトフォルダに表示されます。
  4. プロジェクトフォルダ内のデータオブジェクトを選択して、**【詳細】** > **【ランタイム: 読み取り】** > **【ソースファイルのディレクトリ】** をクリックします。
  5. ファイルパスのファイル名を削除し、フォルダ名はそのまま保持します。

## Avro または Parquet データソースからの複合型ファイルデータオブジェクトの作成

Avro または Parquet データソースから **【ファイル】** または **【接続】** のアクセスタイプで複合型ファイルデータオブジェクトを作成できます。このデータオブジェクトに対して、カラムプロファイルを作成できます。

1. **【オブジェクトエクスプローラ】** ビューで、プロジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【データオブジェクト】** をクリックします。  
**【新規】** ダイアログボックスが表示されます。
3. **【物理データオブジェクト】** > **【複合型ファイルデータオブジェクト】** を選択し、**【次へ】** をクリックします。  
**【新しい複合型ファイルデータオブジェクト】** ダイアログボックスが表示されます。
4. データオブジェクトの名前を入力します。
5. アクセスタイプを **【接続】** または **【ファイル】** から選択できます。
  - アクセスタイプに **【接続】** を選択した場合は、次の手順を実行します。
    1. **【参照】** をクリックして HDFS 接続を選択します。
    2. **【接続の選択】** ダイアログボックスで、データソースを選択して **【OK】** をクリックします。
    3. **【新しい複合型ファイルデータオブジェクト】** ダイアログボックスで、**【完了】** をクリックします。  
データオブジェクトがプロジェクトフォルダに表示されます。
  - アクセスタイプに **【ファイル】** を選択し、リソース形式に **【Binary】** を選択した場合は、次の手順を実行します。
    1. **【参照】** をクリックして、ローカルマシン上の Avro または Parquet ファイルを選択します。
    2. **【新しい複合型ファイルデータオブジェクト】** ダイアログボックスで、**【完了】** をクリックします。  
データオブジェクトがプロジェクトフォルダに表示されます。
    3. プロジェクトフォルダ内のデータオブジェクトを選択して、**【データオブジェクト操作】** ビューをクリックします。
    4. **【データオブジェクト操作】** ビューで、**【読み取り】** > **【詳細】** タブをクリックします。
    5. **【詳細】** タブの **【ファイルパス】** フィールドに、Linux または Windows マシンにあるデータソースのファイルパスを入力します。
    6. ファイル形式に **【カスタム入力】** を指定します。

7. Avro データソース用の【入力形式】フィールドに **com.informatica.avro.AvroToXML** と入力し、Parquet データソース用の【入力形式】フィールドに **com.informatica.parquet.ParquetToXML** と入力します。入力形式を追加すると、データプロセッサトランスフォーメーションによって実行時に Avro または Parquet 形式のデータソースが処理され、XML 形式のデータソースに変換されます。
- アクセスタイプに【ファイル】を選択し、リソース形式に【Avro】または【Parquet】を選択した場合は、次の手順を実行します。
  1. 【参照】をクリックして、ローカルマシンの Avro または Parquet ファイルを選択します。
  2. 【新しい複合型ファイルデータオブジェクト】ダイアログボックスで、【完了】をクリックします。データオブジェクトがプロジェクトフォルダに表示されます。
  3. データオブジェクトを作成したら、【データオブジェクト操作】 > 【読み取り】 > 【詳細】タブに移動して、【ファイルパス】フィールドのファイルパスが、Linux または Windows マシンのデータソースに対応するかどうかを確認します。

注: リソース形式として【Avro】または【Parquet】を選択できるのは、フラットな構造の Avro および Parquet データソースの場合のみです。

複数の Avro ファイルまたは複数の Parquet ファイルが含まれるフォルダを 1 つ選択して、データオブジェクトを作成できます。データオブジェクトを作成したら、【データオブジェクト操作】 > 【読み取り】 > 【詳細】タブに移動して、【ファイルパス】フィールドのファイルパスが、Linux または Windows マシンのデータソースが含まれるフォルダを指すかどうかを確認します。

## 半構造化データソースでのカラムプロファイルの作成

Avro、JSON、Parquet、または XML データソースからフラットファイルデータオブジェクトまたは複合型ファイルデータオブジェクトを作成後、そのデータオブジェクトに対してカラムプロファイルを作成して実行できます。

1. 【オブジェクトエクスプローラ】ビューで、Avro、JSON、Parquet または XML ファイルのデータオブジェクトを選択します。
2. 【ファイル】 > 【新規】 > 【プロファイル】をクリックします。  
【新規】ダイアログボックスが表示されます。
3. 【プロファイル】を選択します。【次へ】をクリックします。  
【新規プロファイル】ダイアログボックスが表示されます。
4. 【新規プロファイル】ダイアログボックスで、プロファイルの名前と説明（省略可能）を入力します。
5. 【拡張ファイル形式の処理】オプションを選択します。【次へ】をクリックします。

次の画像は、【新しいプロファイル】ウィザードで【拡張ファイル形式の処理】オプションが選択された状態を示しています。

1. [拡張ファイル形式の処理]。このオプションを選択し、半構造化データソースを処理します。

**注:** リソース形式を **[Avro]** または **[Parquet]** として選択した場合、Avro および Parquet の各データソースに対しては **[拡張ファイル形式の処理]** オプションは表示されません。

6. **[単一のデータオブジェクトプロファイル]** ページで、必要に応じて、**[カラムの選択]** および **[データドメイン検出]** の下にあるカラムとオプションを選択します。**[完了]** をクリックします。

**注:** Linux マシンに Developer tool がインストールされていて、JSON または XML 物理データオブジェクトが、テキストファイルを含むフラットファイルデータオブジェクトの場合は、次のタスクを実行します。

1. **[概要]** タブの **[精度]** 値を更新して、サーバーのデータソースのファイルパスの文字数が含まれるようにします。
2. フラットファイルデータオブジェクトでプロファイルを作成後、サーバー内の場所へのデータソースファイルパスを更新します。ファイルパスを更新するには、**[詳細]** タブの **[ランタイム: 読み取り]** > **[ソースファイルディレクトリ]** をクリックし、ファイルパスを追加します。
7. プロファイルを右クリックし、**[プロファイルの実行]** を選択します。  
プロファイル結果が表示されます。

## 第 20 章

# Informatica Developer のルール

この章では、以下の項目について説明します。

- [Informatica Developer のルールの概要, 166 ページ](#)
- [Informatica Developer でのルールの作成, 167 ページ](#)
- [Informatica Developer でのルールの適用, 167 ページ](#)

## Informatica Developer のルールの概要

ルールは、カラムプロファイルの実行時にソースデータに適用される条件を定義するビジネスロジックです。ルールをプロファイルに追加して、データを検証できます。ルール、定義済みルール、または再利用可能なルールとして検証されるマップレットをカラムプロファイルで使用できます。

次の方法を使用して、列プロファイルでルールを使用できます。

- Developer tool で、マップレットを作成し、ルールとして検証します。ルールは、Analyst ツールに再利用可能なルールとして表示されます。ルールをカラムプロファイルに適用するには、Analyst ツールと Developer tool を使用します。
- カラムプロファイルでは、定義済みのルールを使用できます。Informatica は、Developer tool と Analyst ツールを使用して定義済みのルールを提供します。
- Analyst ツールで、ルール仕様を作成し、マップレットを生成します。ルール仕様は、Analyst ツールのカラムプロファイルに適用できます。Developer tool で、マップレットをルールとして検証します。このルールは、カラムプロファイルで使用できる再利用可能なルールとして表示されます。

**注:** Developer tool では、カラムプロファイルにルール仕様を追加、編集、または削除することはできません。

ルールは、以下の要件を満たす必要があります。

- 入力トランスフォーメーションと出力トランスフォーメーションを含める必要がある。ルールでデータソースを使用することはできません。
- 式トランスフォーメーション、ルックアップトランスフォーメーション、およびパッシブなデータ品質トランスフォーメーションを含めることができる。他のタイプのトランスフォーメーションを含めることはできません。例えば、一致トランスフォーメーションはアクティブなトランスフォーメーションなので、ルールに含めることはできません。
- 入力グループ間のカーディナリティが指定されていない。

# Informatica Developer でのルールの作成

Developer ツールでルールを作成するには、マプレットをルールとして検証する必要があります。

Developer ツールでマプレットを作成します。

1. マプレットエディタを右クリックします。
2. **【検証】** > **【ルール】** を選択します。

# Informatica Developer でのルールの適用

保存したカラムプロファイルにルールを追加できます。結合分析用に設定したプロファイルにルールを追加することはできません。

1. **【Object Explorer】** ビューを参照して必要なプロファイルを見つけます。
2. プロファイルを右クリックし、**【開く】** を選択します。  
プロファイルがエディタで開かれます。
3. **【定義】** タブをクリックし、**【ルール】** を選択します。
4. **【追加】** をクリックします。  
**【ルールの適用】** ダイアログボックスが開きます。
5. **【参照】** をクリックして、適用するルールを見つけます。  
リポジトリプロジェクトからルールを選択して、**【OK】** をクリックします。
6. **【入力値】** の下の **【値】** カラムをクリックし、ルールの入力ポートを選択します。
7. 必要に応じて、**【出力行】** の下の **【値】** カラムをクリックし、ルールの出力ポートの名前を編集します。  
ルールが **【定義】** タブに表示されます。

## 第 21 章

# マプレットとマッピングのプロファイリング

この章では、以下の項目について説明します。

- [マプレットとマッピングのプロファイリングの概要, 168 ページ](#)
- [マプレットまたはマッピングオブジェクトに対するプロファイルの実行, 168 ページ](#)
- [マッピングまたはマプレットオブジェクトのプロファイルの比較, 169 ページ](#)
- [プロファイルからのマッピングの生成, 169 ページ](#)

## マプレットとマッピングのプロファイリングの概要

マプレットまたはマッピングのオブジェクトに対するカラムプロファイルを定義できます。マッピングまたはマプレットの設計を確認するときにプロファイル結果を保存しない場合は、マプレットまたはマッピングオブジェクトに対してプロファイルを実行します。また、プロファイルからマッピングを生成することもできます。

## マプレットまたはマッピングオブジェクトに対するプロファイルの実行

マプレットまたはマッピングオブジェクトに対してプロファイルを実行すると、プロファイルがすべてのデータカラムに対して実行され、データオブジェクトのステージングされたデータに対するドリルダウン操作が可能になります。複数の出力ポートを使用して、マプレットまたはマッピングオブジェクトに対してプロファイルを実行できます。

プロファイルでは、選択したオブジェクトの出力ポートへのマッピングによってソースデータがトレースされます。マッピングを実行した場合にこれらのポートに表示されるデータが分析されます。

1. マプレットまたはマッピングを開きます。
2. マプレットまたはマッピングが有効であることを確認します。
3. データオブジェクトまたはトランスフォーメーションを右クリックし、[今すぐプロファイル] を選択します。



トランスフォーメーションに複数の出力グループがある場合は、**【出力グループの選択】** ダイアログボックスが表示されます。

4. トランスフォーメーションに複数の出力グループがある場合は、必要な出力グループを選択します。
5. **【OK】** をクリックします。

プロファイルの**【結果】** タブにプロファイル結果が表示されます。プロファイルサマリログファイルおよびマッピングログファイルを表示して、Developer ツールによって実行されたタスクの詳細を取得できます。

**注:** オペレーティングシステムプロファイルを使用してプロファイルを実行した場合、サマリログはデータ統合サービス用に構成されたログディレクトリにあり、マッピングログはオペレーティングシステムのプロファイル用に構成されたログディレクトリにあります。

## マッピングまたはマップレットオブジェクトのプロファイルの比較

マップレットまたはマッピングの2つのオブジェクトを分析し、これらのオブジェクトのカラムプロファイルの結果を比較するプロファイルを作成できます。

単一のマッピングまたはマップレットオブジェクトのカラムプロファイルと同様に、プロファイル比較はすべてのデータカラムに対して実行され、データオブジェクトにステージングされているデータに対してドリルダウン操作を実行できるようにします。ソーステーブルからターゲットテーブルにデータを移動したあと、プロファイルと比較してデータの移行を確認できます。また、時間の経過と共に変化するデータソースのプロファイルと比較することもできます。

単一のマッピングまたはマップレットオブジェクトのプロファイルと同様に、プロファイル比較はすべてのデータカラムに対して実行されます。

1. マップレットまたはマッピングを開きます。
2. マップレットまたはマッピングが有効であることを確認します。
3. エディタで **Ctrl** キーを押しながら2つのオブジェクトをクリックします。
4. いずれかのオブジェクトを右クリックし、**【プロファイルの比較】** を選択します。
5. 必要に応じて、一方のオブジェクトから他方のオブジェクトまでのカラムを照合するようにプロファイル比較を設定します。
6. 必要に応じて、一方のオブジェクトのカラムをクリックして他方のオブジェクトのカラムにドラッグし、カラムを照合します。
7. 必要に応じて、プロファイルですべてのカラムを分析するか一致したカラムのみを分析するかを選択します。
8. **【OK】** をクリックします。

## プロファイルからのマッピングの生成

プロファイルからマッピングオブジェクトを作成できます。作成したマッピングオブジェクトを使用し、有効なマッピングを作成します。作成するマッピングには、プロファイリングしたオブジェクトに基づくデータソ

ースが含まれます。また、プロファイルルールロジックに基づくトランスフォーメーションを含めることもできます。マッピングの作成後に、オブジェクトを追加してマッピングを完成させます。

1. **Object Explorer** ビューで、マッピングを作成するプロファイルを見つけます。
  2. プロファイル名を右クリックし、**【マッピングの生成】** を選択します。  
**【マッピングの生成】** ダイアログボックスが表示されます。
  3. マッピング名を入力します。必要に応じて、マッピングの説明を入力します。
  4. マッピングのフォルダの場所を確認します。  
デフォルトでは、マッピングはプロファイルと同じプロジェクトの **Mappings** フォルダに作成されます。マッピングの場所として別の場所を選択するには、**【参照】** をクリックします。
  5. Developer ツールでマッピングを作成するために使用されるプロファイル定義を確認します。別のプロファイルを使用するには、**【プロファイルの選択】** をクリックします。
  6. **【完了】** をクリックします。  
マッピングが **Object Explorer** に表示されます。
- オブジェクトをマッピングに追加してマッピングを完成させます。

## 第 22 章

# Informatica Developer のカラム プロファイル結果

この章では、以下の項目について説明します。

- [Informatica Developer のカラムプロファイル結果, 171 ページ](#)
- [カラム値のプロパティ, 172 ページ](#)
- [カラムパターンのプロパティ, 172 ページ](#)
- [カラム統計のプロパティ, 172 ページ](#)
- [カラムデータ型のプロパティ, 173 ページ](#)
- [Informatica Developer でのキュレーション, 174 ページ](#)
- [Informatica Developer からのプロファイル結果のエクスポート, 175 ページ](#)

## Informatica Developer のカラムプロファイル結果

カラムプロファイル分析は、データ品質に関して、値の頻度、データのパターン、および統計に焦点を当てた情報を提供します。

以下の表に、それぞれのタイプの分析のプロファイル結果を示します。

プロファイルタイプ	プロファイル結果
カラムプロファイル	<ul style="list-style-type: none"><li>- 一意の値と NULL 値の割合と数に関する統計</li><li>- 推測されたデータ型</li><li>- データソースでデータに対して宣言されているデータ型</li><li>- 最大値と最小値</li><li>- 最後にプロファイルを実行した日時</li><li>- カラム内の一意の各データ要素の割合と数に関する統計</li><li>- カラム内の一意の各文字パターンの割合と数に関する統計</li></ul>
プライマリキープロファイル	<ul style="list-style-type: none"><li>- 推測されるプライマリキー</li><li>- キー違反</li></ul>
機能依存性プロファイル	<ul style="list-style-type: none"><li>- 推測される機能依存性</li><li>- 機能依存性違反</li></ul>

## カラム値のプロパティ

カラム値のプロパティは、プロファイリングしたカラムの値、および各値が各カラムに出現する頻度を示します。頻度は数、割合、および棒グラフで表されます。

カラム値のプロパティを表示するには、**【表示】** リストから **【値】** を選択します。カラム値をダブルクリックし、その値を含む行にドリルダウンします。

以下の表に、カラム値のプロパティを示します。

プロパティ	説明
値	プロファイルのカラムのすべての値を示す一覧。
頻度	カラムの値が出現する回数。
割合	カラムの値が出現する回数（カラム内のすべての値に対する割合で表される）。
チャート	割合の棒グラフ。

## カラムパターンのプロパティ

カラムパターンのプロパティは、プロファイリングしたカラムにおけるデータのパターンおよびパターンが各カラムに出現する頻度を示します。パターンは数、割合、および棒グラフで表されます。

パターン情報を表示するには、**【表示】** リストから **【パターン】** を選択します。パターンをダブルクリックし、そのパターンを含む行にドリルダウンします。

以下の表に、カラム値のパターンのプロパティを示します。

プロパティ	説明
パターン	選択されたカラムのパターン。
頻度	カラムのパターンが出現する回数。
割合	カラムのパターンが出現する回数（カラム内のすべての値に対する割合で表される）。
チャート	割合の棒グラフ。

## カラム統計のプロパティ

カラム統計には、値の最大長と最小長、および最初の値と最後の値などのプロパティが含まれます。

統計情報を表示するには、**【表示】** リストから **【統計】** を選択します。

以下の表に、カラム統計のプロパティを示します。

プロパティ	説明
最大長	カラムの最長値の長さ。
最小の長さ	カラムの最短値の長さ。
下	カラム内の最後の 5 つの値。
上	カラム内の最初の 5 つの値。
合計	数値データ型を持つカラムのすべての値の合計。

**注:** Integer タイプのカラムの場合、プロファイルには平均と標準偏差の統計も表示されます。

## カラムデータ型のプロパティ

カラムデータ型には、プロファイル結果の各カラムの推測されたデータ型がすべて含まれています。

データ型情報を表示するには、**[表示]** リストから **[データ型]** を選択します。特定のデータ型をダブルクリックすると、そのデータ型を含む行にドリルダウンします。

次の表に、カラムデータ型のプロパティを示します。

プロパティ	説明
データ型	プロファイルのカラムのすべての推測されたデータ型の一覧。
頻度	カラムに特定のデータ型が出現する回数（数値で表現される）。
一致率（%）	カラムに特定のデータ型が出現するパーセンテージ。
ステータス	データ型のステータスを示します。ステータスは推測済み、承認済み、または拒否済みです。 <b>推測済み</b> カラムのデータ型が、Developer tool によって推測されたものであることを示します。 <b>承認済み</b> カラムのデータ型が承認済みであることを示します。データ型を承認すると、そのデータ型がモデルリポジトリにコミットされます。 <b>却下済み</b> カラムのデータ型が拒否されたことを示します。

# Informatica Developerでのキュレーション

キュレーションとは、データソースで検出されたメタデータを検証し管理することで、使用とレポートに適したメタデータにするプロセスです。Informatica Developerでメタデータをキュレーションする場合、プロファイル結果の中の推測されたデータ型またはデータドメインを承認、拒否、リセットすることができます。

1つのカラムに対して1つのデータ型またはデータドメインを承認できます。カラムの拒否されたデータ型またはデータドメインは、非表示にできます。推測されたデータ型またはデータドメインを承認または拒否した後でも、データ型またはデータドメインをリセットすると、ステータスを推測済みに戻すことができます。

## データ型の承認

プロファイルの結果には、データソースのカラムごとに推測データ型、頻度、一致率、および推測ステータスが含まれています。各カラムに対して1つのデータ型を選択して承認することができます。

1. **【オブジェクトエクスプローラ】** ビューでプロファイルを選択して、それを開きます。
2. **【結果】** タブが表示されていることを確認します。
3. **【カラムプロファイリング】** ビューで、カラムを選択して値の頻度、パターン、データ型、統計を右側のパネルに表示します。
4. **【詳細】** パネルで、**【データ型】** を**【表示】** リストから選択します。  
カラムの推測されたデータ型が表示されます。
5. 承認する必要があるカラムを右クリックして、**【承認】** をクリックします。  
データ型のステータスが**【承認済み】** に変わります。
6. データ型のステータスを推測済みに戻すには、データ型を右クリックしてから**【リセット】** をクリックします。

## データ型の拒否

Informatica Developerでは、デフォルトで、プロファイル結果に推測されたデータ型が表示されます。推測または承認されたデータ型を拒否することもできます。拒否されたデータドメインを表示するか非表示にするかを選択できます。

1. **【Object Explorer】** ビューで、プロファイルを選択します。
2. プロファイルをダブルクリックして開きます。  
タブにプロファイルが開きます。
3. **【カラムプロファイリング】** ビューで、行を選択します。
4. 推測されたカラムデータ型を拒否するには、右側のパネルで**【データ型】** ビューを選択します。拒否する必要がある推測されたデータ型を選択し、その行を右クリックして**【拒否】** を選択します。  
Informatica Developerでは、拒否されたデータ型がデータ型リストで灰色に表示されます。
5. 拒否されたデータドメインを非表示にするには、その行を右クリックして**【拒否済みを非表示】** を選択します。
6. 拒否されたデータ型を表示するには、行のうち1つを右クリックして**【拒否済みを表示】** を選択します。

# Informatica Developer からのプロファイル結果のエクスポート

カラムプロファイル結果を.csv ファイルまたは Microsoft Excel ファイルにエクスポートできます。プロファイル結果を Microsoft Excel ファイルにエクスポートすると、Developer tool は情報を.xlsx ファイルに保存します。

1. **[Object Explorer]** ビューで、プロファイルを開きます。
2. 必要に応じて、プロファイルを実行してプロファイル結果を更新します。
3. **[結果]** ビューを選択します。
4. カラムを選択します。
5. **[詳細]** で、**[値]**、**[パターン]**、または **[データ型]** を選択し、**[エクスポート]** アイコンをクリックします。  
**[データをファイルにエクスポート]** ダイアログボックスが開きます。
6. デフォルトのファイル名を受け入れるか変更します。
7. エクスポートするデータのタイプを選択します。**[選択したカラムの値]**、**[選択したカラムのパターン]**、**[選択したカラムのデータ型]**、または **[すべて (概要、値、パターン、データ型、統計、プロパティ)]** を選択できます。
8. **[参照]** をクリックして場所を選択し、ファイルをコンピュータ上にローカルに保存します。
9. フィールド名を最初の行としてエクスポートしない場合は、**[フィールド名を最初の行としてエクスポート]** チェックボックスをオフにします。
10. **[OK]** をクリックします。

## 第 23 章

# Informatica Developer のスコアカード

この章では、以下の項目について説明します。

- [Informatica Developer のスコアカードの概要, 176 ページ](#)
- [スコアカードの作成, 176 ページ](#)
- [スコアカードリネージュのリソースファイルのエクスポート, 177 ページ](#)
- [Informatica Developer からのスコアカードリネージュの表示, 177 ページ](#)

## Informatica Developer のスコアカードの概要

スコアカードは、プロファイルでの品質解析結果を視覚的に表したものです。スコアカードは Developer ツールで表示できます。Developer ツールでスコアカードを作成した後、Analyst ツールに接続してそのスコアカードを編集用に開くことができます。データオブジェクトの現在のデータに対して、またはプロファイリングウェアハウスにステージングされているデータに対してスコアカードを実行します。

Analyst ツールでスコアカードを編集または実行したり、メトリックまたはメトリックグループのスコアカードリネージュを表示したりできます。

## スコアカードの作成

スコアカードを作成し、プロファイルからスコアカードにカラムを追加します。スコアカードにカラムを追加する前に、プロファイルを実行する必要があります。

1. **[Object Explorer]** ビューで、スコアカードを作成するプロジェクトまたはフォルダーを選択します。
2. **[ファイル] > [新規] > [スコアカード]** をクリックします。  
**[新しいスコアカード]** ダイアログボックスが表示されます。
3. **[追加]** をクリックします。  
**[プロファイルの選択]** ダイアログボックスが表示されます。追加するカラムが含まれているプロファイルを選択します。
4. **[OK]** をクリックし、**[次へ]** をクリックします。
5. スコアカードに追加するカラムを選択します。



デフォルトでは、プロファイルで定義されているカラムとルールが選択されます。プロファイルに含まれていないカラムは追加できません。

6. **【完了】** をクリックします。  
スコアカードが作成されます。
7. 必要に応じて、**【Informatica Analyst で開く】** をクリックして Analyst ツールに接続し、Analyst ツールでスコアカードを開きます。

## スコアカードリネージュのリソースファイルのエクспорт

スコアカードを含むプロジェクトや、依存オブジェクトを Metadata Manager のリソースファイルとしてエクспортできます。XML 形式でエクспортされたリソースファイルを使用して Metadata Manager のスコアカードリネージュのリソースを作成およびロードします。

1. **エクспортウィザード**を開くには、**【ファイル】** > **【エクспорт】** の順にクリックします。
2. **【Informatica】** > **【Metadata Manager のリソースファイル】** を選択します。
3. **【次へ】** をクリックします。
4. **【参照】** をクリックしてエクспортするスコアカードオブジェクトとリネージュを含むプロジェクトを選択します。
5. **【次へ】** をクリックします。
6. エクспортするスコアカードオブジェクトを選択します。
7. エクспортファイルの名前とファイルの場所を入力します。
8. 選択したオブジェクトとともに**エクспортウィザード**でエクспортされる依存オブジェクトを表示するには、**【次へ】** をクリックします。  
**エクспортウィザード**に依存オブジェクトが表示されます。
9. **【完了】** をクリックします。  
オブジェクトが XML ファイルにエクспортされます。

## Informatica Developer からのスコアカードリネージュの表示

Developer ツールからメトリックまたはメトリックグループのスコアカードリネージュを表示するには、Analyst ツールを起動します。

1. **【Object Explorer】** ビューで、スコアカードが含まれるプロジェクトまたはフォルダを選択します。
2. スコアカードをダブルクリックして開きます。  
タブにスコアカードが表示されます。
3. **【Informatica Analyst で開く】** をクリックします。  
ブラウザウィンドウ内に Analyst ツールが開きます。

4. Analyst ツールの【スコアカード】ビューで、メトリックまたはメトリックグループを選択します。
5. 右クリックして【リネージュを表示する】を選択します。  
スコアカードリネージュダイアグラムがダイアログボックスに表示されます。

## 第 24 章

# Informatica Developer でのデータドメイン検出

この章では、以下の項目について説明します。

- [Informatica Developer でのデータドメイン検出の概要, 179 ページ](#)
- [Informatica Developer のデータドメイングロッサリ, 180 ページ](#)
- [Informatica Developer のデータドメイン検出オプション, 183 ページ](#)
- [Informatica Developer でのデータドメイン検出実行プロファイルの作成, 186 ページ](#)
- [Informatica Developer でのプロファイルの編集, 187 ページ](#)
- [Informatica Developer でのデータドメイン検出実行プロファイルの実行, 187 ページ](#)
- [Informatica Developer のデータドメイン検出結果, 188 ページ](#)

## Informatica Developer でのデータドメイン検出の概要

データドメイングロッサリを使用してデータドメインを管理します。データドメインを作成するために、事前定義されたデータルールやカラム名のルールを使用できます。また、カラムプロファイル結果の中の特定の値やパターンに基づいてデータドメインを生成することもできます。

ソースカラム、カラムデータおよびカラム名と照合するデータドメイン、サンプリングオプション、ドリルダウンオプション、およびランタイム環境を選択できます。データドメイン検出の対象とする最大行数を選択することが可能です。データドメイン検出の適合条件を選択できます。データドメイン検出から NULL 値を除外することが可能です。プロファイルの実行後に、結果に対して検証、作成、ドリルダウンができます。また、Developer tool のエディタから、結果をデータモデルに追加できます。

サンプリングオプションやフィルタを設定したプロファイルを作成して、データドメイン検出を実行できます。プロファイルを実行すると、データソースにサンプリングオプションとフィルタが適用されて、データセットが生成されます。データドメイン検出プロセスは、このデータセットを使用してデータドメインを検出します。

# Informatica Developer のデータドメインングロッサリ

データドメインとデータドメイングループはデータドメインングロッサリで管理します。データドメインとデータドメイングループを追加、編集、および削除できます。特定のデータドメインとデータドメイングループを検索することもできます。

データドメインはデータドメインングロッサリから XML ファイルにエクスポートできます。XML ファイルからデータドメインングロッサリにデータドメインをインポートすることもできます。Personal Health Information (PHI)、Personally Identifiable Information (PII)、またはその他のプロジェクトに関連する概念グループなど、特定のグループにデータドメインを整理するには、データドメイングループを作成します。1つのデータドメインを複数のデータドメイングループに分類できます。例えば、電話番号を PII データドメイングループと PHI データドメイングループの両方に分類できます。

## Informatica Developer でのデータドメイングループの作成

データドメインをデータドメイングループに追加することで、効果的なカラムデータ分析を行えます。

1. **【ウィンドウ】** > **【設定】** をクリックします。  
**【設定】** ダイアログボックスが表示されます。
2. **【設定】** ダイアログボックスで、**【Informatica】** > **【データドメインングロッサリ】** を選択します。  
Developer ツールの **【データドメインングロッサリ】** パネルに、すべてのデータドメインのリストが表示されます。
3. **【表示】** フィールドで、**【データドメイングループ】** を選択します。  
**【データドメインングロッサリ】** パネルのデータドメインリストが、データドメイングループ別に並べ替えられます。
4. **【データドメインングロッサリ】** パネルで、**【データドメイングループ】** を選択します。
5. **【追加】** をクリックします。  
**【データドメイングループ】** ダイアログボックスが表示されます。
6. 名前と説明を入力します。
7. **【次へ】** をクリックします。
8. **【選択】** をクリックして、**【データドメインの選択】** ダイアログボックスを開きます。
9. データドメイングループに追加するデータドメインを選択し、**【OK】** をクリックします。  
Developer ツールの **【データドメインの選択】** パネルに、選択したデータドメインが一覧表示されます。
10. **【完了】** をクリックします。  
データドメイングループがデータドメインングロッサリに追加されます。

## Informatica Developer でのデータドメインの作成

データドメインを作成し、データドメインングロッサリに追加することができます。データドメインを 1 つ以上のデータドメイングループに追加することもできます。

1. **【ウィンドウ】** > **【設定】** をクリックします。  
**【設定】** ダイアログボックスが表示されます。
2. **【設定】** ダイアログボックスで、**【Informatica】** > **【データドメインングロッサリ】** を選択します。  
Developer ツールの **【データドメインングロッサリ】** パネルに、すべてのデータドメインのリストが表示されます。

3. **【データドメインングロッサリ】** パネルで、**【データドメイン】** を選択します。
4. **【追加】** をクリックします。  
**【データドメイン】** ダイアログボックスが表示されます。
5. 名前と説明を入力します。
6. **【データルールを使用】** をクリックして、カラムデータに基づいてデータドメインを検出します。**【カラム名ルールを使用】** を選択して、データソース内のカラム名に基づいてデータドメインを検出することもできます。  
**【参照】** ボタンが有効になります。
7. **【参照】** をクリックして、**【場所の選択】** ダイアログボックスを開きます。
8. 適切なルールを選択し、**【OK】** をクリックします。  
データドメインを作成すると、Developer ツールがデータドメインに関連付けられているルールや他の依存オブジェクトがデータドメインングロッサリにコピーします。データドメインに関連付けられているルールを編集するには、元のルールに移動してそこで変更を加える必要があります。その後、変更したルールをデータドメインに再び関連付けます。  
選択したルールが、**【データルール】** フィールドと **【カラム名ルール】** フィールドに表示されます。
9. **【次へ】** をクリックします。
10. **【選択】** をクリックして、**【データドメイングループの選択】** ダイアログボックスを開きます。
11. データドメインを含めるデータドメイングループを選択し、**【OK】** をクリックします。  
Developer ツールの **【データドメイングループに割り当て】** ペインに、選択したデータドメイングループが表示されます。
12. **【完了】** をクリックします。  
Developer ツールがデータドメインをデータドメインングロッサリに追加されます。

## Informatica Developer でのプロファイル結果からのデータドメインの作成

カラムプロファイルを実行すると、ソースデータの値およびパターンを表示できます。次に、これらからデータドメインを作成できます。

1. カラムプロファイルを実行して、その結果を表示します。
2. データドメインを作成するベースとなるパターンの値を選択します。  
**【結果】** ビューに値、パターン、および統計が表示されます。
3. 値またはパターンを右クリックし、**【宛先】** > **【新規ドメイン】** を選択します。  
**【データドメイン】** ダイアログボックスが表示されます。
4. データドメイン名と説明（省略可能）を入力します。  
場所はデフォルトでデータドメインングロッサリに設定されます。
5. **【完了】** をクリックします。  
データドメインがデータドメインングロッサリに追加されます。

## Informatica Developer でのデータドメインの検出

データドメインングロッサリには、デフォルトですべてのデータドメインが表示されます。特定のデータドメインとデータドメイングループを検索できます。

**【設定】** ダイアログボックスの **【データドメインングロッサリ】** ペインに、すべてのデータドメインとデータドメイングループが表示されます。次の方法で、データドメインおよびデータドメイングループについての詳細情報を検索したり表示したりすることができます。

**データドメインとデータドメイングループを検索する。**

**【データドメインングロッサリ】** パネルの上部にあるフィールドに、データドメイン名またはデータドメイングループ名の一部を入力します。**【データドメイングループビュー】** で作業している場合は、Developer ツールが名前に検索文字列を含むデータドメインと、それらに関連付けられているデータドメイングループを一覧表示します。**【データドメインビュー】** で作業している場合は、Developer ツールが名前に検索文字列を含むすべてのデータドメインを一覧表示します。

**データドメイングループとグループ内のデータドメインを表示する。**

**【表示】** フィールドで、**【データドメイングループ】** を選択します。

**すべてのデータドメインを表示する。**

**【表示】** フィールドで、**【データドメイン】** を選択します。

**データドメインのプロパティを表示する。**

データドメイン名をクリックして、**【データドメインングロッサリ】** パネルの下にそのプロパティを表示します。説明および関連付けられているルールを表示できます。データドメインが属しているドメイングループを表示するには、**【データドメイングループで表示】** をクリックします。

**データドメイングループのプロパティを表示する。**

データドメイングループ名をクリックして、**【データドメインングロッサリ】** パネルの下にその説明を表示します。

## データドメインのインポート

Developer ツールでは、ソース XML ファイルからデータドメインングロッサリにデータドメインをインポートできます。インポートするデータドメインの情報がファイルに含まれていることを確認する必要があります。

1. データドメインングロッサリを開きます。
2. **【データドメイン】** または **【データドメイングループ】** が選択されていることを確認します。
3. **【インポート】** をクリックします。  
**【インポート】** ダイアログボックスが表示されます。
4. **【ファイル名】** フィールドに、データドメインをインポートする XML ファイル名を入力します。  
**【参照】** をクリックしてファイルを選択します。
5. **【次へ】** をクリックします。  
**【インポートするオブジェクトの選択】** ペインが表示され、ソースとターゲットを指定できます。
6. **【ソース】** パネルで、インポートするデータドメインを選択します。  
**注:** 複数のデータドメインを選択するには、Shift キーを押しながら選択します。
7. **【ターゲットへの自動一致】** をクリックして、データドメインを **【ターゲット】** パネルに移動します。  
選択しているソースの下位オブジェクトが選択したターゲットの名前、タイプ、および親階層で個別に照合され、一致するオブジェクトが追加されます。
8. **【解決】** をクリックして、重複するオブジェクトの処理方法を指定します。

インポートするオブジェクトの名前を変更するか、インポートするオブジェクトで既存のオブジェクトを上書きするか、既存のオブジェクトを再利用することができます。デフォルトでは、重複するすべてのオブジェクトの名前が変更されます。

9. **【次へ】** をクリックします。

確認のためにインポート設定のサマリが表示されます。 **【追加のインポート設定】** ペインで、追加のインポート設定を指定できます。

10. **【完了】** をクリックします。

## データドメインのエクスポート

データドメインおよびデータドメインルールを Developer ツールのデータドメインングロッサリから XML ファイルにエクスポートすることができます。

1. データドメインングロッサリを開きます。
2. **【データドメイン】** または **【データドメイングループ】** が選択されていることを確認します。
3. **【エクスポート】** をクリックします。  
**【エクスポート】** ダイアログボックスが表示されます。
4. データドメインをエクスポートするには、**【データドメインをエクスポート】** を選択します。**【データドメインルールをエクスポート】** を選択してデータドメインルールをエクスポートします。
5. **【次へ】** をクリックします。
6. **【ファイルにエクスポート】** パネルで、エクスポートするデータドメインまたはデータドメインルールを選択します。
7. データドメインをエクスポートするには、**【参照】** をクリックしてエクスポートファイルとその場所を選択します。Model Repository Service の別のプロジェクトにデータドメインをエクスポートするには、**【プロジェクトにコピー】** を選択して、データドメインルールのコピー先のプロジェクトを選びます。
8. **【次へ】** をクリックします。  
**【依存関係】** ペインに、依存オブジェクトのリストが表示されます。
9. **【次へ】** をクリックします。  
**【コンテンツのエクスポート設定】** ペインが表示されます。エクスポートする関連付けられた参照テーブルを選択できます。
10. **【完了】** をクリックします。

参照テーブルを使用するルールをデータドメインに関連付けると、そのデータドメインを作成した Developer ツールセッションでは参照テーブルをエクスポートできない場合があります。データドメインングロッサリで **【エクスポート】** をクリックした後、Model Repository Service から接続を切断し、再度接続してから、参照テーブルが使用されるルールをインポートしてください。

## Informatica Developer のデータドメイン検出オプション

プロファイルを作成してデータドメイン検出を実行する際には、ソースカラム、データドメイン、および推測オプションを選択できます。データ型やデータの長さに基づいてデータドメイン検出からカラムを除外することもできます。

## Informatica Developer でのデータドメインの選択

**データドメインの選択**オプションを選択すると、データドメイングローサリのすべてのドメインが一覧表示されます。特定のデータドメインを検索して選択し、データドメイン検出の実行対象にすることができます。

以下の表に、データドメイン検出の**データドメインの選択**オプションを示します。

オプション	説明
"プロファイルの実行"アクションの一部として有効化	プロファイルの実行時にデータドメイン検出オプションを含めます。
名前	データドメイン名。
説明	データドメインの説明。
データドメイングループ	データドメインが属しているデータドメイングループの名前。
データドメイングループを階層で表示	すべてのデータドメイングループをデータドメイングループ別に一覧表示します。

## Informatica Developer でのデータドメインカラムの選択

データドメイン検出の実行対象とするカラムを選択するには、**カラムの選択**オプションを使用します。

以下の表に、データドメイン検出の**カラムの選択**オプションを示します。

オプション	説明
カラム	カラム名。
データ型	カラムのデータ型。
精度	カラムの最大精度。
スケール	カラムの位取り。
NULL 可能	NULL 値を指定できるカラムを示します。
説明	カラムの説明。



## Informatica Developer のデータドメイン推測オプション

推測オプションでは、ドメイン検出の実行対象を、カラムデータ、カラム名、またはその両方から選択できます。データソース内のすべての行をプロファイルで処理する必要があるかどうかを指定できます。データドメインの一致に適合条件を選択し、データドメイン検出から NULL を除外するように選択できます。

以下の表に、データドメイン検出の推測オプションを示します。

オプション	説明
デフォルトの推測オプションをオーバーライド	定義済みの推測オプションを変更できるようにします。
データ	プロファイルのカラムデータに対して実行します。
カラム名	プロファイルのカラムタイトルに対して実行します。
データとカラム名	プロファイルのカラムデータとカラムタイトルの両方に対して実行します。
プロファイルする最大行数	プロファイルを実行できる最大行数。Developer tool では、ソースの先頭行から行が選択されます。
行の最小割合	データドメインの一致に必要なデータセット内の行の最小一致率。
行の最小数	データドメインの一致に必要なデータセット内の行の最小数。
データドメイン検出から NULL 値を除外	データドメイン検出用のデータセットから NULL 値を除外します。

### 最小一致率

データドメイン検出の適合条件として、データセット内の行の最小割合を選択できます。

一致率は、一致する行の数を行の合計数で除算した割合です。

**注:** Developer tool では、NULL 値は一致しない行とみなされます。NULL 値の数が多いカラムでは、最小一致率を低く指定しないと、結果がデータドメイン推測にならない場合があります。

#### 例

行数が 10,000 のデータソースがあり、そのうちの 2,500 行のコメントカラムに社会保障番号が含まれている場合を考えます。カラムプロファイルとデータドメイン検出を作成し、適合条件として、行の最小割合を 30% に設定します。プロファイルを実行すると、結果には、推測されたデータドメインとして社会保障番号が表示されません。これは、最小の適合条件が、データソースに含まれる行の 30%または 3,000 行に設定されているためです。

### 適合する行の最小数

データドメイン検出の適合条件として、データセット内の行の最小数を選択できます。

#### 例

行数が 10,000 のデータソースがあり、そのうちの 3 行のコメントカラムに電子メールアドレスが含まれている場合を考えます。カラムプロファイルとデータドメイン検出プロファイルを作成し、適合条件として、行の最小数を 1 に設定します。プロファイルを実行すると、結果には、他の推測されたデータドメインと一緒に、3 つの適合する行を持つ推測されたデータドメインとして、電子メールアドレスが表示されます。

## NULL 値の除外

データソースに対してデータドメイン検出を実行する際に、NULL 値を除外することができます。NULL 値を除外するオプションを指定して行の最小割合を選択した場合、一致率は、合計行数からカラム内の NULL 値を除いた数で一致行数を除算した割合になります。

**【データドメイン検出から NULL 値を除外】** オプションと各種のサンプリングオプションやフィルタを選択すると、データドメイン検出プロセスの結果は異なるものになります。

次のシナリオでは、NULL 値を除外するオプションと、1 つのサンプリングオプションおよびフィルタを組み合わせて選択した場合のデータドメイン検出結果について説明しています。

- サンプリングオプションとして **【すべての行】** を選択し、フィルタは選択しない場合。データドメイン検出では、カラム内のすべての NULL 値が無視されます。
- サンプリングオプションを 1 つ選択し、フィルタは選択しない場合。サンプリングされたデータに含まれるすべての NULL 値が無視され、サンプリングされたデータの残りのデータに対してデータドメイン検出が実行されます。
- サンプリングオプションとして **【すべての行】** を選択し、フィルタを複数選択する場合。フィルタされたデータに含まれるすべての NULL 値が無視され、フィルタされたデータの残りのデータに対してデータドメイン検出が実行されます。
- サンプリングオプションを 1 つ選択し、フィルタを複数選択する場合。サンプル内のフィルタされたデータに含まれる NULL 値が無視され、フィルタされたデータの残りのデータに対してデータドメイン検出が実行されます。

### 例

行数が 10,000 のデータソースがあり、そのうちの 3,000 行のコメント列に社会保障番号が含まれている場合を考えます。カラムプロファイルとデータドメイン検出を作成し、次のオプションを選択します。

- **【データドメイン検出から NULL 値を除外】** オプションを選択します。
- サンプリングオプションとして **【すべての行】** を選択します。
- **【行の最小割合】** オプションを選択して、オプションを 12% に設定します。

プロファイルを実行すると、プロファイルはデータセットに対して実行され、データドメイン検出では NULL 値が無視されます。

## Informatica Developer でのデータドメイン検出実行プロファイルの作成

データソース内のデータドメインを、単独のデータオブジェクトプロファイルの一部として、またはエンタープライズ検出プロファイルとして検出できます。データドメイン検出の実行後、Developer ツールのエディタから結果の検証とドリルダウンを行ったり、結果をデータモデルに追加したりすることができます。

1. **【Object Explorer】** ビューで、プロファイル用のデータオブジェクトが含まれるプロジェクトを選択します。
2. データオブジェクトを右クリックし、**【プロファイル】** を選択します。  
**新規作成ウィザード**が表示されます。
3. **【プロファイル】** を選択します。
4. **【次へ】** をクリックします。

別のペインが表示され、プロファイルの全般プロパティを設定できます。

5. 必要に応じて、プロファイル名と説明を変更します。データオブジェクトの追加や削除も可能です。
6. **【次へ】** をクリックします。
7. データドメイン検出を行うカラムと、そのカラムと照合するデータドメインを選択します。
8. 必要に応じて、デフォルトの推測オプションを変更します。
9. **【完了】** をクリックして、プロファイルを作成します。

## Informatica Developer でのプロファイルの編集

プロファイルを実行してデータドメイン検出を実行した後、プロファイルに変更を加えることができます。特定のデータ型のカラムを除外したり、カラム選択、データドメイン選択、および推測の各オプションを変更したりすることができます。

1. **【Object Explorer】** ビューで、編集するプロファイルが含まれるプロジェクトまたはフォルダーを選択します。
2. プロファイルをダブルクリックして開きます。  
タブにプロファイル定義が表示されます。
3. 必要に応じて、カラム選択、データドメイン選択、および推測の各オプションに変更を加えます。
4. **【カラムの選択】** セクションで、**【カラムの除外】** をクリックして、データ型に基づいて除外オプションを設定します。  
**【カラムの除外】** ダイアログボックスが表示されます。
5. 変更を保存します。

## Informatica Developer でのデータドメイン検出実行 プロファイルの実行

プロファイルは作成後すぐに実行されるよう指定できます。作成後にプロファイルを手動で実行することもできます。

1. **【Object Explorer】** ビューで、実行するプロファイルを含むプロジェクトまたはフォルダを選択します。  
プロファイルを自動的に実行するには、プロファイルの作成時に**新しいプロファイルウィザード**で**【終了時にプロファイルを実行】**を選択します。
2. プロファイルをダブルクリックして開きます。  
タブにプロファイル定義が表示されます。
3. プロファイルを右クリックし、**【プロファイルの実行】**を選択します。  
**【プロファイルの実行】** ダイアログボックスが表示され、プロファイル実行ステータスが表示されます。

# Informatica Developer のデータドメイン検出結果

データドメイン検出結果には、データドメインに一致するカラムに関する統計が表示されます。この情報には、データドメイン一致への適合条件や、カラム名がデータドメインに一致するかどうかなどが含まれます。

分析のためにさらに結果をドリルダウンすることができます。データソースのすべての行の結果を検証したり、Developer tool のエディタからデータモデルに結果を追加したりすることもできます。結果は、データドメイン、データドメイングループ、およびカラムでソートできます。データドメイン検出結果は Microsoft Excel ファイルにエクスポートできます。

以下の表に、データドメイン検出結果を示します。

カラム名	説明
名前	[データドメイン]、[データドメイングループ]、または [カラム] ビューのどれを選択したかに応じた、データドメイン、データドメイングループ、またはカラムの名前。
接続	接続の名前。
ステータス	カラムの推測ステータス。
データ一致率	データドメインの一致に必要な行の最小一致率。
行数を合わせる	データドメインの一致に必要な行の最小数。
NULL の行の%	カラムの NULL 値の割合。
合計行数	行の合計数です。
カラム名の一致	カラム名がデータドメイン名に一致するかどうかを示します。
データドメイングループ	データドメインが属しているデータドメイングループ。
文書化されたデータ型	プロファイルオブジェクト内のカラムに対して宣言されたデータ型。
ドリルダウン	選択すると、行にドリルダウンします。
検証済み	データソースのすべての行についてのデータドメインの一致の検証を示します。
最終実行時刻	最後にプロファイルを実行した日時。

## データドメイングループ別の表示

データドメイン検出結果をデータドメイングループでソートして表示できます。

1. プロファイルを実行して、その結果を表示します。
2. [結果] をクリックします。
3. [データドメイン検出] をクリックします。  
データドメイン検出結果が右側のパネルに表示されます。
4. [表示] フィールドで [データドメイン] オプションが選択されていることを確認します。

5. **【データドメイングループの階層を表示】** を選択して、データドメイングループでソートされた結果を表示します。

## コラム別の表示

データドメインに一致するソースカラムでソートされたデータドメイン検出結果を表示できます。

1. プロファイルを実行して、その結果を表示します。
2. **【結果】** をクリックします。
3. **【データドメイン検出】** をクリックします。

データドメイン検出結果が右側のパネルに表示されます。

4. **【カラム】** を選択して、データドメインに一致するソースカラムでソートされた結果を表示します。

## 結果の検証

プロファイルを実行すると、データソースのサンプルが分析されて、プロファイル結果が推測されます。ソースデータのすべての行についてプロファイルを実行して、推測結果を検証できます。

1. プロファイルを実行して、その結果を表示します。
2. **【結果】** をクリックします。
3. **【データドメイン検出】** をクリックします。

データドメイン検出結果が右側のパネルに表示されます。

4. 右側のパネルで、検証するカラムを選択します。
5. カラムを右クリックし、**【検証】** を選択して、データソースのすべての行に対してプロファイルを実行します。

結果の検証後、**【データ一致率】** の値または **【行数を合わせる】** の値の変化を確認できる場合があります。

6. 複数のカラムの推測結果を検証するには、複数のカラムを選択します。それから右クリックすると、**【すべて検証】** を選択することができます。

## データドメインの承認

単一のデータオブジェクトプロファイルでデータドメイン検出を実行する場合は、推測されたデータドメインを複数カラムに対して一度に承認することができます。エンタープライズ検出の一部としてデータドメイン検出を実行する場合は、一度に1つのソースカラムのデータドメインを承認することができます。エンタープライズ検出の後に複数カラムのデータドメインを承認する場合は、データオブジェクトプロファイルのタスクを個別に開いてデータドメインを承認することができます。

1. **【Object Explorer】** ビューで、プロファイルを選択します。
2. プロファイルをダブルクリックして開きます。  
タブにプロファイルが開きます。
3. 単一のデータオブジェクトプロファイルを実行した場合、**【データドメイン検出】** ビューを選択してから行を選択します。その行には、カラムごとのデータドメイン検出の結果が入っています。
4. その行を右クリックして、**【承認】** を選択します。

データドメインのインタフェースステータスが **【承認済み】** に変わります。

5. エンタープライズ検出を実行した場合は、**【データドメイン】** ビューを選択してからデータドメインを選択します。

データドメインが一致するカラムが、右側のパネルに表示されます。

- 承認する必要があるカラムを右クリックして、**【承認】** を選択します。必要に応じて、拒否されたカラムを複数選択してそれらを承認することもできます。  
データドメインのインタフェースステータスが **【承認済み】** に変わります。
- データドメインのステータスを推測済みに戻すには、その行を右クリックしてから **【リセット】** をクリックします。

## データドメインの却下

Informatica Developer では、デフォルトで、プロファイル結果に推測されたデータドメインが表示されます。推測または承認されているデータドメインは、却下することもできます。却下されたデータドメインを表示するか非表示にするか選択することができます。

- 【Object Explorer】** ビューで、プロファイルを選択します。
- プロファイルをダブルクリックして開きます。  
タブにプロファイルが開きます。
- 【データドメイン検出】** ビューまたは **【データドメイン】** ビューで、行を選択します。
- 推測されたデータドメインを却下するには、行を右クリックして **【却下】** を選択します。  
データドメイン検出の結果の中の拒否されたデータドメインが、Informatica Developer によって灰色に表示されます。
- 却下されたデータドメインを非表示にするには、その行を右クリックして **【却下されたものを非表示】** を選択します。
- 拒否されたデータドメインを表示するには、行のうち 1 つを右クリックして **【拒否済みを表示】** を選択します。

## Informatica Developer からのデータドメイン検出結果のエクスポート

Informatica Developer でデータドメイン検出結果を .xlsx にエクスポートすることで、そのファイルをサーバーや、クライアントマシン内の特定の場所に保存することができます。

- プロファイルを実行してデータドメイン検出を実行する。
- 【結果】** ビューをクリックします。
- 【ファイルへの結果のエクスポート】** アイコンをクリックします。  
**【ファイルへのデータのエクスポート】** ダイアログボックスが表示されます。
- ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
- 【保存】** で、**【クライアントに保存】** を選択し、**【参照】** をクリックして場所を選択し、ファイルをローカルコンピュータ上に保存します。デフォルトでは、Informatica Developer は Informatica Administrator の Data Integration Service プロパティに設定されているサーバ内の場所にファイルを書き込みます。
- 【OK】** をクリックします。

## 第 25 章

# Informatica Developer でのエンタープライズ検出

この章では、以下の項目について説明します。

- [Informatica Developer でのエンタープライズ検出の概要, 191 ページ](#)
- [エンタープライズ検出プロセス, 192 ページ](#)
- [エンタープライズ検出のプロファイルオプション, 193 ページ](#)
- [Informatica Developer でのエンタープライズ検出プロファイルの作成, 198 ページ](#)
- [プロファイルの編集, 199 ページ](#)
- [エンタープライズ検出プロファイルの実行, 200 ページ](#)
- [外部キー検出, 201 ページ](#)
- [結合分析, 202 ページ](#)
- [重複検出, 204 ページ](#)
- [DDL スクリプトファイル, 206 ページ](#)
- [エンタープライズ検出プロファイルの同期, 206 ページ](#)

## Informatica Developer でのエンタープライズ検出の概要

エンタープライズ検出は、大量のデータソース内のカラムプロファイル統計、データドメイン、プライマリキー、および外部キーを検出するプロセスです。エンタープライズ検出は、複数の接続またはスキーマに対して実行できます。

企業のデータアナリストは、大量のデータソースにまたがる重要なデータ特性を検出しなければならないことがあります。たとえば、リレーショナルデータアセットの特定、検出されたデータアセットに対するカラムプロファイルの実行、企業内の重要なデータ特性、プライマリキー、および候補キーの検出などが求められます。また、検出されたリレーションに基づいてデータモデルを導出できるように、データソースにまたがって存在する外部キーリレーションの表示が必要になることもあります。

エンタープライズ検出は、企業の情報資産に含まれる問題、パターン、傾向、および重要なデータ特性を検出します。モデルリポジトリにインポートしたデータソースと、外部リレーショナル接続の両方のデータソースを選択できます。データ検出プロセスには、カラムプロファイル統計、データドメイン分析、候補キーを含む



データオブジェクト構造、および外部キーを含むデータオブジェクトリレーションの検出が含まれます。Developer tool でエンタープライズ検出を実行すると、各データソースに対して次のタスクが実行されます。

- カラムプロファイルを実行する。
- データドメインを検出する。
- プライマリキーを推測する。

カラムプロファイル、データドメイン検出、およびプライマリキープロファイルの実行後、すべてのデータソースに対して外部キープロファイルが実行されます。プロファイリングと検出のタスクが完了すると、結果がサマリーにまとめられ、グラフィック形式とテーブル形式の両方で表示されます。

Informatica Developer でオペレーティングシステムプロファイルを選択できます。オペレーティングシステムプロファイルを選択すると、データ統合サービスによってエンタープライズ検出プロファイルが作成および実行されます。この処理は、オペレーティングシステムプロファイルに定義した、オペレーティングシステムユーザーの権限に基づいて行われます。

## エンタープライズ検出プロセス

Developer ツールでエンタープライズ検出プロファイルを実行して、エンタープライズ検出を行うことができます。プロファイルを実行する前に、プロファイルタイプごとにデータ検出オプションを設定する必要があります。

Developer ツールによって、選択したデータソースに対してデータオブジェクトが作成され、各データオブジェクトに対してプロファイルタスクが作成されます。次にプロファイルタスクが実行され、プロファイル結果が生成されます。

エンタープライズ検出を実行するには、以下の手順を実行します。

1. モデルリポジトリにインポートされた複数のデータオブジェクト、および複数の外部リレーショナル接続にまたがるデータソースを選択して、エンタープライズ検出プロファイルを作成します。
2. データドメイン検出、カラムプロファイル、プライマリキープロファイル、および外部キープロファイルの設定を定義します。
3. エンタープライズ検出プロファイルを実行します。
4. モデルリポジトリサービスを更新する。

**注:** 外部接続に対するメタデータのインポートがモデルリポジトリで発生するため、この操作を行う必要があります。Developer ツールによってモデルリポジトリの変更が反映されるように、モデルリポジトリサービスを更新する必要があります。

5. プロファイルの実行を監視します。また、Developer ツールによって実行されるプロファイルタスクのステータスを必要に応じて確認します。
6. エンタープライズ検出結果のサマリーを確認します。サマリーには、対話形式のグラフィカルビューとテーブル形式のビューが含まれます。



# エンタープライズ検出のプロファイルオプション

プロファイルを実行してエンタープライズ検出を実行する前に、プロファイルオプションを設定します。プロファイルオプションには、データドメイン検出オプション、カラムプロファイルサンプリングオプション、およびプライマリキーと外部キーの推測オプションが含まれます。

プロファイルオプションを設定したら、エンタープライズ検出プロファイルを実行できます。プロファイルの設定後、プロファイルを実行せずに、プロファイルタスクを作成することもできます。

## エンタープライズ検出でのデータドメインの選択

推測オプションでは、データドメイン検出の実行対象を、カラムデータ、カラム名、またはその両方から選択できます。データソース内のすべての行をプロファイルで処理する必要があるかどうかを指定し、データドメインの一致に適合条件を選択できます。

以下の表に、エンタープライズ検出に対して設定するデータドメイン推測オプションを示します。

オプション	説明
デフォルトの推測オプションをオーバーライド	定義済みの推測オプションを変更します。
データ	プロファイルをカラムデータに対して実行します。
カラム名	プロファイルをカラムタイトルに対して実行します。
データとカラム名	プロファイルをカラムデータとカラムタイトルの両方に対して実行します。
行の最小割合	データドメインの一致に必要なデータセット内の行の最小一致率。一致率は、一致行数を合計行数で除算した値です。 注: Developer tool では、NULL 値は一致しない行とみなされます。
行の最小数	データドメインの一致に必要なデータセット内の行の最小数。
データドメイン検出から NULL 値を除外	データドメイン検出用のデータセットから NULL 値を除外します。
すべての行	プロファイルをデータソースのすべての行に対して実行します。
サンプリング最初の	プロファイルを実行できる最大行数。Developer tool では、ソースの先頭行から行が選択されます。
以降のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します	次のプロファイル実行で、データ型とデータドメイン推測から、承認済みのデータ型とデータドメインを除外します。

## エンタープライズ検出のカラムプロファイルのサンプリングオプション

サンプリングオプションは、Developer tool でカラムプロファイルをデータソースのすべての行に対して実行するか、または一部の行に対して実行するかを指定します。

以下の表に、エンタープライズ検出に対して設定するカラムプロファイルサンプリング推測オプションを示します。

オプション	説明
すべての行	データオブジェクト内のすべての行でプロファイルを実行します。 ネイティブ、Blaze、および Spark ランタイム環境でサポートされます。
最初の<number>行をサンプリング	データオブジェクト内の行のうち、先頭からのサンプル行でプロファイルを実行します。行は最大で 2,147,483,647 行まで選択できます。 ネイティブおよび Blaze ランタイム環境でサポートされます。
Limit N <number>行	データオブジェクトの行数に基づいてプロファイルを実行します。Hadoop 検証環境でプロファイルを実行するように選択すると、Spark エンジンではデータオブジェクトの複数のパーティションからサンプルを収集し、これらのサンプルを単一のノードにプッシュしてサンプルサイズを計算します。Limit n のサンプリングオプションは、Oracle、SQL Server および DB2 データベースをサポートします。Limit n のサンプリングオプションには、詳細フィルタは適用できません。 Spark ランタイム環境でサポートされます。
ランダムパーセンテージ	データオブジェクト内の行の割合に対してプロファイルを実行します。 Spark ランタイム環境でサポートされます。
データ型が承認されているカラムについてはデータ型の推定を除外する	カラムプロファイル実行のデータ型推測から、承認されたデータ型のカラムを除外します。

## ランタイム環境オプション

[ネイティブ] または [Hadoop] ランタイム環境オプションを選択します。Hadoop ランタイム環境では、[Blaze] または [Spark] オプションを選択できます。Informatica Developer では、ランタイム環境の選択

後に、プロファイル定義にランタイム環境が設定されます。ランタイム環境はプロファイル結果には影響しません。

以下の表に、エンタープライズ検出プロファイルのランタイム環境オプションを示します。

オプション	説明
Native	Developer tool が、プロファイリングサービスモジュールにプロファイルジョブを送信します。次に、プロファイリングサービスモジュールが、プロファイルジョブを一連のマッピングに分割します。データ統合サービスがこれらのマッピングを実行し、プロファイル結果をプロファイルウェアハウスに書き込みます。
Blaze	データ統合サービスがプロファイルのロジックを Hadoop クラスタの Blaze エンジンにプッシュして、プロファイルを実行します。
Spark	データ統合サービスがプロファイルのロジックを Hadoop クラスタの Spark エンジンにプッシュして、プロファイルを実行します。

## エンタープライズ検出のプライマリキー推測オプション

エンタープライズ検出のデフォルトのプライマリキー推測オプションは上書きできます。オプションには、プロファイルを実行できる最大行数と最小一致率が含まれます。

以下の表に、エンタープライズ検出に対して設定するプライマリキー推測オプションを示します。

オプション	説明
デフォルトの推測オプションをオーバーライド	プライマリキー推測のカスタム設定を実行できる。
最大キーカラム数	プライマリキーを構成できる最大カラム数。
最大行数	プロファイルを実行できる最大行数。
最小割合	プライマリキーの一致に必要なカラムデータの最小一致率。
最大違反行数	プロファイルでプライマリキーの判別時に許容されるキー違反のある行の最大数。

## エンタープライズ検出の外部キー推測オプション

外部キー推測オプションを設定し、データオブジェクト間の外部キーリレーションを検出するためのカラム設定を定義します。外部キー推測結果は、エンタープライズ検出に対して設定したプライマリキー推測オプション、記述されているプライマリキー、およびユーザー定義のプライマリキーによって異なります。

次の方法のいずれかを使用して、Informatica Developer で外部キーを推測できます。

- デフォルト値を使用する。
- 外部キーの推測オプションを設定する。
- 外部キーの構成ファイルを使用して、自動キュレーションパラメータを設定する。

以下の表に、エンタープライズ検出に対して設定する外部キー推測オプションを示します。

オプション	説明
デフォルトの推測オプションをオーバーライド	定義済みの推測オプションを変更します。
比較に使用するデータ型	プライマリキーと外部キーの比較で使用されるデータ型。 注: このオプションは、外部キー推測の前にデータソースに対してカラムプロファイルを実行する場合に適用されます。
比較での大文字小文字の区別	カラムデータを比較するときに大文字小文字が区別されます。
比較の前に値を切り捨てる	Developer tool での処理時にカラムデータの先頭または末尾の空白を残すかどうかを指定します。
比較で使用されている推測されるプライマリキー 上位ランク _ 位までのキーの使用	Developer tool がすべてのデータソースに対して外部キープロファイルを実行する場合に、外部キー推測で、上位ランクの何位までのプライマリキーを使用するかを示す数。Developer tool では、上位ランク方式とともに、記述されているプライマリキーおよびユーザー定義プライマリキーを使用して、外部キーリレーションを推測します。 推測されるキーの上位ランクは、小数点第 2 位を丸めた一致率の高い順に決定されます。例えば、Developer tool では、一致率 99.75 を 99.8、99.74 を 99.7 とみなします。 デフォルト値は 1 です。Developer tool で、外部キー推測で推測されたすべてのキーを使用する場合は、値を -1 に設定します。 注: プライマリキーのデータソースでプライマリキーが承認された場合、Developer tool では外部キーの推測に、推測されるプライマリキーを使用しません。
データオブジェクト間の最大外部キー数	プロファイルの実行後に Developer tool によって返される、外部キー検出に必要な推測カラムの最大数。
最小一致率	外部キー結果にカラムを含める基準となる、適格性の最小値（パーセント）。
シングニチャの再生成	ソースデータが変更された場合に、カラムのシングニチャをリロードします。

## 外部キー推測の自動キュレーションパラメータ

自動キュレーションパラメータを設定して、手動操作なしにプライマリキーと外部キーのリレーションを推測できます。自動キュレーションパラメータは、ユーザー定義のカスタム属性で、特定の条件に基づいてデータリレーションを識別するために設定できます。

検出結果にプライマリキーと外部キーのリレーションが大量に含まれる場合に、何百ものデータリレーションから重要なデータリレーションを識別することが難しいことがあります。また、データの一致やデータ型などの特定の条件に基づいてリレーションを監督することが困難になる場合があります。この問題を解決するには、自動キュレーションパラメータを設定し、エンタープライズ検出プロファイルを実行します。

データソースに複数の外部キー候補があり、外部キー候補を選択するルールを指定する場合、次の操作を実行できます。

- エンタープライズ検出プロファイルウィザードで、[データオブジェクト間の最大外部キー] および [最小一致率] オプションを設定します。
- 自動キュレーションパラメータの重みおよびスコアを ForeignKeyConfig.xml ファイルに設定します。

管理者は外部キー構成ファイルを編集して保存できます。外部キー構成ファイルに自動キュレーションパラメータを設定してください。このアルゴリズムにより、自動キュレーションパラメータに基づいて、複数のデータオブジェクト間でプライマリキーと外部キーのリレーションが推測されます。

外部キー構成ファイルである ForeignKeyConfig.xml は、次のディレクトリにあります。

<Informatica インストールディレクトリ>\services\DataIntegrationService\modules\ProfilingService

自動キュレーションパラメータには、データ重複の一致、カラム名の一致、リレーションタイプの一致、およびデータ型の一致があります。

## データ重複の一致

データ重複の一致は、プライマリキーと外部キーの間にある値の推定の重複です。エンタープライズ検出プロファイルウィザードで **【最小一致率】** オプションを使用して重複の一致を設定できます。デフォルトでは、**【最小一致率】** オプションは 90 に設定されています。

データ重複の一致が最小一致率に満たない場合、外部キーは自動キュレーションの対象にはなりません。データ重複の一致が最小一致率を満たす場合、残りのパラメータは調整されたスコアの計算に使用されます。

## 名前の一致

名前の一致パラメータはオプションのパラメータです。エディット距離アルゴリズムを使用し、プライマリキーカラムと外部キーカラムの名前がどれだけ近いかに判断し、0~1 のスコアを設定します。このパラメータを使用してプライマリキーと外部キーのリレーションを判断しない場合、名前の一致のウェイトを 0 に設定します。

## リレーションタイプの一致

リレーションタイプの一致は、プライマリキーカラムと外部キーカラムの間のリレーションのタイプを判断し、0~1 の固定スコアを割り当てます。リレーションタイプの一致は、外部キーカラムのカラムタイプに基づいて計算されます。

次のリレーションタイプの一致は、ForeignKeyConfig.xml ファイルで設定できます。

- 外部キーカラムがキー以外のカラムである、プライマリキーと外部キーのリレーション。このリレーションタイプの一致のデフォルトは 1 です。このリレーションは多くのデータソースで確認できます。
- 外部キーカラムがプライマリキーカラムである、プライマリキーとプライマリキーのリレーション。このリレーションタイプのデフォルトは 0.25 です。このリレーションタイプは縦にパーティション化されているテーブルを表すため確認できるのはまれです。
- 外部キーカラムがプライマリキーカラムであり、カラムのデータ型がシーケンスデータ型である、プライマリキーとプライマリキーのシーケンスリレーション。例えば、Order テーブル OrderID カラムはシーケンスデータ型です。このリレーションタイプのデフォルトはゼロです。理由は、シーケンスキーによって誤検知の外部キーが複数発生する可能性があり、この障害をプライマリキーとプライマリキーのアルゴリズムが避けようとするためです。データソースにいくつかのシーケンスデータ型が含まれることが分かっている場合、リレーションタイプの一致を高いスコアに設定できます。

## データ型の一致

データ型の一致は、プライマリキーカラムと外部キーカラムのデータ型を比較し、カラムのデータ型がどれだけ近いかに基づいて規定の一致率を割り当てます。

以下の表に、プライマリキーおよび外部キーの各組み合わせに対するデータ型の規定のマッチ率を示します。

	数値外部キー	日付外部キー	文字列外部キー
数値プライマリキー	1.0	0.5	0.0
日付プライマリキー	0.5	1.0	0.5
文字列プライマリキー	0.0	0.0	1.0

必要に応じて、データ型のデフォルト一致率を変更できます。

# Informatica Developer でのエンタープライズ検出プロファイルの作成

プロファイルは、複数の接続の複数のデータソースに対して作成できます。Developer ツールは、ソースごとに個別のプロファイルタスクを作成します。

1. **【オブジェクトエクスプローラ】** ビューで、プロファイルを実行する複数のデータオブジェクトを選択します。
2. **【ファイル】 > 【新規】 > 【プロファイル】** をクリックしてプロファイルウィザードを開きます。
3. **【エンタープライズ検出プロファイル】** を選択して、**【次へ】** をクリックします。
4. プロファイルの名前を入力し、プロジェクトの場所を確認します。必要に応じて、新しい場所を参照します。
5. 選択したデータオブジェクトの名前が**【データオブジェクト】** セクション内に表示されることを確認します。さらにデータオブジェクトを選択する必要がある場合は、**【選択】** をクリックして選択します。
6. **【次へ】** をクリックします。  
**【プロファイル定義へのリソースの追加】** ペインが表示されます。このペインから複数の外部リレーショナル接続とデータソースを選択できます。
7. **【選択】** をクリックして、**【リソースの選択】** ダイアログボックスを開きます。  
**【リソース】** ペインに、Informatica ドメインの下に、内部接続、外部接続、およびデータオブジェクトがすべて一覧表示されます。
8. **【OK】** をクリックしてダイアログボックスを閉じます。
9. **【次へ】** をクリックします。
10. 実行するプロファイルタイプを設定します。以下のプロファイルタイプを設定できます。
  - データドメイン検出
  - カラムプロファイル
  - プライマリキープロファイル
  - 外部キープロファイル

**注:** エンタープライズ検出プロファイルの一部として実行するプロファイルタイプに対して **「[エンタープライズ検出プロファイルの実行] アクションの一部として有効化」** を選択します。 カラムプロファイリングはデフォルトで有効になっています。

11. プロファイルのオプションを確認します。  
カラムプロファイルではサンプリングオプションを編集できます。 また、データドメイン、プライマリキー、および外部キープロファイル用の推測オプションを編集できます。
12. **「プロファイルの作成」** を選択します。  
Developer ツールによって、個々のデータソースのプロファイルが作成されます。
13. プロファイル設定を完了したら、**「終了時にエンタープライズ検出プロファイルを実行」** を選択してプロファイルを実行します。 すべてのプロファイリング操作を有効にすると、選択したすべてのデータソースに対して、カラムプロファイル、データドメインプロファイル、およびプライマリキープロファイルが実行されます。 その後、すべてのデータソースに対して外部キープロファイルが実行されます。
14. **「完了」** をクリックします。  
エンタープライズ検出プロファイルを実行したら、結果を表示する前にモデルリポジトリサービスを更新する必要があります。 モデルリポジトリで外部接続に対してメタデータのインポートが行われるため、この手順が必要となります。 Developer ツールによってモデルリポジトリの変更が反映されるように、モデルリポジトリサービスを更新する必要があります。

## プロファイルの編集

エンタープライズ検出プロファイルの設定後、そのプロファイルに変更を加えることができます。 特定のデータ型のカラムを除外したり、カラム選択、データドメイン選択、および推測の各オプションを変更したりすることができます。

1. **「オブジェクトエクスプローラ」** ビューで、編集するプロファイルが含まれるプロジェクトまたはフォルダーを選択します。
2. **「チーム」** > **「チェックアウト」** をクリックして、プロファイルをチェックアウトします。
3. プロファイルをダブルクリックして開きます。
4. **「プロパティ」** ビューをクリックします。  
[プロパティ] ビューは [デフォルト] ビューの下にあります。
5. **「プロファイル」** をクリックして、プロファイルタスクを表示します。
6. 右側のペインで編集するプロファイルタスクを選択し、**「開く」** をクリックします。  
タブにプロファイル定義が表示されます。
7. エンタープライズ検出プロファイルのグローバル設定を変更するには、**「プロファイリングタスクのリスト」** 上部のプロファイルを選択して、**「設定」** をクリックします。
8. プロファイル定義オプションに必要な変更を加えます。
9. 変更を保存します。
10. **「チーム」** > **「チェックイン」** をクリックしてプロファイルをチェックインします。



# エンタープライズ検出プロファイルの実行

エンタープライズ検出プロファイルは複数の方法で実行できます。プロファイルは、[Object Explorer] ビューまたは [プロパティ] ウィンドウの [プロファイル] タブから実行できます。エンタープライズ検出プロファイルの一部を構成する個別または複数のプロファイルタスクを選択して実行できます。

1. [Object Explorer] ビューで、実行するプロファイルを含むプロジェクトまたはフォルダーを選択します。

プロファイルを自動的に実行するには、プロファイルの作成時に [新しいエンタープライズ検出] ウィザードで [終了時にエンタープライズ検出プロファイルを実行] を選択します。

2. プロファイルをダブルクリックして開きます。

タブにプロファイルが開きます。

3. [Object Explorer] ビューでプロファイルを右クリックし、[エンタープライズ検出プロファイルの実行] を選択します。

または、[プロパティ] ウィンドウで [プロファイル] を選択し、[プロファイリングタスクのリスト] でプロファイル名を選択して、[実行] をクリックすることもできます。

**注:** エンタープライズ検出プロファイルを実行する場合、結果を表示する前にモデルリポジトリサービスを更新する必要があります。モデルリポジトリで外部接続に対してメタデータのインポートが行われるため、この手順が必要となります。Developer ツールによってモデルリポジトリの変更が反映されるように、モデルリポジトリサービスを更新する必要があります。

4. [実行] ダイアログボックスが表示されます。このダイアログボックスで、プロファイルのグローバル設定に変更を加えることができます。

デフォルトでは、変更内容は、エンタープライズ検出プロファイルに新たに追加されるデータオブジェクトに適用されます。

5. エンタープライズ検出プロファイル内のすべてのデータオブジェクトプロファイルのタスクとシステムが生成した外部キープロファイルのタスクに変更を適用するには、[現在のプロファイルにグローバル設定を使用します] を選択します。

変更された設定に基づいて、すべてのデータオブジェクトプロファイルタスクと外部キープロファイルタスクが更新されます。

6. プロファイルタスクを個別に実行するには、タスクを選択し、[実行] をクリックします。

7. 複数のプロファイルタスクを実行するには、[複数実行] をクリックします。

[複数実行] ダイアログボックスが表示されます。

**ヒント:** エンタープライズ検出結果のロードに時間がかかる場合、プロファイリングウェアハウスデータベースの統計を更新することを推奨します。エンタープライズ検出プロファイルを複数実行すると、データの容量とカラムの値が大きく変わる可能性があります。統計を更新すると、データベースは最新の統計に基づいて SQL クエリの実行プランを実行し、データベースの操作を最適化します。

8. デフォルトではすべてのタスクが選択されています。実行しないタスクの選択を取り消し、[OK] をクリックします。



# 外部キー検出

カラムのデータ値が別のデータオブジェクトのプライマリキーカラム値と一致する場合、そのカラムは外部キーです。

Developer tool では、複数のデータオブジェクトに対して外部キー検出を実行できます。エンタープライズ検出プロファイルを作成してデータオブジェクトを選択し、プロファイルを定義します。

外部キー検出を実行する前に、エンタープライズ検出プロファイルで親データオブジェクトと子データオブジェクトを特定する必要があります。プロファイルでは、親オブジェクトの1つ以上のキー（プライマリキーを含む）を使用して、子オブジェクトの外部キーを検出します。親オブジェクトと子オブジェクトを定義し、親オブジェクトのキーを特定したら、プロファイルを作成して実行します。

## オブジェクトの親子関係の定義

2つのデータオブジェクト間の外部キー関係を検出するには、親データオブジェクトを選択し、そのオブジェクトのプライマリキーを指定する必要があります。

1. 分析するデータオブジェクトが含まれているエンタープライズ検出プロファイルを開きます。
2. 親オブジェクトを選択します。
3. 親オブジェクトのプライマリキーを選択します。
  - **【プロパティ】** タブをクリックし、**【キー】** をクリックします。
  - **【追加】** をクリックし、**【新しいキー】** ダイアログボックスでプライマリキーカラムを選択します。
  - **【新しいキー】** ダイアログボックスで **【OK】** をクリックします。 **【選択したフィールド】** ペインにプライマリキーが表示されていること、および **【プライマリキー】** オプションが選択されていることを確認します。

外部キープロファイルを作成して、子オブジェクトに外部キーがあるかどうかを分析します。

## データオブジェクト間の外部キー関係の検出

Developer ツールでエンタープライズ検出プロファイルを使用して、2つのデータオブジェクト間のキー関係を検出します。

プライマリキーを含むデータオブジェクトが親オブジェクトで、外部キーを含むデータオブジェクトが子オブジェクトです。

1. 分析するデータオブジェクトが含まれているエンタープライズ検出プロファイルを開きます。
2. データオブジェクトの名前を右クリックし、**【外部キープロファイル】** を選択します。
3. プロファイルの名前を入力し、プロジェクトの場所を確認します。必要に応じて、新しい場所を参照します。必要に応じて、プロファイルの説明を入力します。
4. プロファイルで子オブジェクトの外部キーを検出するために使用される親オブジェクトのキーを選択します。
5. プロファイルを保存して実行します。

## 外部キー分析結果

外部キープロファイルを実行したら、モデリングエディタの下にあるプロファイル名をクリックして分析の結果を表示します。

結果ビューには、定義したプライマリキーと外部キーの推測条件を満たすカラムが一覧表示されます。推測の設定を編集するには、**【オプション】** ボタンをクリックします。推測されたキーがデータオブジェクトの有効なキーであることを確認するには、カラム名をクリックして **【検証】** を選択します。

以下の表に、外部キー分析のプロパティを示します。

プロパティ	説明
親のプライマリキー	プロファイルで子オブジェクトの外部キーを検出するために使用された親データオブジェクトのプライマリキーカラム。
子の外部キー	プロファイルで、現在の行の親のプライマリキーに対する外部キーとして推測されたカラム。
算入%	プライマリキーと外部キーの間で一致するデータ値の数（割合で表される）。 注: 外部キーの結果に示されている推測されたカラムの <b>【算入%】</b> の値は、検証後の値とは異なる可能性がある。推測されたカラムの <b>【算入%】</b> は、親オブジェクトの一意のプライマリキーカラム値と一致する子オブジェクトの一意の外部キーカラム値の数である。推測されたカラムを検証すると、 <b>【算入%】</b> は、親オブジェクトのプライマリキーカラム値と一致する子オブジェクトの外部キーカラム値の数になる。
関係タイプ	プロファイルの実行前にプライマリキーカラムと外部キーカラムに対して定義したリレーションのタイプ。 プロファイルの実行前にリレーションを定義した場合は、算入の割合の数値がプロファイルに対して設定されている信頼度のしきい値を満たさない場合でも、そのリレーションのデータが返される。
検証済み	ユーザーがプライマリキーと外部キーのリレーションを検証したことを示す。
最終実行時刻	最後にプロファイルを実行した日時。
リレーションタイプ (モデル内)	プロファイルでカラム間のリレーションが検証されたことを示す。

## 結合分析

結合分析は、2つのデータカラム間に見込まれる結合度を示します。結合プロファイルを使用して、単一のデータソース内、または複数のデータソース全体のカラム結合を分析します。

結合プロファイルの結果は、ベン図および数値と割合の値で表示されます。エンタープライズ検出プロファイルから結合プロファイルを作成および実行します。

## 結合プロファイルの作成

エンタープライズ検出プロファイルのデータオブジェクト間の結合の可能性を分析できます。結合プロファイルでは、分析はモデルリポジトリに格納されます。

1. エンタープライズ検出プロファイルを作成するか、開きます。

2. エンタープライズ検出プロファイルに、必要なデータオブジェクトが含まれていることを確認します。  
結合プロファイルにデータオブジェクトを追加するには、そのオブジェクトを **[Object Explorer]** ビューからモデリングエディタにドラッグします。
3. プロファイリングするデータオブジェクトを選択します。
4. オブジェクトを右クリックし、**[結合プロファイル]** を選択します。  
プロファイルウィザードが開きます。
5. プロファイルの名前を入力します。必要に応じて、プロファイルの説明を入力します。
6. データオブジェクトの名前がウィザードの **[データオブジェクト]** に表示されることを確認します。
7. **[終了時にプロファイルを実行]** オプションを選択するか、選択を取り消します。
8. **[次へ]** をクリックします。
9. プロファイルに含めるデータカラムを選択し、**[次へ]** をクリックします。  
必要に応じて、データオブジェクトをスクロールダウンして使用可能なすべてのカラムを表示します。プロファイルはデフォルトですべてのカラムに対して実行されます。
10. **[追加]** をクリックします。  
**[結合条件]** ダイアログボックスが表示されます。
11. **[新規]** をクリックしてカラム選択フィールドをアクティブにします。
12. 検証するデータオブジェクトおよびカラムを選択します。  
2つのカラム間の結合条件を定義します。1つ以上のデータオブジェクトに対して複数の結合条件を定義できます。
13. **[OK]** をクリックして結合条件を作成します。  
必要に応じて、**[追加]** をクリックして追加の条件を定義します。
14. **[左]** と **[右]** の結合カラムに正しいデータオブジェクト名のプレフィックスが付いていることを確認します。
15. **[完了]** をクリックします。

## 結合分析結果

結合分析の **[結果]** タブには、親の孤立行、子の孤立行、および結合行の数と割合に関する情報が表示されます。また、結合分析結果には、カラム間のリレーションを示すベン図も含まれます。

次の表に、**[結果]** タブに表示されるプロパティを示します。

プロパティ	説明
左テーブル	左テーブルの名前と、結合分析で使用されたカラム。
右テーブル	右テーブルの名前と、結合分析で使用されたカラム。
左のみの行	左テーブルの結合できない行の数。
右のみの行	右テーブルの結合できない行の数。
行の結合	結合に含まれる行の数。

カラム間のリレーションを示すベン図を表示するには、結合条件を選択します。ベン図の下の領域にも、カラム内の孤立値、NULL 値、および結合値の数と割合が表示されます。

ベン図内のセクションをダブルクリックすると、そのセクションが表すレコードが表示されます。これらのレコードはデータビューアビューで開きます。

**注:** レコードの一覧をデータビューアビューからフラットファイルにエクスポートできます。

## 結合プロファイル結果のファイルへのエクスポート

結合条件に応じて返されたデータ行を区切りファイルにエクスポートできます。左と右のソース間の重複する行、またはソース内の孤立行をエクスポートします。

1. **[Object Explorer]** ビューで、結合分析を含むエンタープライズ検出プロファイルを開きます。
2. 結合プロファイルを実行します。
3. **[結合結果]** ビューを選択します。
4. **[データビューア]** タブで、**[ドリルダウン結果をファイルにエクスポート]** アイコンをクリックします。  
**[データのエクスポート]** ダイアログボックスが表示されます。
5. ファイル名を入力し、**[保存]** をクリックします。

## 重複検出

重複検出を行うと、1 つまたは複数のデータソース内のカラムのペア間で発生している重複データについての情報を得ることができます。重複データは、エンタープライズ検出プロファイルから検出できます。プロファイル結果の検証と表示は、ベン図で行うことができます。

重複検出では、デフォルトの設定または指定した設定に基づいて重複データを特定します。デフォルトの設定を上書きし、推測オプション（重複の割合に基づいて重複検出プロセスが返すトップペアの最大数など）を指定できます。また、重複検出に対する適格性を決定する信頼度レベルを指定することもできます。

## 重複検出の結果

**[重複検出]** タブには、関係するカラムについての情報と、重複の割合が表示されます。重複検出結果には、カラムのペア間の重複データと前回重複検出が行われた日時を示したベン図が含まれます。

カラムをクリックして **[検証]** を選択すると、結果をベン図で確認することができます。

次の表は、重複検出プロパティの説明です。

プロパティ	説明
左カラム	重複分析時に他のカラムと比較されるプライマリカラム。
右カラム	プライマリカラムと比較されるカラム。
重複の割合	2 つのカラム間の重複の割合。
検証済み	重複結果の行を検証したことを示す。
最終実行時刻	前回重複検出が行われた日時。

Informatica Developer は、それぞれの重複したペアを重複検出結果に 2 回表示します。データソースの ITEMS および ORDERS を検討します。ITEMS には「m」と「n」というカラムがあります。ORDERS には「p」と「q」というカラムがあります。

次のテーブルに、ITEMS と ORDERS で重複している検出結果を示します。

左カラム	右カラム
ITEMS	-
m	ORDERS.p
m	ORDERS.q
n	ORDERS.p
n	ORDERS.q
ORDERS	-
p	ITEMS.m
p	ITEMS.n
q	ITEMS.m
q	ITEMS.m

## 重複データの検出

カラムのペア間の重複データは、エンタープライズ検出プロファイルで特定できます。重複分析はカラム内の一意の値に基づいて行われ、NULL 値は考慮されません。

1. データオブジェクトを含むエンタープライズ検出プロファイルを作成または開きます。
2. 重複データを検出するデータオブジェクトを選択します。  
データオブジェクトを 1 つ選択してカラムのペア間の重複データを検出することも、複数のデータオブジェクトを選択することもできます。
3. オブジェクトを右クリックし、**【重複検出】** を選択します。  
**【新規重複検出】** ダイアログボックスが表示されます。
4. 名前を入力します。
5. 必要に応じて、その重複分析について説明するテキストを入力します。
6. データオブジェクトの名前がウィザードの **【データオブジェクト】** に表示されることを確認します。
7. 必要に応じて、設定を完了した時点でプロファイルを実行するように、**【終了時にプロファイルを実行】** を選択します。
8. **【次へ】** をクリックします。
9. 重複検出を行うカラムを選択します。
10. **【次へ】** をクリックします。  
ダイアログボックスに、デフォルトの推測オプションが表示されます。
11. 必要に応じて、重複検出の推測オプションを指定し、デフォルト設定を上書きします。
12. **【完了】** をクリックします。

# DDL スクリプトファイル

データ定義言語（DDL）スクリプトファイルには、Create、Alter、および Drop の各 SQL 文を指定します。

スクリプトファイルを生成する場合、ファイル名、場所、およびターゲットデータベースタイプを指定できます。Developer ツールはスクリプトファイル名に "\_create" や "\_drop" ラベルを追加します。仮想カラムは DDL スクリプトファイルには含まれません。

## エンタープライズ検出プロファイルからの DDL スクリプトの作成

エンタープライズ検出プロファイルから DDL スクリプトファイルを生成する場合、スクリプトファイルを保存する場所を選択できます。また、スクリプトの実行対象にするデータベースタイプも選択できます。DDL スクリプトを生成する前に、必ずエンタープライズ検出プロファイル内の必要な変更をすべて確認し、確定します。

1. **[Object Explorer]** ビューで、エンタープライズ検出プロファイルを選択します。
2. プロファイルを右クリックし、**[DDL の生成]** を選択します。  
**[DDL の生成]** ダイアログボックスが表示されます。
3. **[参照]** をクリックして、**[名前を付けて保存]** ダイアログボックスを開きます。  
デフォルトのファイル拡張子は .sql です。
4. ファイルの場所を選択し、ファイル名を入力します。
5. ターゲットデータベースのタイプを選択します。
6. **[OK]** をクリックします。

指定した場所に、Developer ツールにより DDL スクリプトファイルが生成されます。

## エンタープライズ検出プロファイルの同期

エンタープライズ検出プロファイルは、Developer tool で同期できます。

バージョン 9.5 以前からバージョン 9.6 以降にアップグレードした後、以前のバージョンからアップグレードしたバージョンにプロファイルを移行できます。エンタープライズ検出プロファイルの場合、以前のバージョンでユーザー定義キー、文書化されたキー、またはリレーションシップを追加した場合、キーとリレーションシップ情報は、プロファイルウェアハウスではなく、モデルリポジトリ内でのみ存続します。アップグレードされたバージョンでは、Developer tool でエンタープライズ検出プロファイルを開くと、文書化された、またはユーザー定義のキーとリレーションシップがプロファイルのキュレートされた結果には含まれません。

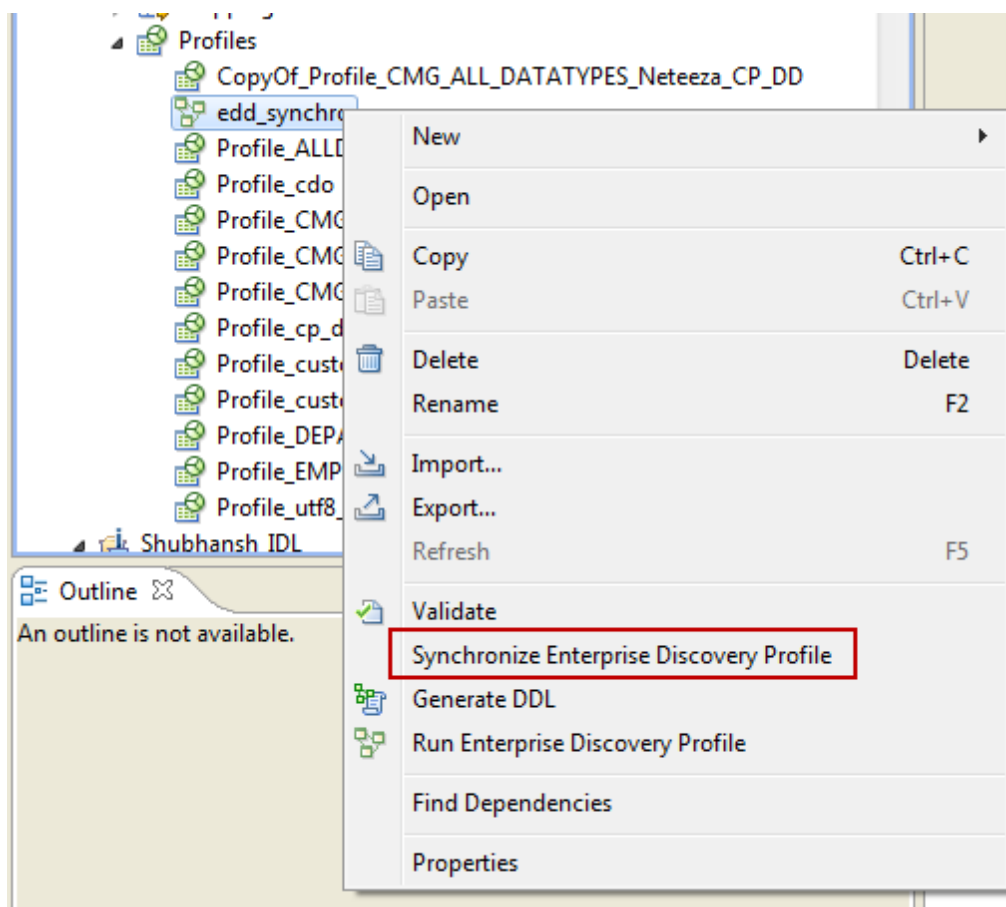
モデルリポジトリ内のユーザー定義キー、文書化されたキー、およびリレーションシップをプロファイリングウェアハウスに同期するには、Developer tool の **[エンタープライズ検出プロファイルを同期]** オプションを使用します。エンタープライズ検出プロファイルを同期した後、ユーザー定義と文書化されたキーおよび関係が **[承認]** に設定され、Developer tool でキュレートされた結果を表示できます。

## エンタープライズ検出プロファイルの同期

Informatica Developer は、バージョン 9.5 以前からバージョン 9.6 以降にアップグレードした後、エンタープライズ検出プロファイルのキュレートされた結果を同期することができます。

1. **[Object Explorer]** ビューで、エンタープライズ検出プロファイルを選択します。

2. プロファイルを右クリックし、**「エンタープライズプロファイルを同期」** を選択します。
- 次の図は、Developer tool の「エンタープライズ検出プロファイルを同期」 オプションを示しています。



プロファイルのキュレートされた結果が同期されます。

## 第 26 章

# エンタープライズ検出結果

この章では、以下の項目について説明します。

- [エンタープライズ検出結果の概要, 208 ページ](#)
- [\[リレーション\] ビュー, 209 ページ](#)
- [外部キープロファイリングビュー, 210 ページ](#)
- [テーブルビュー, 213 ページ](#)
- [\[データドメイン\] ビュー, 215 ページ](#)
- [\[カラムプロファイル\] ビュー, 216 ページ](#)
- [エンタープライズ検出の実行中のカラムプロファイル結果の表示, 216 ページ](#)
- [エンタープライズ検出の実行中のデータドメイン検出結果の表示, 217 ページ](#)
- [エンタープライズ検出のランタイムステータスの表示, 217 ページ](#)
- [エンタープライズ検出エクスポートファイル, 218 ページ](#)

## エンタープライズ検出結果の概要

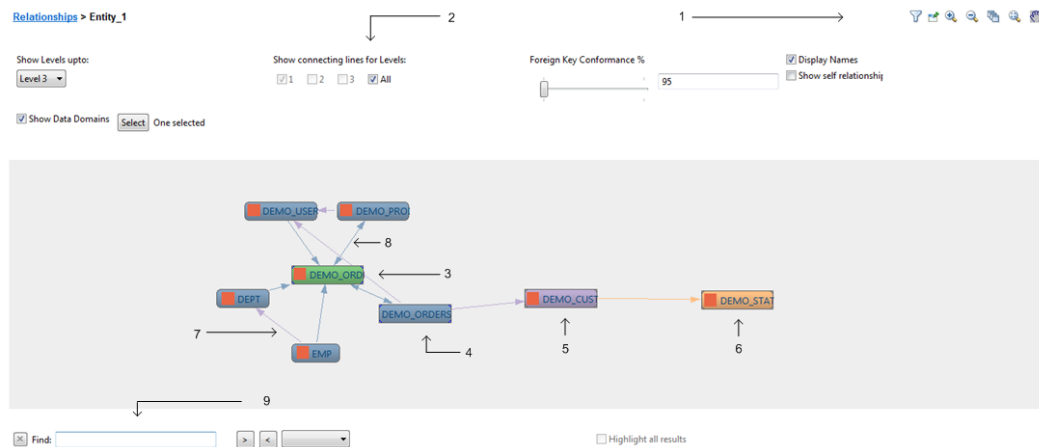
エンタープライズ検出結果は複数のビューで表示できます。ビューには、**[リレーション]**、**[データドメイン]**、**[カラムプロファイル]**、**[結合プロファイル]**、および**[重複検出]**の5つがあります。

**[リレーション]** ビューでは、データオブジェクトのグループが円で表示されます。このビューから外部キープロファイル結果を表示できます。外部キープロファイル結果は、グラフィカルビューとテーブルビューの両方で表示できます。**[データドメイン]** ビューには、データドメイン検出結果が表示されます。**[カラムプロファイル]** ビューには、各データオブジェクトのカラムプロファイル結果が表示されます。**[結合プロファイル]** ビューには、親の孤立行数、子の孤立行数、および結合に含まれる行数が表示されます。**[重複検出]** ビューには、関係するカラムについての情報と、重複の割合が示されます。

データオブジェクトは、互いに複数のリレーションを持つことができます。外部キー結果のグラフィカルビューには、一致率が最も高いデータオブジェクトリレーションが表示されます。



次の図に、エンタープライズ検出結果からのいくつかのサンプルデータオブジェクトのグラフィカルビューを示します。



1. ツールバーアイコン - フィルタ、データオブジェクトを固定、ズームイン、ズームアウト、すべて整列、ウィンドウに合わせる、パン。
2. フィルタオプション - 様々なレベルのデータオブジェクトリレーションの表示、結合ラインの表示、自己関連データオブジェクトの表示など。
3. ビジュアルエディタがデータオブジェクト間のその他のリレーションを表示するのに使用する、選択されているデータオブジェクト。
4. データオブジェクトリレーションの最初のレベル。
5. データオブジェクトリレーションの 2 番目のレベル。
6. データオブジェクトリレーションの 3 番目のレベル。
7. データオブジェクト間のコネクタ。矢印が 1 つのコネクタは、2 つのデータオブジェクト間のプライマリキーから外部キーへのリレーションを示します。矢印はプライマリキーのあるデータオブジェクトを指しています。
8. データオブジェクト間のコネクタ。矢印が 2 つのコネクタは、2 つのデータオブジェクト間のプライマリキーからプライマリキーへのリレーションを示します。コネクタ上にカーソルを置くと、推測されるリレーションに対して一致率が最大の結合カラムが表示されます。
9. Ctrl+F キーを押して [検索] フィールドを表示し、ワイルドカード文字としてアスタリスク (\*) を使用してグラフィカルビュー中のデータオブジェクトを検索します。

## [リレーション] ビュー

[リレーション] ビューでは、エンタープライズ検出結果の概要をエンティティを含めて表示できます。エンティティはデータオブジェクトのグループで、円で表されます。エンティティには、ソースデータベースの複数の接続とスキーマの、リレーションのあるデータオブジェクト、自己リレーションのデータオブジェクト、およびリレーションのないデータオブジェクトがあります。

自己リレーションデータオブジェクトは、そのデータオブジェクト内のカラム間にリレーションがあります。リレーションのないデータオブジェクトは、ソースデータベース内の他のデータオブジェクトとの間にも、データオブジェクト内のカラム間にもリレーションがありません。エンタープライズ検出結果のデータオブジェクトのエンティティのリレーション図は、データソースに記述されているリレーションではなく、推測されるリレーションに基づいています。

## データオブジェクトの検索

データオブジェクトは、[リレーション] ビューまたは [外部キープロファイリング] ビューで検索できます。アスタリスク (\*) をワイルド文字として使用して、データオブジェクトを検索することができます。

1. [リレーション] ビューまたは [外部キープロファイリング] ビューを表示していることを確認します。
2. 検索するデータオブジェクト名の一部を入力し、検索する内容に応じて検索文字列の先頭または末尾に\*ワイルドカード文字を追加します。例えば、文字列「CA」で始まるすべてのデータオブジェクトを検索するには、「CA\*」と入力し、**Enter** キーを押します。名前に文字列「ZIP」が含まれるすべてのデータオブジェクトを検索するには、「\*ZIP\*」と入力します。  
検索では、大文字と小文字が区別されます。

## [外部キープロファイリング] ビューへの移動

[外部キープロファイリング] ビューには、プロファイルを実行した複数のデータオブジェクトにまたがる外部キーリレーションがまとめて表示されます。ビュー内の円は、エンティティ、自己リレーションオブジェクト、および参照されていないオブジェクトを表します。

1. [リレーション] タブが表示されていることを確認します。  
右側のペインに外部キープロファイルのリンクが表示されています。
2. [外部キープロファイル] をクリックして、ビューを開きます。  
新しいタブにビューが表示されます。このビューでは、データオブジェクトが、それぞれのリレーションタイプに基づいて異なるセットの円に表示されます。統合された外部キービューに含まれるデータオブジェクトの合計数も表示されます。
3. 必要に応じて、[リレーション] リンクをクリックして [リレーション] ビューに戻ることができます。

## 外部キープロファイリングビュー

[外部キープロファイリング] ビューでは、エンタープライズ検出結果のサマリーをグラフィック形式で表示できます。データオブジェクトのプロファイル結果とカラムレベルのリレーションをテーブル形式でビューから開くことができます。

データオブジェクトのプロファイル結果には、カラムプロファイル、プライマリキー推測、機能依存性の推測、およびデータドメイン検出結果が含まれます。データオブジェクトのカラムレベルのリレーションを開いたら、データリレーションを検証およびキューレーションすることができます。データリレーションを検証する際、Developer ツールにより、ソースデータの全行に対してプロファイルが実行されて、推測結果が検証されます。[外部キープロファイリング] ビューでは、データリレーションを承認、拒否、およびリセットすることができます。

**注:** Hive データソースを使用してエンタープライズ検出プロファイルを作成した場合、[外部キープロファイリング] ビューにデータオブジェクトが表示されません。

## データオブジェクトリレーションの表示

データオブジェクト間のリレーションはグラフィカル形式で表示できます。エンティティの円をダブルクリックすると、エンティティ内のテーブルとそのリレーションが表示されます。

1. [外部キープロファイリング] ビューが表示されていることを確認します。
2. データドメインを統合グラフィカルビューに含めるには、[データドメインを表示] を選択します。

【選択】 ボタンが有効になります。

3. 【選択】 をクリックして、グラフィカルビューに含めるデータドメインを選択します。

【データドメインの選択】 ダイアログボックスが表示されます。

4. 必要なデータドメインを選択して、【OK】 をクリックします。

Developer ツールにより、選択したデータドメインを含むエンティティの円が強調表示されます。

5. エンティティの円をダブルクリックすると、エンティティ内のテーブルのリレーションが視覚的に表示されます。テーブルは、各データオブジェクトがエンティティ内の他のデータオブジェクトに対して持つリレーションを表したグラフィカルなビューで表示されます。

他のデータオブジェクトとのリレーション数が最大のデータオブジェクトまたは移動を開始するデータオブジェクトが緑色で強調表示されます。データドメインを含めた場合、データドメインの選択が各データオブジェクトの視覚表示の左側に強調表示されます。

6. 【外部キープロファイリング】 ビューの右側のペインで、直接リレーション情報とデータドメイン情報を確認します。
7. 必要に応じて、【リレーション】 リンクをクリックして【リレーション】 ビューに戻ることができます。

## ビューのズームインとズームアウト

【外部キープロファイリング】 ビューでは、見やすくするために、データオブジェクトリレーションのグラフィカルビューをズームインすることができます。ズームインすると、画像が拡大されます。ズームアウトすると、画像が縮小されます。

1. 【外部キープロファイリング】 ビューが表示されていることを確認します。
2. 画像の拡大レベルを増加させるには、ビューを右クリックして【ズームイン】を選択します。
3. グラフレイアウトの拡大レベルを減少させるには、【ズームアウト】を選択します。

## データオブジェクトの検索

外部キー結果のグラフィカルビューから、データオブジェクトを検索できます。アスタリスク (\*) をワイルドカード文字として使用して、データオブジェクトを検索することができます。

1. 【外部キープロファイリング】 ビューが表示されていることを確認します。  
必ずこのビューで外部キー結果のグラフィカルビューを開きます。
2. Ctrl+F を押して、【検索】 フィールドを表示します。
3. 【検索】 フィールドで、検索するデータオブジェクト名の一部を入力し、検索する内容に応じて検索文字列の先頭または末尾に \* ワイルドカード文字を追加します。例えば、文字列「EMP」で始まるすべてのデータオブジェクトを検索するには、「EMP\*」と入力し、**Enter** キーを押します。名前に文字列「ZIP」が含まれるすべてのデータオブジェクトを検索するには、「\*ZIP\*」と入力します。
4. 次の一致データオブジェクトに移動する場合は、【次の一致】 ボタンをクリックします。  
前の一致データオブジェクトに移動する場合は、【前の一致】 ボタンを使用します。
5. すべての一致データオブジェクトを強調表示する場合は、【すべての結果を強調表示する】を選択します。
6. 【検索】 フィールドの検索文字列を消去するには、フィールドの横の【消去】 ボタンをクリックします。

## カラムリレーションの表示

データオブジェクト内の各カラムと関連データオブジェクト内のカラムのリレーションを表示できます。データオブジェクトリレーションを検証し、データモデルにコミットすることもできます。

1. 【外部キープロファイリング】 ビューが表示されていることを確認します。

2. データオブジェクトを右クリックし、**【カラムリレーションの表示】** を選択します。  
カラムリレーションはテーブルビューで表示されます。ビューには、ソースデータオブジェクト、関連データオブジェクト、関連データオブジェクトのカラムなどのリレーション情報が表示されます。
3. 推測ステータス、検証ステータス、およびキュレーションステータスを確認します。
4. 親エンティティ内のすべてのデータオブジェクトとそれらのカラムリレーション情報を表示するには、**【グループのすべてのデータオブジェクト】** を選択します。  
デフォルトでは、ビュー内で選択したデータオブジェクトのリレーション情報が表示されます。
5. 必要に応じて、キャンバスの上部にある **【エンティティ】** リンクをクリックして、データオブジェクトのグラフィカルビューに戻ることができます。

## イメージとしてのエンティティリレーションダイアグラムの保存

エンタープライズ検出結果内のデータオブジェクトのエンティティリレーションダイアグラムを ".png" ファイルとして保存できます。

1. プロファイルを実行してエンタープライズ検出を実行します。
2. **【外部キープロファイリング】** ビューが表示されていることを確認します。
3. ビューで、グラフィック形式のデータオブジェクトリレーションに切り替えます。
4. 右クリックして、**【イメージとして保存】** を選択します。  
**【名前を付けて保存】** ダイアログボックスが表示されます。デフォルトでは、イメージは ".png" ファイルとして保存されます。
5. ファイルの場所を選択し、ファイル名を入力します。
6. **【保存】** をクリックします。

## 【外部キープロファイリング】 ビューからのデータオブジェクトプロファイル結果の表示

**【外部キープロファイリング】** ビューから、選択したデータオブジェクトのカラムプロファイル、プライマリキー、およびデータドメイン検出結果を表示することができます。データオブジェクトを選択するには、必ずキャンバス内で選択テーブルとして固定してください。

1. **【外部キープロファイリング】** ビューが表示されていることを確認します。
2. データオブジェクトを右クリックし、**【データオブジェクトをフォーカスとして固定】** を選択してテーブルを選択します。  
または、**【固定】** アイコンを使用してデータオブジェクトを選択します。
3. キャンバス内を右クリックし、**【データオブジェクトプロファイルの表示】** を選択します。  
データオブジェクトプロファイル結果がタブに表示されます。

# テーブルビュー

**【外部キープロファイリング】** ビューを開くと、デフォルトでは結果がグラフィカルビューで表示されます。テーブルとリレーションの詳細をテーブル形式で表示するには、テーブル表示に切り替えます。

エンティティ内のデータオブジェクトの数、関連テーブルの名前、それらの接続情報、および2つのデータオブジェクト間のリレーションの数を表示できます。カラムリレーションを検証し、それらをデータモデルに追加することもできます。

## [テーブル詳細] ペイン

データオブジェクトの詳細は、エンタープライズ検出結果のグラフィカルビューとテーブルビューの両方で表示できます。グラフィカルビューでは、[テーブル詳細] ペインに、選択したデータオブジェクトとの直接リレーションのあるデータオブジェクトの数とデータオブジェクト名が表示されます。

以下の表に、テーブルビューの [テーブル詳細] ペインのカラムを示します。

カラム名	説明
テーブル名	左側のペインで選択したデータオブジェクトと直接リレーションのあるデータオブジェクトの名前。
接続	関連データオブジェクトの接続の名前。
リレーション	左側のペインで選択したデータオブジェクトと [テーブル詳細] ペインの関連データオブジェクトの間のリレーションの数。

## エンタープライズ検出結果の検証

エンタープライズ検出結果を検証すると、データソースのすべての行に対してプロファイルが実行されます。検証後、一致率の値は、データソース内のすべての行でカラム値に応じて変化する可能性があります。

1. プロファイルを実行後、開きます。
2. **【外部キープロファイリング】** ビューが表示されていることを確認します。
3. ビューの上部にある **【テーブル表示】** アイコンをクリックします。  
テーブル表示では左側のペインにエンティティが表示されます。
4. 左ペインでデータオブジェクトを右クリックし、**【カラムリレーションの表示】** を選択します。  
選択したデータオブジェクト内のカラムと他のデータオブジェクト内のカラムのリレーションを表示できます。右にスクロールすると、リレーションタイプ、一致率、検証ステータス、コミットステータスなどの詳細が表示されます。
5. 行を右クリックし、**【検証】** を選択します。  
**【プロファイルの実行】** ダイアログボックスが表示されます。検証の完了後、行を選択し、プライマリキーと外部キーのリレーションの重なりをベン図に表示します。

## カラムのリレーションのキューレーション

**【外部キープロファイリング】** ビューでは、データリレーションを承認、拒否、およびリセットすることができます。

1. プロファイルを実行後、開きます。

2. **【外部キープロファイリング】** ビューが表示されていることを確認します。
3. グラフィカルビューで推測されたカラムリレーションを拒否するには、データオブジェクトを選択し、右クリックして次のオプションのいずれかを選択します。
  - **【次の対象とのすべてのリレーションを拒否する】 > 【推測されるプライマリキー】**。このオプションを選択すると、データオブジェクト内で推測されるプライマリキーを持つカラムと、他の接続されたデータオブジェクト内で推測される外部キーを持つカラムのすべてのリレーションを拒否できます。
  - **【次の対象とのすべてのリレーションを拒否する】 > 【推測される外部キー】**。このオプションを選択すると、データオブジェクト内で推測される外部キーを持つカラムと、他の接続されたデータオブジェクト内で推測されるプライマリキーを持つカラムのすべてのリレーションを拒否できます。
  - **【次の対象とのすべてのリレーションを拒否する】 > 【推測されるプライマリキーと外部キー】**。このオプションを選択すると、データオブジェクト内で推測されるプライマリキーを持つカラムと、他の接続されたデータオブジェクト内で推測される外部キーを持つカラムのすべてのリレーションを拒否できます。また、データオブジェクト内で推測される外部キーを持つカラムと、他の接続されたデータオブジェクト内で推測されるプライマリキーを持つカラムのすべてのリレーションを拒否できます。
4. グラフィカルビューでデータオブジェクトを右クリックし、**【カラムリレーションの表示】** を選択します。
5. キュレーションするデータオブジェクトのリレーションを選択します。
6. カラムのリレーションを承認するには、右クリックして **【承認】** をクリックします。  
行のステータスが **【承認済み】** に変わります。
7. カラムのリレーションのステータスを推測済みに戻すには、右クリックしてから **【リセット】** をクリックします。
8. 却下されたカラムのリレーションを表示するには、行のうち1つを右クリックしてから **【拒否済みを表示】** を選択します。
9. 拒否されたデータ型を非表示にするには、行のうち1つを右クリックして、**【拒否済みを非表示】** を選択します。

## モデルリポジトリへの結果のコミット

プロファイルの実行後、データオブジェクト間のカラムリレーションをモデルリポジトリに保存できます。**【外部キープロファイリング】** ビューのテーブルビューから、モデルリポジトリにリレーションをコミットすることができます。

1. プロファイルを実行後、開きます。
2. **【外部キープロファイリング】** ビューが表示されていることを確認します。
3. 左ペインでデータオブジェクトを右クリックし、**【カラムリレーションの表示】** を選択します。  
選択したデータオブジェクト内のカラムと他のデータオブジェクト内のカラムのリレーションを表示できます。
4. 行を右クリックし、**【承認】** を選択します。

# [データドメイン] ビュー

[データドメイン] ビューには、Developer ツールでエンタープライズ検出の一部として検出されるデータドメインおよび一致するカラム統計が一覧表示されます。 [データドメイン] ビューから、カラムを検証したり、行をドリルダウンしたり、データオブジェクトプロファイル結果を表示したりすることができます。

## データドメイン検出結果の表示

データドメイン検出結果は、[データドメイン] タブで表示できます。データドメインを検索し、データドメイングループでソートして表示することができます。

1. プロファイルを実行してエンタープライズ検出を実行します。
2. プロファイルを開きます。
3. [データドメイン] タブをクリックして、データドメイン検出結果を表示します。  
データオブジェクトプロファイル結果が右側のペインに表示されます。
4. 検索フィールドにデータドメイン名の一部を入力して、特定のデータドメインを検索します。  
Select [データドメイングループの階層を表示] を選択して、データドメイングループでデータドメインのリストをソートします。

## データドメイン検出結果の検証

プロファイルを実行すると、データソースのサンプルが分析されて、プロファイル結果が推測されます。ソースデータのすべての行についてプロファイルを実行して、推測結果を検証できます。

1. プロファイルを実行後、開きます。
2. [データドメイン] タブをクリックして、結果を表示します。  
データドメイン検出結果が右側のパネルに表示されます。
3. 右側のパネルで、検証するカラムを選択します。
4. カラムを右クリックし、[検証] を選択して、データソースのすべての行に対してプロファイルを実行します。  
結果の検証後、[データ一致率] の値または [行数を合わせる] の値の変化を確認できる場合があります。

## 行のドリルダウン

さらにデータを分析するために、データドメイン検出結果をドリルダウンすることができます。

1. プロファイルを実行後、開きます。
2. [データドメイン] タブをクリックして、結果を表示します。  
データドメイン検出結果が右側のパネルに表示されます。
3. 右側のパネルで、ドリルダウンする行を選択します。
4. カラムを右クリックし、[ドリルダウン] を選択してソース行をドリルダウンします。



## [データドメイン] ビューからのデータオブジェクトプロファイル結果の表示

[データドメイン] ビューから、選択したデータオブジェクトのデータオブジェクトプロファイル結果を表示することができます。

1. [データドメイン] ビューが表示されていることを確認します。
2. [プロファイルされたドメイン] ペインでデータドメインを選択します。
3. 右側の [カラム] ペインで、カラムを選択します。
4. カラムを右クリックし、[データオブジェクトプロファイルを開く] を選択します。  
データオブジェクトプロファイル結果がタブに表示されます。

## [カラムプロファイル] ビュー

[カラムプロファイル] ビューには、Developer ツールが Enterprise Discovery の一環として実行する単一データオブジェクトプロファイルのカラムプロファイル結果のサマリが表示されます。データオブジェクト内の各カラムの一意の値、NULL 値、データ型、最大値と最小値などのカラム統計を表示できます。

### データオブジェクトプロファイル結果の表示

エンタープライズ検出には、データオブジェクトプロファイルの実行によるカラムデータ統計、プライマリキーと候補キー、およびデータドメインの検出が含まれます。[カラムプロファイル] ビューから、選択したデータオブジェクトのデータオブジェクトプロファイル結果を表示することができます。

1. [カラムプロファイル] ビューが表示されていることを確認します。
2. [プロファイルされたデータオブジェクト] ペインでデータオブジェクトを選択します。
3. 右側の [カラム] ペインで、カラムを選択します。
4. カラムを右クリックし、[データオブジェクトプロファイルの表示] を選択します。  
データオブジェクトプロファイル結果がタブに表示されます。デフォルトで、カラムプロファイル結果が表示されます。
5. プライマリキープロファイルを表示するには、[プライマリキー推測] をクリックします。
6. 機能依存性検出結果を表示するには、[機能依存性推測] をクリックします。
7. データドメイン検出結果を表示するには、[データドメイン検出] をクリックします。

## エンタープライズ検出の実行中のカラムプロファイル結果の表示

エンタープライズ検出の完了にかかる時間は、プロファイルタスクの数、データソースのサイズ、およびプロファイルタイプによって異なります。Developer ツールでデータ検出タスクが実行されている間、データ検出の初期段階で完了したカラムプロファイルの結果を表示できます。

1. プロファイルの実行後、[プロパティ] ウィンドウの [プロファイル] をクリックします。



2. 結果を表示するカラムプロファイルを選択します。【プロパティ】ウィンドウで、プロファイル実行のステータスが【成功】であることを確認します。
3. 【開く】をクリックして、別のタブに結果を表示します。
4. 【結果】セクションで、【カラムプロファイリング】を選択して右側のペインに結果を表示します。

## エンタープライズ検出の実行中のデータドメイン検出結果の表示

Developer ツールでエンタープライズ検出の一部であるデータ検出タスクの実行中、エンタープライズ検出の初期段階で完了したデータドメイン検出の結果を表示できます。

1. プロファイルの実行開始後、【プロパティ】ウィンドウの【プロファイル】をクリックします。
2. データドメイン結果を表示するプロファイルを選択します。【プロパティ】ペインで、プロファイル実行のステータスが【成功】であることを確認します。
3. 【開く】をクリックして、別のタブに結果を表示します。
4. 【結果】セクションで、【データドメイン検出プロファイリング】を選択して右側のペインに結果を表示します。

## エンタープライズ検出のランタイムステータスの表示

Developer ツールの【進行状況】ビューに、プロファイル実行などの操作の進行状況が表示されます。【進行状況】ビューから、エンタープライズ検出タスクのランタイムステータスを表示することができます。

1. プロファイルを実行してデータソースでエンタープライズ検出を実行した後、Developer ツールの右下隅にある【進行状況ビュー】ボタンをクリックします。  
【進行状況】ペインがまだ開かれていなかった場合は、表示されます。
2. 【エンタープライズ検出実行中:タスクステータスの表示】リンクをクリックして、サブタスクダイアログボックスを開きます。  
ダイアログボックスには、エンタープライズ検出の一部である複数のプロファイルタスクが一覧表示されます。プロファイル名、プロファイルタイプ、およびプロファイルのステータスを表示できます。
3. カラムヘッダをクリックして、プロファイルタスクをソートします。例えば、プロファイルタスクをそのステータスでソートするには、【ステータス】カラムヘッダをクリックします。
4. 特定のプロファイルタスクをキャンセルする必要がある場合は、そのタスクを選択し、【キャンセル】をクリックします。  
キャンセルしたタスクのステータスは【ターミネート済み】に変わります。

# エンタープライズ検出エクスポートファイル

エンタープライズ検出プロファイルの実行後、すべてのデータオブジェクトリレーション、データドメイン、および個々の外部キータスク結果などの情報をエクスポートすることができます。データオブジェクトリレーションのグラフィカルイメージは .jpg ファイルとして保存できます。

プロファイル結果をエクスポートすると、Developer ツールにより、すべてのエンタープライズ検出結果が複数の Microsoft Excel ファイルに保存されます。データオブジェクトリレーション、カラムプロファイル結果、データドメイン検出結果、エンティティ、および個々の外部キータスク結果は、個別のファイルで表示できます。

## エンタープライズ検出結果のエクスポート

エクスポートできるのは、エンティティのリスト、各エンティティの構成、エンティティの全データオブジェクトおよびカラムレベルデータオブジェクトリレーション、データドメイン、ならびに カラムプロファイリング結果です。

1. プロファイルを実行してエンタープライズ検出を実行します。
2. **[リレーション]**、**[データドメイン]**、または **[カラムプロファイル]** ビューで、ウィンドウの右上領域にある **[エクスポート]** アイコンをクリックします。  
**[ファイルへのデータのエクスポート]** ダイアログボックスが表示されます。
3. ファイル名を入力します。必要に応じて、デフォルトのファイル名を使用します。
4. **[保存]** で、**[クライアントに保存]** を選択し、**[参照]** をクリックして場所を選択し、ファイルをコンピュータ上にローカルに保存します。デフォルトでは、Informatica Developer は Informatica Administrator の Data Integration Service プロパティに設定されている場所にファイルを書き込みます。
5. **[OK]** をクリックします。

## 第 27 章

# Informatica Developer での Business Glossary Desktop

この章では、以下の項目について説明します。

- [ビジネス用語集の検索, 219 ページ](#)
- [ビジネス用語のルックアップ, 219 ページ](#)
- [ビジネス用語のルックアップ用にホットキーをカスタマイズ, 220 ページ](#)

## ビジネス用語集の検索

Developer ツールのオブジェクト名の意味を、ビジネス用語として Business Glossary Desktop でルックアップし、そのビジネス要件と現在の実装を理解します。

ビジネス用語集とは、ビジネスユーザー向けに、ビジネス言語を使用して概念を定義した用語の集まりです。ビジネス用語は、概念のビジネス上の定義および使用方法を説明したものです。Business Glossary Desktop は、ビジネス用語集をホストする Metadata Manager サービスに接続するクライアントです。Business Glossary Desktop を使用して、ビジネス用語集内のビジネス用語をルックアップします。

Business Glossary Desktop がご使用のマシンにインストールされている場合は、Developer ツールでオブジェクトを選択し、ホットキーか [検索] メニューを使用して、ビジネス用語集でオブジェクトの名前をルックアップします。オブジェクトの名前をルックアップするには、**[Object Explorer]** ビューなど、Developer tool のビューを使用します。カラム、プロファイル、トランスフォーメーションポートの名前をルックアップするにはエディタを使用します。

例えば、開発者が Developer ツール内の Sales\_Audit データオブジェクトに対応するビジネス用語をビジネス用語集で検索するとします。開発者は、そのビジネス用語の詳細を確認して、ビジネス要件と、Developer ツールにおける Sales\_Audit オブジェクトの現在の実装を理解したいと考えています。これは、そのデータオブジェクトの意味や、オブジェクトへの実装が必要な可能性がある変更を把握ことができます。

## ビジネス用語のルックアップ

ビジネス要件と現在の実装を理解するため、Developer ツールのオブジェクト名をビジネス用語としてビジネス用語集デスクトップでルックアップします。

ビジネス用語集デスクトップは、使用するマシンにあらかじめインストールしておく必要があります。

1. オブジェクトを選択します。
  2. ホットキーを使用するか、[検索] メニューを使用するかを選択して、ビジネス用語集デスクトップを開きます。
    - ホットキーを使用する場合は、以下のホットキーの組み合わせを使用します。  
CTRL+Shift+F
    - [検索] メニューを使用する場合は、[検索] > [Business Glossary] をクリックします。
- ビジネス用語集デスクトップが表示され、オブジェクト名に一致するビジネス用語が表示されます。

## ビジネス用語のルックアップ用にホットキーをカスタマイズ

ホットキーをカスタマイズして、ビジネス用語集デスクトップを開くキーの組み合わせを変更します。

1. Developer ツールのメニューから、[ウィンドウ] > [設定] > [全般] > [キー] をクリックします。
2. [Business Glossary の検索] をコマンドのリストで見つけたり、検索したりするには、以下のいずれかを選択します。
  - 検索してキーを探すには、検索ボックスに「ビジネス用語集検索」と入力します。
  - スクロールしてキーを探すには、[コマンド] カラムでスクロールして [Business Glossary の検索] コマンドを見つけます。
3. [Business Glossary の検索コマンド] をクリックします。
4. [コマンドのバインド解除] をクリックします。
5. [バインディング] フィールドで、キーの組み合わせを入力します。
6. [適用] をクリックしてから、[OK] をクリックします。

## 付録 A

# プロファイリングウェアハウス接続に基づく機能のサポート

- [プロファイリング機能のサポート, 221 ページ](#)

## プロファイリング機能のサポート

JDBC 接続またはネイティブ接続を介してプロファイリングウェアハウスに接続できます。プロファイリングウェアハウスの接続に基づいて、プロファイリングで特定の機能を実行できます。

以下の表に、Data Engineering Quality と Data Engineering Integration で選択したプロファイリングウェアハウス接続のタイプに基づいて実行できる機能を示します。

機能	JDBC 接続	ネイティブ接続
単一のデータオブジェクトのカラムプロファイル	サポート	サポート
単一のデータオブジェクトのデータドメイン検出	サポート	サポート
プライマリキー検出を使用した単一のデータオブジェクトプロファイル	サポートされていません	サポートされていません
機能依存性検出を使用した単一のデータオブジェクトプロファイル	サポートされていません	サポートされていません
[サンプリング最初の<number>行] サンプリングを使用したプロファイル	サポート	サポート
[ランダムサンプリング<number>行] サンプリングを使用したプロファイル	サポート	サポート
スコアカードのメトリックと値のコスト	サポートされていません	サポート
プライマリキーおよび外部キー検出を使用したエンタープライズ検出プロファイル	サポートされていません	サポートされていません

機能	JDBC 接続	ネイティブ接続
結合分析と重複検出を使用したエンタープライズ検出プロファイル	サポートされていません	サポートされていません
推測された値と値頻度のドリルダウン	サポート	サポート
プロファイル結果のエクスポート	サポート	サポート
行数マッピング*	サポートされていません	サポート
*統計ヘルパーで行数の報告に失敗すると、プロファイリングは行数マッピングを実行します。		

# 索引

## A

Analyst ツールでのエンタープライズ検出

- 概要 [118](#)
- サマリビュー [125](#)
- データ型の競合 [127](#)
- プロセス [119](#)
- プロファイルビュー [128](#)

Analyst ツールでのカラムプロファイル結果

- インタフェース [60](#), [73](#), [74](#)
- カラムの詳細 [61](#), [75](#)
- サマリ [59](#)

Analyst ツールでの検出検索

- プロセス [130](#)

Analyst ツールでの検出検索

- 一致タイプ [134](#)
- オプション [131](#)
- 概要 [129](#)
- 間接一致 [134](#)
- 関連するアセット [135](#)
- 直接一致 [134](#)

Analyst ツールでの検出検索結果

- 概要 [132](#)

## I

Informatica Analyst

- カラムプロファイル結果 [58](#), [70](#)
- カラムプロファイルの概要 [36](#), [72](#)
- ルール [48](#)
- ロックとバージョン管理 [41](#)

Informatica Analyst でのエンタープライズ検出の結果

- 概要 [124](#)

Informatica Developer

- プロファイルのビュー [142](#)
- プロファイルの概要 [141](#)
- ルール [166](#)

## M

Metadata Manager のビジネス用語

- ビジネス用語の管理 [139](#)
- プロジェクト [138](#)

## S

Sqoop 設定

- プロファイリング [40](#), [148](#)

## い

異常値

- 検出 [66](#)

## え

export

- XML へのスコアカードリネージュ [177](#)

エンタープライズ検出結果

- イメージとして保存 [212](#)
- エクスポート [218](#)

- 概要 [208](#)

エンタープライズ検出プロファイル

- DDL スクリプト [206](#)
- DDL スクリプトの作成 [206](#)

- 実行 [200](#)

エンタープライズ検出

- Analyst ツールでの編集 [122](#)
- Informatica Analyst での実行 [121](#)
- 外部キープロファイリングビュー [210](#)
- 概要 [191](#)
- [カラムプロファイル] ビュー [216](#)
- データオブジェクトリレーションの表示 [210](#)
- [データドメイン] ビュー [215](#)
- テーブルビュー [213](#)
- プロセス [192](#)
- 編集 [199](#)
- ランタイムステータス [217](#)
- [リレーション] ビュー [209](#)

## か

外部キー検出

- 概要 [201](#)

外部キープロファイル

- 検出 [201](#)

カラムプロファイル

- Informatica Developer [145](#)
- オプション [26](#)
- オペレーティングシステムプロファイル [41](#), [152](#)
- ドリルダウン [76](#)
- プロセス [37](#)
- 概要 [25](#)

カラムプロファイルの作成

- プロファイル [42](#)

カラムプロファイル結果

- Informatica Developer [171](#)

## き

機能依存性検出

- 概要 [151](#)

## キューレーション

- Informatica Analyst [77](#)
- Informatica Developer [174](#)
- 概念 [32](#)
- タスク [33](#)
- プロセス [33](#)

## け

### 傾向グラフ

- Informatica Analyst からのエクスポート [95](#)
- コスト [93](#)
- スコア [93](#)
- 表示 [94](#)

### 結合分析

- 概要 [202](#)

### 検索

- ビジネス用語集 [219](#)

### 検出検索

- 前提条件 [130](#)

### 検出検索結果

- インタフェース [132](#)

## し

### 式ルールの作成

- ルール [50](#)

### 重複検出

- 概要 [204](#)
- 結果 [204](#)
- 実行 [205](#)

## す

### スコアカード

- Informatica Analyst [83](#)
- Informatica Analyst でのプロセス [84](#)
- Informatica Developer [176](#)
- 概要 [27](#)
- グローバル通知設定の実行 [105](#)
- 傾向グラフ [93](#)
- 固定費 [90](#)
- しきい値の定義 [90](#)
- 実行中 [87](#)
- スコアカードへのカラムの追加 [86](#)
- スコアの移動 [91](#)
- 通知 [103](#)
- 通知の設定 [105](#)
- ドリルダウン [92](#)
- 表示 [88](#)
- 変動費 [90](#)
- 無効なデータのコスト [89](#)
- メトリック [89](#)
- メトリックウェイト [89](#)
- メトリックグループの削除 [92](#)
- メトリックグループの作成 [91](#)
- メトリックグループの編集 [91](#)
- 編集 [88](#)
- スコアカード実行傾向ペイン
- Informatica Analyst [98](#)
- スコアカードダッシュボード
- Informatica Analyst [96](#)
- スコアカード付きデータオブジェクト
- Informatica Analyst [99](#)

### スコアカードの結果

- Excel へのエクスポート [103](#)
- Informatica Analyst からのエクスポート [102](#)
- エクスポート [102](#)
- スコアカードリネージュ
- Informatica Analyst での表示 [106](#)
- Informatica Developer からの表示 [177](#)

## せ

### 設定オプション

- Analyst ツールでのエンタープライズ検出 [119](#)

## て

### 定義済みルール

- プロセス [49](#)

### データオブジェクトプロファイル

- エンタープライズ検出 [198](#)
- 概要 [144](#)
- コメント [158](#)
- 単一のプロファイルの作成 [153](#)
- 複数のプロファイルの作成 [154](#)

### データ検出

- 概要 [21](#)
- プロセス [19](#)

### データドメイン

- Informatica Analyst での検索 [109](#)
- Informatica Analyst でのプロファイル結果からの作成 [109](#)
- Informatica Analyst での作成 [108](#)
- Informatica Developer での検索 [182](#)
- Informatica Developer でのプロファイル結果からの作成 [181](#)
- Informatica Developer での作成 [180](#)
- インポート [182](#)
- エクスポート [183](#)
- 概要 [29](#)

### データドメイングループ

- Informatica Analyst での作成 [108](#)
- Informatica Developer での作成 [180](#)
- 概要 [29](#)

### データドメイングロッサリ

- Informatica Analyst [107](#)
- Informatica Developer [180](#)
- 概要 [30](#)

### データドメイン検出

- Informatica Analyst の概要 [107](#)
- Informatica Developer の概要 [179](#)
- 概要 [28](#)
- プロセス [30](#)

### データドメイン検出オプション

- Informatica Developer [183](#)

### データドメイン検出結果

- Informatica Analyst [115](#)
- Informatica Analyst からのエクスポート [117](#)
- Informatica Analyst でのエクスポート [116](#)
- Informatica Developer [188](#)
- Informatica Developer からのエクスポート [190](#)

### データドメイン検出プロファイル結果

- Microsoft Excel [116](#)

### テーブルデータオブジェクト

- 同期 [47](#)



## ひ

### ビジネス用語

- ビジネス用語のルックアップ [139](#)
- ホットキーのカスタマイズ [220](#)
- ルックアップ [219](#)

## ふ

### フィルタ

- 概要 [53](#)

### プライマリキー検出

- 概要 [149](#)

### フラットファイルデータオブジェクト

- 同期 [45](#)

### プロジェクト

- Metadata Manager のビジネス用語 [138](#)

### プロジェクト別のスコアカードペイン

- Informatica Analyst [97](#)

### プロファイリング

- アーキテクチャ [18](#)

- 概要 [16](#)

- ロックとバージョン管理 [26](#)

### プロファイル

- Avro または Parquet 形式 [161](#)
- XML 形式と JSON 形式 [160](#), [161](#)
- カラムプロファイルの編集 [43](#)
- カラムプロファイルの作成 [42](#)
- コンポーネント [22](#)
- 実行中 [44](#), [71](#), [72](#), [115](#)
- フィルタの作成 [53](#)
- フィルタの編集 [56](#)

### プロファイルオプション

- エンタープライズ検出 [193](#)

### プロファイル結果

- Developer tool でのデータドメインの拒否 [174](#)
- Excel [78](#)
- Informatica Analyst からのエクスポート [79](#)
- Informatica Developer でのエクスポート [175](#)
- Informatica Developer でのカラムのリレーションのキュレーション [213](#)
- Informatica Developer でのコメントの追加 [158](#)
- Informatica Developer でのデータドメインの拒否 [190](#)
- Informatica Developer でのデータドメインの承認 [189](#)
- Informatica Developer でのデータ型の承認 [174](#)

### プロファイル結果 (続く)

- エクスポート [78](#)
- カラム値 [67](#)
- カラムデータ型 [64](#), [173](#)
- カラムパターン [66](#)
- コメント [81](#)
- サマリ [73](#), [76](#)
- サマリビュー [60](#)
- 詳細ビュー [62](#)
- タグ [82](#)
- データ型の拒否 [77](#)
- データ型の承認 [77](#)
- データドメインの却下 [116](#)
- データドメインの承認 [115](#)
- ドリルダウン [76](#)
- ビジネス用語 [81](#)

## ま

### マッピングオブジェクト

- プロファイルの実行 [168](#)

### マップレットとマッピングのプロファイリング

- 概要 [168](#)

## ら

### ランタイム環境

- Analyst ツール [39](#)
- Hadoop [39](#), [148](#)

## る

### 累積メトリックペイン

- Informatica Analyst [100](#)

### ルール

- Informatica Developer での作成 [167](#)
- Informatica Developer での適用 [167](#)
- PowerCenter Express での適用 [167](#)
- 式 [50](#)
- 式ルールの作成 [50](#)
- 定義済み [48](#)
- 定義済みルールの適用 [49](#)
- ルール仕様を使用した式ルールの作成 [51](#)