



Informatica®
10.1

プロファイリング入門ガイド

Informatica プロファイリング入門ガイド

10.1

2016 年 6 月

© 著作権 Informatica LLC 1998, 2018

本ソフトウェアおよびマニュアルには、Informatica LLC の所有権下にある情報が収められています。これらは使用および開示の制限等を定めた使用許諾契約のもとに提供され、著作権法により保護されています。当該ソフトウェアのリバースエンジニアリングは禁じられています。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。このソフトウェアは、米国および/または国際的な特許、およびその他の出願中の特許によって保護されています。

合衆国政府によるソフトウェアの使用、複製または開示は、DFARS 227.7202-1 (a) および 227.7702-3 (a) (1995 年)、DFARS 252.227-7013(C) (1) (ii) (1988 年 10 月)、FAR 12.212 (a) (1995 年)、FAR 52.227-19、または FAR 52.227-14 (ALT III) に記載されているとおりに、当該ソフトウェア使用許諾契約に定められた制限によって規制されます。

本製品または本書の情報は、予告なしに変更されることがあります。お客様が本製品または本書内に問題を発見された場合は、書面に当社までお知らせください。

Informatica、Informatica Platform、Informatica Data Services、PowerCenter、PowerCenterRT、PowerCenter Connect、PowerCenter Data Analyzer、PowerExchange、PowerMart、Metadata Manager、Informatica Data Quality、Informatica Data Explorer、Informatica B2B Data Transformation、Informatica B2B Data Exchange、Informatica On Demand、Informatica Identity Resolution、Informatica Application Information Lifecycle Management、Informatica Complex Event Processing、Ultra Messaging、Informatica Master Data Management、および Live Data Map は、Informatica LLC の米国および世界中の管轄地での商標または登録商標です。その他のすべて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。

本ソフトウェアまたはドキュメントの一部は、次のサードパーティが有する著作権に従います（ただし、これらに限定されません）。Copyright DataDirect Technologies. All rights reserved. Copyright (C) Sun Microsystems. All rights reserved. Copyright (C) RSA Security Inc. All rights reserved. Copyright (C) Ordinal Technology Corp. All rights reserved. Copyright (C) Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright (C) Meta Integration Technology, Inc. All rights reserved. Copyright (C) Intalio. All rights reserved. Copyright (C) Oracle. All rights reserved. Copyright (C) Adobe Systems Incorporated. All rights reserved. Copyright (C) DataArt, Inc. All rights reserved. Copyright (C) ComponentSource. All rights reserved. Copyright (C) Microsoft Corporation. All rights reserved. Copyright (C) Rogue Wave Software, Inc. All rights reserved. Copyright (C) Teradata Corporation. All rights reserved. Copyright (C) Yahoo! Inc. All rights reserved. Copyright (C) Glyph & Cog, LLC. All rights reserved. Copyright (C) Thinkmap, Inc. All rights reserved. Copyright (C) Clearpace Software Limited. All rights reserved. Copyright (C) Information Builders, Inc. All rights reserved. Copyright (C) OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright (C) International Organization for Standardization 1986. All rights reserved. Copyright (C) ej-technologies GmbH. All rights reserved. Copyright (C) Jaspersoft Corporation. All rights reserved. Copyright (C) International Business Machines Corporation. All rights reserved. Copyright (C) yWorks GmbH. All rights reserved. Copyright (C) Lucent Technologies. All rights reserved. Copyright (C) University of Toronto. All rights reserved. Copyright (C) Daniel Veillard. All rights reserved. Copyright (C) Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright (C) MicroQuill Software Publishing, Inc. All rights reserved. Copyright (C) PassMark Software Pty Ltd. All rights reserved. Copyright (C) LogiXML, Inc. All rights reserved. Copyright (C) 2003-2010 Lorenzi Davide, All rights reserved. Copyright (C) Red Hat, Inc. All rights reserved. Copyright (C) The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright (C) EMC Corporation. All rights reserved. Copyright (C) Flexera Software. All rights reserved. Copyright (C) Jinfonet Software. All rights reserved. Copyright (C) Apple Inc. All rights reserved. Copyright (C) Telerik Inc. All rights reserved. Copyright (C) BEA Systems. All rights reserved. Copyright (C) PDFlib GmbH. All rights reserved. Copyright (C) Orientation in Objects GmbH. All rights reserved. Copyright (C) Tanuki Software, Ltd. All rights reserved. Copyright (C) Ricebridge. All rights reserved. Copyright (C) Sencha, Inc. All rights reserved. Copyright (C) Scalable Systems, Inc. All rights reserved. Copyright (C) jQWidgets. All rights reserved. Copyright (C) Tableau Software, Inc. All rights reserved. Copyright (C) MaxMind, Inc. All rights reserved. Copyright (C) T-Mate Software s.r.o. All rights reserved. Copyright (C) MapR Technologies Inc. All rights reserved. Copyright (C) Amazon Corporate LLC. All rights reserved. Copyright (C) Highsoft. All rights reserved. Copyright (C) Python Software Foundation. All rights reserved. Copyright (C) BeOpen.com. All rights reserved. Copyright (C) CNRI. All rights reserved.

本製品には、Apache Software Foundation (<http://www.apache.org/>) によって開発されたソフトウェア、およびさまざまなバージョンの Apache License（まとめて「License」と呼んでいます）の下に許諾された他のソフトウェアが含まれます。これらのライセンスのコピーは、<http://www.apache.org/licenses/> で入手できます。適用法にて要求されないか書面に合意されない限り、ライセンスの下に配布されるソフトウェアは「現状のまま」で配布され、明示的あるいは黙示的かを問わず、いかなる種類の保証や条件も付帯することはありません。ライセンス下での許諾および制限を定める具体的文言については、ライセンスを参照してください。

本製品には、Mozilla (<http://www.mozilla.org/>) によって開発されたソフトウェア、ソフトウェア copyright The JBoss Group, LLC、コンテンツの無断複製・転載を禁じます、ソフトウェア copyright, Red Hat Middleware, LLC、コンテンツの無断複製・転載を禁じます、Copyright (C) 1999-2006 by Bruno Lowagie and Paulo Soares および GNU Lesser General Public License Agreement (<http://www.gnu.org/licenses/lgpl.html> を参照) に基づいて許諾されたその他のソフトウェアが含まれています。資料は、Informatica が無料で提供しており、一切の保証を伴わない「現状渡し」で提供されるものとし、Informatica LLC は市場性および特定の目的の適合性の黙示の保証などを含めて、一切の明示的及び黙示的保証の責任を負いません。

製品には、ワシントン大学、カリフォルニア大学アーバイン校、およびバンダービルト大学の Douglas C. Schmidt および同氏のリサーチグループが著作権を持つ ACE (TM) および TAO (TM) ソフトウェアが含まれています。Copyright (C) 1993-2006, All rights reserved.

本製品には、OpenSSL Toolkit を使用するために OpenSSL Project が開発したソフトウェア（copyright The OpenSSL Project. コンテンツの無断複製・転載を禁じます）が含まれています。また、このソフトウェアの再配布は、<http://www.openssl.org> および <http://www.openssl.org/source/license.html> にある使用条件に従います。

本製品には、Curl ソフトウェア Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>が含まれます。All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://curl.haxx.se/docs/copyright.html> にある使用条件に従います。すべてのコピーに上記の著作権情報とこの許諾情報が記載されている場合、目的に応じて、本ソフトウェアの使用、コピー、変更、ならびに配布が有償または無償で許可されます。

本製品には、MetaStuff, Ltd. のソフトウェアが含まれます。Copyright 2001-2005 (C) MetaStuff, Ltd. All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://www.dom4j.org/license.html> にある使用条件に従います。

製品には、The Dojo Foundation のソフトウェアが含まれます。Copyright (C) 2004-2007. All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://dojotoolkit.org/license> にある使用条件に従います。

本製品には、ICU ソフトウェアおよび他のソフトウェアが含まれます。Copyright International Business Machines Corporation. All rights reserved. 本ソフトウェアに関する許諾および制限は、<http://source.icu-project.org/repos/icu/icu/trunk/license.html> にある使用条件に従います。

本製品には、Per Bothner のソフトウェアが含まれます。Copyright (C) 1996-2006. All rights reserved. お客様がこのようなソフトウェアを使用するための権利は、ライセンスで規定されています。<http://www.gnu.org/software/kawa/Software-License.html> を参照してください。

本製品には、OSSP UUID ソフトウェアが含まれます。Copyright (C) 2002 Ralf S. Engelschall, Copyright (C) 2002 The OSSP Project Copyright (C) 2002 Cable & Wireless Deutschland. 本ソフトウェアに関する許諾および制限は、<http://www.opensource.org/licenses/mit-license.php> にある使用条件に従います。

本製品には、Boost (<http://www.boost.org/>) によって開発されたソフトウェア、または Boost ソフトウェアライセンスの下で開発されたソフトウェアが含まれます。本ソフトウェアに関する許諾および制限は、http://www.boost.org/LICENSE_1_0.txt にある使用条件に従います。

本製品には、University of Cambridge のが含まれます。Copyright (C) 1997-2007. 本ソフトウェアに関する許諾および制限は、<http://www.pcre.org/license.txt> にある使用条件に従います。

本製品には、The Eclipse Foundation のソフトウェアが含まれます。Copyright (C) 2007. All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://www.eclipse.org/org/documents/epl-v10.php> および <http://www.eclipse.org/org/documents/edl-v10.php> にある使用条件に従います。

本製品には、<http://www.tcl.tk/software/tcltk/license.html>、<http://www.bosrup.com/web/overlib/?License>、<http://www.stlport.org/doc/license.html>、<http://www.asm.ow2.org/license.html>、<http://www.cryptix.org/LICENSE.TXT>、<http://hsqldb.org/web/hsqldbLicense.html>、<http://httpunit.sourceforge.net/doc/license.html>、<http://jung.sourceforge.net/license.txt>、http://www.gzip.org/zlib/zlib_license.html、<http://www.openldap.org/software/release/license.html>、<http://www.libssh2.org>、<http://slf4j.org/license.html>、<http://www.sente.ch/software/OpenSourceLicense.html>、<http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>、<http://antlr.org/license.html>、<http://aopalliance.sourceforge.net/>、<http://www.bouncycastle.org/license.html>、<http://www.jgraph.com/jgraphdownload.html>、<http://www.jcraft.com/jsch/LICENSE.txt>、http://jotm.objectweb.org/bsd_license.html に基づいて許諾されたソフトウェアが含まれています。<http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>、<http://www.slf4j.org/license.html>、<http://nanoxml.sourceforge.net/orig/copyright.html>、<http://www.json.org/license.html>、<http://forge.ow2.org/projects/javaservice/>、<http://www.postgresql.org/about/licence.html>、<http://www.sqlite.org/copyright.html>、<http://www.tcl.tk/software/tcltk/license.html>、<http://www.jaxen.org/faq.html>、<http://www.jdom.org/docs/faq.html>、<http://www.slf4j.org/license.html>、<http://www.iodbc.org/dataspace/iodbc/wiki/iODBC/License>、<http://www.keplerproject.org/md5/license.html>、<http://www.toedter.com/en/jcalendar/license.html>、<http://www.edankert.com/bounce/index.html>、<http://www.net-snmp.org/about/license.html>、<http://www.openmdx.org/#FAQ>、http://www.php.net/license/3_01.txt、<http://srp.stanford.edu/license.txt>、<http://www.schneider.com/blowfish.html>、<http://www.jmock.org/license.html>、<http://xsom.java.net>、<http://benalman.com/about/license/>、<https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>、<http://www.h2database.com/html/license.html#summary>、<http://jsoncpp.sourceforge.net/LICENSE>、<http://jdbc.postgresql.org/license.html>、<http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>、<https://github.com/rantav/hector/blob/master/LICENSE>、<http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>、<http://jibx.sourceforge.net/jibx-license.html>、<https://github.com/lyokato/libgeohash/blob/master/LICENSE>、<https://github.com/hjiang/jsonxx/blob/master/LICENSE>、<https://code.google.com/p/lz4/>、<https://github.com/jedisct1/libsodium/blob/master/LICENSE>、<http://one-jar.sourceforge.net/index.php?page=documents&file=license>、<https://github.com/EsotericSoftware/kryo/blob/master/license.txt>、<http://www.scala-lang.org/license.html>、<https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>、<http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>、<https://aws.amazon.com/asl/>、<https://github.com/twbs/bootstrap/blob/master/LICENSE>、および <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>。

本製品には、Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>)、Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>)、Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>)、Sun Binary Code License Agreement Supplemental License Terms、BSD License (<http://www.opensource.org/licenses/bsd-license.php>)、BSD License (<http://opensource.org/licenses/BSD-3-Clause>)、MIT License (<http://www.opensource.org/licenses/mit-license.php>)、Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>)、Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>) に基づいて許諾されたソフトウェアが含まれています。

本製品には、ソフトウェア copyright (C) 2003-2006 Joe Walnes, 2006-2007 XStream Committers が含まれています。All rights reserved.本ソフトウェアに関する許諾および制限は、<http://j.org/license.html> にある使用条件に従います。本製品には、Indiana University Extreme! Lab によって開発されたソフトウェアが含まれています。詳細については、<http://www.extreme.indiana.edu/> を参照してください。

本製品には、ソフトウェア Copyright (C) 2013 Frank Balluffi and Markus Moeller が含まれています。All rights reserved.本ソフトウェアに関する許諾および制限は、MIT ライセンスの使用条件に従います。

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

免責: 本文書は、一切の保証を伴わない「現状渡し」で提供されるものとし、Informatica LLC は他社の権利の非侵害、市場性および特定の目的への適合性の黙示の保証などを含めて、一切の明示的および黙示の保証の責任を負いません。Informatica LLC では、本ソフトウェアまたはドキュメントに誤りのないことを保証していません。本ソフトウェアまたはドキュメントに記載されている情報には、技術的に不正確な記述や誤植が含まれる場合があります。本ソフトウェアまたはドキュメントの情報は、予告なしに変更されることがあります。

NOTICES

この Informatica 製品（以下「ソフトウェア」）には、Progress Software Corporation（以下「DataDirect」）の事業子会社である DataDirect Technologies からの特定のドライバ（以下「DataDirect ドライバ」）が含まれています。DataDirect ドライバには、次の用語および条件が適用されます。

1. DataDirect ドライバは、特定物として現存するままの状態提供され、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。
2. DataDirect または第三者は、予見の有無を問わず発生した ODBC ドライバの使用に関するいかなる直接的、間接的、偶発的、特別、あるいは結果的損害に対して責任を負わないものとします。本制限事項は、すべての訴訟原因に適用されます。訴訟原因には、契約違反、保証違反、過失、厳格責任、詐称、その他の不法行為を含みますが、これらに限るものではありません。

発行日: 2018-06-26

目次

序文	7
Informatica のリソース	7
Informatica Network	7
Informatica ナレッジベース	7
Informatica マニュアル	8
Informatica 製品可用性マトリックス	8
Informatica Velocity	8
Informatica Marketplace	8
Informatica グローバルカスタマサポート	8
第 1 章 : スタートガイド概要	9
プロファイリングの概要	9
Informatica Analyst について	10
Informatica Analyst ユーザーインターフェース	11
Informatica Developer のユーザーインターフェース	13
Informatica Developer のようこそページ	14
チートシート	14
Informatica Developer のタスク	14
チュートリアルの内容	15
チュートリアルの構造	15
チュートリアルの前提条件	16
Informatica Analyst のチュートリアル	16
Informatica Developer tool	17
第 I 部 : Informatica Analyst の基本操作	18
第 2 章 : レッスン 1. Informatica Analyst の設定	19
Informatica Analyst の設定の概要	19
タスク 1. Informatica Analyst へのログイン	20
タスク 2. プロジェクトの作成	20
タスク 3. フォルダーの作成	21
Informatica Analyst の設定のまとめ	21
第 3 章 : レッスン 2. データオブジェクトの作成	22
データオブジェクトの作成の概要	22
タスク 1. フラットファイルデータオブジェクトの作成	23
タスク 2. データオブジェクトのプロパティの表示	23
データオブジェクトの作成のまとめ	24

第4章 : レッスン 3. デフォルトプロファイルの作成	25
デフォルトプロファイルの作成の概要	25
タスク 1. デフォルトプロファイルの作成と実行	26
タスク 2. サマリビューでのプロファイル結果の表示	26
タスク 3. 詳細ビューでのプロファイル結果の表示	27
デフォルトプロファイルの作成のサマリ	28
第5章 : レッスン 4. カスタムプロファイルの作成	29
カスタムプロファイルの作成の概要	29
タスク 1. カスタムプロファイルの作成	30
タスク 2. プロファイルの実行	31
タスク 3. プロファイル結果でのドリルダウン	31
カスタムプロファイルの作成のまとめ	32
第6章 : レッスン 5. 式ルール of 作成	33
式ルールの作成の概要	33
タスク 1. 式ルールの作成とプロファイルの実行	34
タスク 2. 式ルールの出力の表示	34
タスク 3. 式ルールの編集	35
式ルールの作成のまとめ	35
第7章 : レッスン 6. スコアカードの作成と実行	36
スコアカードの作成と実行の概要	36
タスク 1. プロファイル結果からのスコアカードの作成	37
タスク 2. スコアカードの実行	38
タスク 3. スコアカードの表示	38
タスク 4. スコアカードの編集	39
タスク 5. しきい値の設定	40
タスク 6. スコアの傾向グラフの表示	40
スコアカードの作成と実行のまとめ	40
第8章 : レッスン 7. プロファイルカラムからの参照テーブルの作成	42
プロファイルカラムからの参照テーブルの作成の概要	42
タスク 1. プロファイルカラムからの参照テーブルの作成	43
タスク 2. 参照テーブルの編集	44
プロファイルカラムからの参照テーブルの作成のまとめ	44
第9章 : レッスン 8. 参照テーブルの作成	45
参照テーブルの作成の概要	45
タスク 1. 参照テーブルの作成	46
参照テーブルの作成のまとめ	46

第 II 部 : Informatica Developer の基本操作	47
第 10 章 : レッスン 1. Informatica Developer の設定	48
Informatica Developer の設定の概要	48
タスク 1. Informatica Developer の起動	49
タスク 2. ドメインの追加	49
タスク 3. モデルリポジトリの追加	50
タスク 4. プロジェクトの作成	50
タスク 5. フォルダの作成	50
タスク 6. デフォルトの Data Integration Service の選択	51
Informatica Developer の設定のまとめ	51
第 11 章 : レッスン 2: 物理データオブジェクトのインポート	52
物理データオブジェクトのインポートの概要	52
タスク 1. Boston_Customers フラットファイルデータオブジェクトのインポート	53
タスク 2. LA_Customers フラットファイルデータオブジェクトのインポート	58
タスク 3. All_Customers フラットファイルデータオブジェクトのインポート	59
物理データオブジェクトのインポートのまとめ	59
第 12 章 : レッスン 3. データソースでのプロファイルの実行	60
データのプロファイリングの概要	60
タスク 1. 2 つのデータソースでの結合分析の実行	61
タスク 2. 結合分析結果の表示	62
タスク 3. データソースでのプロファイルの実行	62
タスク 4. カラムプロファイル結果の表示	63
データのプロファイリングのまとめ	63
索引	64

序文

『Informatica プロファイリング入門ガイド』は、データ品質とデータサービスの開発者およびアナリストを対象としています。Informatica Developer および Informatica Analyst を初めて使用するユーザーのためのチュートリアルを提供します。本書では、データ品質の概念、フラットファイルとリレーショナルデータベースの概念、および作業環境のデータベースエンジンについて理解していることを前提としています。

Informatica のリソース

Informatica Network

Informatica Network は、Informatica グローバルカスタマサポート、Informatica ナレッジベースなどの製品リソースをホストします。Informatica Network には、<https://network.informatica.com> からアクセスしてください。

メンバーは以下の操作を行うことができます。

- 1つの場所からすべての Informatica のリソースにアクセスできます。
- ドキュメント、FAQ、ベストプラクティスなどの製品リソースをナレッジベースで検索できます。
- 製品の提供情報を表示できます。
- 自分のサポート事例を確認できます。
- 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。

メンバーは以下の操作を行うことができます。

- 1つの場所からすべての Informatica のリソースにアクセスできます。
- ドキュメント、FAQ、ベストプラクティスなどの製品リソースをナレッジベースで検索できます。
- 製品の提供情報を表示できます。
- 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

ドキュメント、ハウツー記事、ベストプラクティス、PAM などの製品リソースを Informatica Network で検索するには、Informatica ナレッジベースを使用します。

ナレッジベースには、<https://kb.informatica.com> から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) です。

Informatica マニュアル

使用している製品の最新のドキュメントを取得するには、
https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx にあ
る Informatica ナレッジベースを参照してください。

このマニュアルに関する質問、コメント、ご意見の電子メールの送付先は、Informatica マニュアルチーム
(infa_documentation@informatica.com) です。

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 Network メンバである場合は、
PAM
(<https://network.informatica.com/community/informatica-network/product-availability-matrices>) に
アクセスできます。

Informatica Velocity

Informatica Velocity は、Informatica プロフェッショナルサービスによって開発されたヒントおよびベスト
プラクティスのコレクションです。数多くのデータ管理プロジェクトの経験から開発された Informatica
Velocity には、世界中の組織と協力して優れたデータ管理ソリューションの計画、開発、展開、および維持を
行ってきた弊社コンサルタントの知識が集約されています。

Informatica Network メンバである場合は、Informatica Velocity リソース
(<http://velocity.informatica.com>) にアクセスできます。

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 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 製品を強化したり拡張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。Informatica の開発者およびパートナーの何百ものソリューションを利用して、プロ
ジェクトで実装にかかる時間を短縮したり、生産性を向上させたりできます。Informatica Marketplace には、
<https://marketplace.informatica.com> からアクセスできます。

Informatica グローバルカスタマサポート

Informatica Network の電話またはオンラインサポートから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
(<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>) を参
照してください。

Informatica Network メンバである場合は、オンラインサポート (<http://network.informatica.com>) を使用
できます。

第 1 章

スタートガイド概要

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, 9 ページ](#)
- [Informatica Analyst について, 10 ページ](#)
- [Informatica Developer のユーザーインターフェース, 13 ページ](#)
- [チュートリアルの内容, 15 ページ](#)
- [チュートリアルの構造, 15 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。

Informatica Analyst について

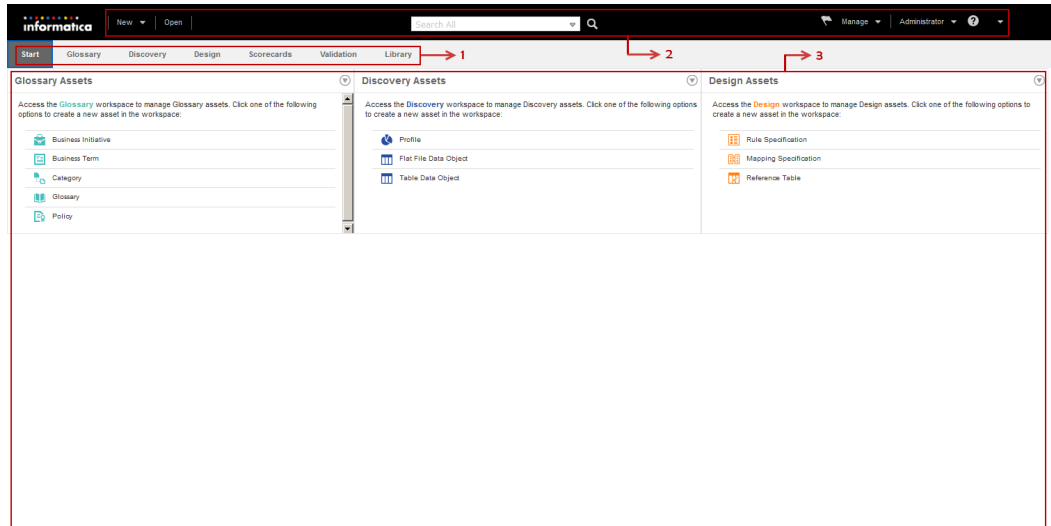
Informatica Analyst は、アナリストがエンタープライズ内のデータを分析、クレンジング、標準化、プロファイル、およびスコアリングするために使用できる Web ベースのアプリケーションクライアントです。

ライセンスによっては、ビジネスアナリストと開発者は、Analyst ツールをデータドリブンコラボレーションに使用します。カラムプロファイリングとルールプロファイリング、スコアカードの作成、および不良レコードと重複レコードの管理を実行できます。データ品質ソリューションでは開発者に対して、参照データの管理と提供も行うことができます。

Informatica Analyst ユーザーインターフェース

Analyst ツールのユーザーインターフェースは、データ統合ソリューションとデータ品質データ連携ソリューションの作成に使用する複数のワークスペースで構成されます。ワークスペースは、ライセンスと特権に基づいてアクセスできる Web ページです。

次は Analyst ツールユーザーインターフェースの画像です。



1. ワークスペースアクセスパネル
2. ヘッダ領域
3. ワークスペースのタブ

【開始】 ワークスペースに Analyst ツールが開きます。**【開始】** ワークスペースには、ワークスペースアクセスパネルで使用するためのライセンスと特権のあるワークスペースが表示されます。

Analyst ツールには次のヘッダーメニューが表示されます。

新規

【用語集】、**【検出】**、**【設計】** の各ワークスペースにアセットを作成します。

開く

【ライブラリ】 ワークスペースを開きます。

通知アラート

用語集アセットに関する通知を表示します。

管理

一時的なワークスペースと **【通知】** を開きます。**【接続】**、**【データドメイン】**、**【ジョブステータス】**、**【プロジェクト】**、および **【Business Glossary のセキュリティ】** ワークスペースを開くことができます。

ユーザー名

パスワードを変更したり、Analyst ツールからログアウトしたりするために、ユーザー設定を設定します。

ヘルプ

現在のワークスペースのヘルプにアクセスします。

Analyst ツールには、次の永続的なワークスペースが含まれます。

開始

ライセンスでアクセス権が付与された他のワークスペースにワークスペースアクセスパネルからアクセスします。例外管理を行うライセンスがある場合は、ワークスペースの【**マイタスク**】パネルにタスクが表示されます。

用語集

組織にとって重要なビジネスコンセプトを定義して説明します。ビジネス用語、カテゴリ、用語集、ポリシーの作成と管理が可能です。

検出

ソースシステムのデータとメタデータの品質を分析します。プロファイル、フラットファイルデータオブジェクト、およびテーブルデータオブジェクトの作成と管理が可能です。

設計

アナリストと開発者の連携を支援するビジネスロジックを設計します。マッピング仕様、参照テーブル、ルール定義などの作成と管理が可能です。

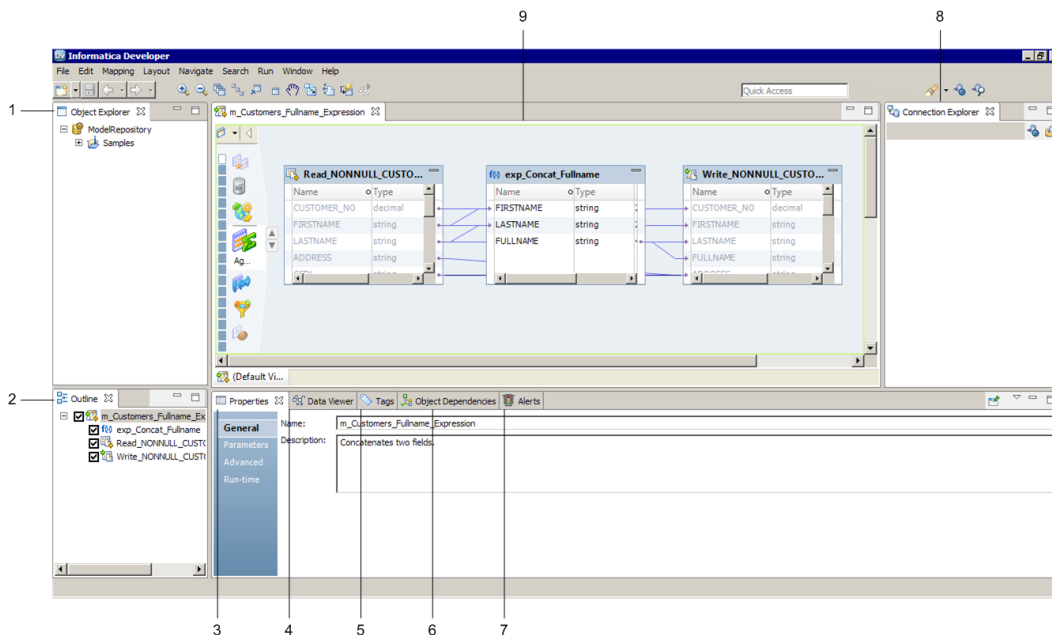
スコアカード

プロファイル結果から作成したスコアカードを開き、編集して実行します。メトリックの追加、カラムでのドリルダウン、スコアカードフィルタの追加、スコアカードの傾向グラフの表示が可能です。

Informatica Developer のユーザーインターフェース

Developer ツールのユーザーインターフェースはデータ統合およびデータ品質 Data Federation ソリューションの作成に使用する複数のビューを持つワークベンチから構成されています。

次の画像で Developer ツールのユーザーインターフェースを示します。



1. [Object Explorer] ビュー
2. [アウトライン] ビュー
3. [プロパティ] ビュー
4. [データビューア] ビュー
5. [タグ] ビュー
6. [オブジェクトの依存関係] ビュー
7. 警告ビュー
8. [接続エクスプローラ] ビュー
9. エディタ

Developer ツールは次のビューを表示できます。

[オブジェクトエクスプローラ] ビュー

プロジェクト、フォルダ、およびプロジェクトとフォルダ内のオブジェクトが表示されます。

[接続エクスプローラ] ビュー

リレーショナルデータベースへの接続が表示されます。

[アウトライン] ビュー

[オブジェクトエクスプローラ] ビューで選択したオブジェクトに応じたオブジェクトが表示されます。

[チートシート] ビュー

開いているチートシートが表示されます。チートシートを開くには、[ヘルプ] > [チートシート] をクリックして、チートシートを選択します。

[データビューア] ビュー

ソースデータ、プロファイル結果を表示し、トランスフォーメーションの出力をプレビューします。Web サービスメッセージをプレビューしたり、データビューアビューから SQL クエリを実行できます。

【オブジェクトの依存関係】 ビュー

オブジェクトの表示、編集、または削除を行うときにオブジェクトの依存関係が表示されます。

【警告】 ビュー

接続ステータス警告が表示されます。

【検証ログ】 ビュー

オブジェクトの検証エラーが表示されます。

Developer ツールは他のビューも表示できます。ビューを非表示にしたり、Developer ツールのワークベンチ内の別の場所に移動したりすることができます。表示するビューを選択するには、**【ウィンドウ】 > 【ビューの表示】** をクリックします。

Informatica Developer のようこそページ

Developer ツールを初めて開くと、ようこそページが表示されます。ようこそページを使用して、Developer ツールに関する知識を深め、Developer ツールを設定し、Developer ツールでの作業を開始します。

ようこそページには次のオプションが表示されます。

- 概要: データ品質およびデータサービスソリューションの概要を表示するには、**【概要】** ボタンをクリックします。
- 最初の手順: Developer ツールの設定および Informatica Data Quality Service と Informatica Data Services のレッスンへのアクセスに関する知識を深めるには、**【最初の手順】** ボタンをクリックします。
- チュートリアル: データ品質およびデータサービスソリューションのチュートリアルレッスンを表示するには、**【チュートリアル】** ボタンをクリックします。
- Web リソース: mysupport.informatica.com にリンクする **【Web リソース】** ボタンをクリックすると、Informatica How-To ライブラリにアクセスできます。Informatica How-To ライブラリには、Informatica Data Quality、Informatica Data Services、および他の Informatica 製品に関する記事が掲載されています。
- 新機能: Developer tool の最新機能を表示するには、**【新機能】** ボタンをクリックします。

チートシート

Developer ツールには、オンラインヘルプの一部としてチートシートが含まれています。チートシートは段階を追った手引きで、Developer ツールで 1 つまたは複数のタスクを完了するために役立ちます。

チートシートを完了して、タスクを完了して結果を確認します。例えば、チートシートを完了してリレーショナルデータオブジェクトをインポートおよびプレビューすると、リレーショナルデータベーステーブルはインポートされ、Developer ツールでのデータはプレビューされています。

チートシートにアクセスするには、**【ヘルプ】 > 【チートシート】** をクリックします。

Informatica Developer のタスク

Developer ツールのデータ品質機能を使用して、データの内容および構造を分析します。ビジネスニーズに合わせてデータを強化できます。

Developer ツールを使用して、次の目標を実現するプロセスを実行します。

- データソースでプロファイルを実行します。プロファイリングにより、データの内容と構造が明らかになります。プロファイリングは、どのようなデータプロジェクトにおいても重要です。プロファイリングを実施すると、データの長所と短所が明らかになり、プロジェクト計画の定義に役立ちます。

- スコアカードを作成して、データ品質を確認します。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。
- データ品質ルールの作成と実行。Informatica には、事前作成されたルールが用意されています。このルールは、そのまま実行することも、プロジェクト目標に合わせて編集することもできます。Developer ツールでルールを作成できます。
- Informatica ユーザーとの共同作業。モデルリポジトリに追加されたルールは、Developer ツールおよび Analyst ツールで使用できます。ユーザーはプロジェクトを共同で実施できます。また、プロジェクトのさまざまな段階でさまざまなユーザーがオブジェクトの所有権を取得できます。

チュートリアルの内容

HypoStores Corporation は、本社が Boston で、いくつかの州に店舗を持つ全国的な小売組織です。全国の店舗の業務データを本社のデータストアと定期的に統合しています。最近、Los Angeles に店舗を開店しました。

本社には、すべてのコンポジットおよび BI アプリケーションに共通のデータサービス層を提供する、管理者、開発者、およびアーキテクトから成る中央の ICC チームがあります。BI アプリケーションには、請求やマーケティングに使用されるマスター顧客データファイルが含まれる CRM システムが組み込まれています。

HypoStores Corporation は、データセットを統合する前に、Boston と Los Angeles のデータをプロファイルする必要があります。プロファイル操作により、統合前に HypoStores が修正できるデータ品質の問題が特定されます。

チュートリアルの構造

スタートガイドには、レッスンとタスクから構成されるチュートリアルが含まれています。

レッスン

各レッスンでは、タスクを理解してレッスンを完了するために役立つ概念を紹介します。レッスンでは、内容全体のビジネス要件が示されています。レッスンの目標では、ビジネス要件を満たすために完了するタスクの概略を説明します。レッスンごとに、完了までの推定時間が示されています。レッスン内のタスクを完了すると、レッスンのサマリを確認できます。

ツール内の環境が設定されていない場合は、各チュートリアルの最初のレッスンでそれを行う方法が示されています。

タスク

タスクでは、段階を追った説明が示されています。示された順序ですべてのタスクを完了すると、レッスンは完了です。

チュートリアル の前提条件

チュートリアル のレッスンを開始する前に、少なくとも 1 つのノード設定で Informatica ドメインが実行されている必要があります。アナリストサービス、モデルリポジトリサービス、およびデータ統合サービスが実行されていることを確認する必要があります。

インストーラには、レッスンを完了するために使用するチュートリアルファイルが含まれています。すべてのファイルは、クライアントのインストールディレクトリとサーバーのインストールディレクトリの両方に置かれています。

- チュートリアルファイルは、Developer ツールのインストールパスの次の場所に置かれています。

<Informatica Installation Directory>\clients\DeveloperClient\Tutorials

- チュートリアルファイルは、サービスのインストールパスの次の場所に置かれています。

<Informatica Installation Directory>\server\Tutorials

チュートリアル のレッスンでは次のファイルが必要です。

- All_Customers.csv
- Boston_Customers.csv
- Customer_Order.xsd
- LA_customers.csv
- orders.csv

Informatica Analyst のチュートリアル

このチュートリアルでは、アナリストが Analyst ツールにログインし、プロジェクトとフォルダを作成し、プロファイルとルールを作成し、データのスコアカードを作成して、参照テーブルを作成します。

以下の表に、実行可能なレッスンを示します。

レッスン	説明
レッスン 1。 Informatica Analyst の設定	Analyst ツールにログインし、チュートリアル のレッスン用のプロジェクトとフォルダーを作成します。
レッスン 2。 データオブジェクトの作成	フラットファイルをデータオブジェクトとしてインポートし、データをプレビューします。
レッスン 3。 クイックプロファイルの作成	クイックプロファイルを作成して、データ品質を迅速に把握します。
レッスン 4。 カスタムプロファイルの作成	カスタムプロファイルを作成して、カラム、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設定します。
レッスン 5。 式ルールの作成	式ルールを作成して、カラム値を変更およびプロファイルします。
レッスン 6。 スコアカードの作成と実行	スコアカードを作成および実行して、長期間にわたってデータ品質の進捗を測定します。

レッスン	説明
レッスン 7。プロファイル結果からの参照テーブルの作成	ソースデータの標準化に使用できる参照テーブルを作成します。
レッスン 8。参照テーブルの作成	参照テーブルを作成して、ソースデータと有効な標準値の間にリレーションシップを確立します。

Informatica Developer tool

このチュートリアルでは、Developer tool を使用していくつかのデータ品質操作を実行します。

Developer tool を使用して、データの内容および構造を分析するプロファイルを作成し、実行できます。

以下の表に、実行可能なレッスンを示します。

レッスン	説明
レッスン 1。Informatica Developer の設定	ドメイン内でモデルリポジトリサービスによって管理されるモデルリポジトリへの接続を作成します。チュートリアルのレッスンの作業を格納するプロジェクトとフォルダを作成します。デフォルトのデータ統合サービスを選択します。
レッスン 2。物理データオブジェクトのインポート	物理データオブジェクトに関連付けられた顧客のデータファイルのデータ品質プロセスを定義します。
レッスン 3。データソースでのプロファイルの実行	プロファイルを実行すると、プロファイルの結果により、データの内容および構造が明らかになります。 プロファイルでは、結合分析（2 つのデータカラム間で有効な結合が可能であるかどうかを判断する分析の形式）を実行することができます。

パート I: Informatica Analyst の 基本操作

この部には、以下の章があります。

- [レッスン 1. Informatica Analyst の設定, 19 ページ](#)
- [レッスン 2. データオブジェクトの作成, 22 ページ](#)
- [レッスン 3. デフォルトプロファイルの作成, 25 ページ](#)
- [レッスン 4. カスタムプロファイルの作成, 29 ページ](#)
- [レッスン 5. 式ルールの作成, 33 ページ](#)
- [レッスン 6. スコアカードの作成と実行, 36 ページ](#)
- [レッスン 7. プロファイルカラムからの参照テーブルの作成, 42 ページ](#)
- [レッスン 8. 参照テーブルの作成, 45 ページ](#)

第 2 章

レッスン 1. Informatica Analyst の設定

この章では、以下の項目について説明します。

- [Informatica Analyst の設定の概要, 19 ページ](#)
- [タスク 1. Informatica Analyst へのログイン, 20 ページ](#)
- [タスク 2. プロジェクトの作成, 20 ページ](#)
- [タスク 3. フォルダーの作成, 21 ページ](#)
- [Informatica Analyst の設定のまとめ, 21 ページ](#)

Informatica Analyst の設定の概要

このチュートリアルレッスンを開始する前に、Analyst ツールを設定する必要があります。Analyst ツールを設定するには、Analyst ツールにログインし、作業内容を格納するプロジェクトおよびフォルダーを作成します。

Informatica ドメインは、Informatica 環境を定義するノードとサービスの集合です。ドメイン内のサービスとして、アナリストサービスとモデルリポジトリサービスがあります。アナリストサービスは Analyst ツールを実行するサービスで、モデルリポジトリサービスはモデルリポジトリを管理するサービスです。Analyst ツールでの作業で作成したアセットは、モデルリポジトリに格納されます。

Analyst ツールでアセットを作成する前に、プロジェクトを作成する必要があります。プロジェクトには、Analyst ツールのアセットが含まれます。プロジェクトには、同じビジネス要件の一部であるデータオブジェクトなどの関連するアセットを格納するフォルダーも含まれます。

目標

このレッスンでは、以下の作業を完了します。

- Analyst ツールにログインする。
- Analyst ツールで作成したアセットを格納するプロジェクトを作成する。
- 関連するアセットを格納するプロジェクト内にフォルダーを作成する。

要件

このレッスンを開始する前に、次の要件を確認してください。

- 管理者が Administrator ツールでモデルリポジトリサービスとアナリストサービスの設定を完了している。

- Analyst ツールのホスト名とポート番号がわかっている。
- アナリストサービスにアクセスするためのユーザー名とパスワードがわかっている。この情報は管理者から提供されます。

時間

このレッスンの完了には、5～10 分見ておいてください。

タスク 1. Informatica Analyst へのログイン

Analyst ツールにログインし、チュートリアルを開始します。

1. Microsoft Internet Explorer または Google Chrome ブラウザを起動します。
2. [アドレス] フィールドに、次に示す Informatica Analyst の URL を入力します。
`http[s]://<fully qualified host name>:<port number>/analyst`
3. ドメインで LDAP またはネイティブ認証を使用する場合、ログインページ用のユーザー名とパスワードを入力します。
4. [ネイティブ] または特定のセキュリティドメインの名前を選択します。
Informatica ドメインで LDAP または Kerberos 認証を使用している場合は、[セキュリティドメイン] フィールドが表示されます。ユーザーアカウントが所属するセキュリティドメインがわからない場合は、Informatica ドメイン管理者に問い合わせてください。
5. [ログイン] をクリックします。
Analyst ツールにより、[開始] ワークスペースが開きます。

タスク 2. プロジェクトの作成

このタスクでは、Analyst ツールで作成したアセットを含むプロジェクトを作成します。プロジェクトのフォルダーを含むようにチュートリアルプロジェクトを作成します。

1. [管理] ヘッダーの [プロジェクト] をクリックします。
[プロジェクト] ワークスペースが表示されます。
2. [アクション] メニューから、[新規] > [プロジェクト] を選択します。
[新規プロジェクト] ウィンドウが表示されます。
3. プロジェクト名として「Tutorial_自分の名前」を入力します。
4. [OK] をクリックします。

タスク 3. フォルダーの作成

このタスクでは、関連するアセットを格納するフォルダーを作成します。フォルダーは、プロジェクトまたは別のフォルダー内に作成することができます。Data Quality プロジェクトに関連するアセットを格納する Customers という名前のフォルダーを作成します。

1. **【プロジェクト】** パネルで、チュートリアルプロジェクトを選択します。
2. **【アクション】** メニューから、**【新規】** > **【フォルダー】** をクリックします。
【新しいフォルダー】 ウィンドウが表示されます。
3. フォルダー名に「Customers」を入力します。
4. **【OK】** をクリックします。
フォルダーにはチュートリアルプロジェクトの下に表示されます。

Informatica Analyst の設定のまとめ

このレッスンでは、Analyst ツールではプロジェクトおよびフォルダーにアセットが格納されることを学習しました。プロジェクトとフォルダーはモデルリポジトリに格納されます。Analyst ツールは、アナリストサービスで実行されます。モデルリポジトリサービスでは、モデルリポジトリが管理されます。アナリストサービスとモデルリポジトリサービスは、Informatica ドメイン内のアプリケーションサービスです。

Analyst ツールにログインし、プロジェクトとフォルダーを作成しました。

これで、Analyst ツールを使用してこのチュートリアルその他のレッスンを実行できます。

第 3 章

レッスン 2. データオブジェクトの作成

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の作成の概要, 22 ページ](#)
- [タスク 1. フラットファイルデータオブジェクトの作成, 23 ページ](#)
- [タスク 2. データオブジェクトのプロパティの表示, 23 ページ](#)
- [データオブジェクトの作成のまとめ, 24 ページ](#)

データオブジェクトの作成の概要

Analyst ツールで、データオブジェクトはフラットファイルまたはリレーショナルデータベーステーブルに基づいたデータの表現です。フラットファイルまたはテーブルオブジェクトを作成し、そのフラットファイルまたはリレーショナルデータベーステーブル内のデータに対してプロファイルを実行します。Analyst ツール内でフラットファイルデータオブジェクトを作成すると、Analyst ツールを実行するマシン上のフラットファイルキャッシュにファイルをアップロードすることや、フラットファイルが格納されるネットワーク上の場所を指定することができます。

内容

HypoStores は、Los Angeles の顧客データをフラットファイルに保存しています。HypoStores は、そのデータをプロファイルおよび分析し、データ品質タスクを実行する必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. フラットファイルをフラットファイルキャッシュの場所にアップロードし、データオブジェクトを作成します。
2. フラットファイルデータオブジェクトのデータをプレビューします。

要件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 を完了している。
- LA_Customers.csv フラットファイルがある。このファイルは、<Installation Root Directory>\<Release Version>\clients\DeveloperClient\Tutorials フォルダにあります。

時間

このタスクの完了には、5～10 分見ておいてください。

タスク 1. フラットファイルデータオブジェクトの作成

このタスクでは、**フラットファイルの追加**ウィザードを使用して、LA_Customers データファイルからフラットファイルデータオブジェクトを作成します。

1. **プロジェクトナビゲータ**で、チュートリアルプロジェクト内の Customers フォルダを選択します。

注: フラットファイルデータオブジェクトの作成前に、それを作成するプロジェクトまたはフォルダを選択する必要があります。

2. **【新規】 > 【フラットファイルデータオブジェクト】** をクリックします。

【検出】 ワークスペースに、**フラットファイルの追加**ウィザードが表示されます。

3. **【参照およびアップロード】** を選択し、**【参照】** をクリックします。

4. LA_Customers.csv の場所を参照し、**【開く】** をクリックします。

5. **【次へ】** をクリックします。

【インポートの種類の選択】 ページには、**【区切り】** オプションと **【固定長】** オプションが表示されます。デフォルトのオプションは **【区切り】** です。

6. **【次へ】** をクリックします。

7. **【データで使用されている区切り文字とテキスト修飾子を指定する】** で、テキスト修飾子として **【二重引用符】** を選択します。

8. **【インポートする行を指定します】** で、**【先頭行からインポート】** を選択して空ではない先頭行の列名をインポートします。

9. **【次へ】** をクリックします。

【列属性】 セクションに、列ごとの名前、データ型、精度、スケール、および形式が表示されます。

10. **【次へ】** をクリックします。

【名前】 フィールドに **LA_Customers** が表示されます。

11. 必要に応じて、ファイルの名前を変更し、説明を追加します。

12. **【フォルダー】** セクションで **Customers** フォルダを選択します。

13. **【完了】** をクリックします。

LA_Customers データオブジェクトが **【検出】** ワークスペースにソースデータのプレビューとともに表示されます。

タスク 2. データオブジェクトのプロパティの表示

このタスクでは、LA_Customers データオブジェクトのプロパティを表示できます。

1. **【開く】** をクリックして **【ライブラリ】** ワークスペースを開きます。

2. **【ライブラリ】** ワークスペースで、**【アセット】** パネルの **【データオブジェクト】** をクリックします。
【データオブジェクト】 パネルにデータオブジェクトのリストが表示されます。
3. **LA_Customers** フラットファイルをクリックします。
【データプレビュー】 パネルとともに、LA_Customers データオブジェクトから取得したデータが表示されます。
4. **【データプレビュー】** パネルで、LA_Customers データオブジェクトの構造および内容を確認します。
Analyst ツールによって、フラットファイルデータオブジェクトの最初の 100 行が表示されます。
5. **【プロパティ】** をクリックします。
【プロパティ】 パネルには、データオブジェクトの名前、タイプ、説明、および場所が表示されます。データオブジェクトのカラム名とカラムプロパティも表示できます。

データオブジェクトの作成のまとめ

このレッスンでは、データオブジェクトがフラットファイルまたはリレーショナルデータベースソースに基づいたデータの表現であることを学習しました。フラットファイルデータオブジェクトを作成し、その中のデータをプレビューできることを学習しました。

フラットファイルをアップロードしてフラットファイルデータオブジェクトを作成し、データオブジェクトのデータをプレビューし、データオブジェクトのプロパティを表示しました。

データオブジェクトの作成後、レッスン 3 でデータオブジェクトのデフォルトプロファイルを作成し、レッスン 4 でデータオブジェクトのカスタムプロファイルを作成します。

第 4 章

レッスン 3. デフォルトプロファイルの作成

この章では、以下の項目について説明します。

- [デフォルトプロファイルの作成の概要, 25 ページ](#)
- [タスク 1. デフォルトプロファイルの作成と実行, 26 ページ](#)
- [タスク 2. サマリビューでのプロファイル結果の表示, 26 ページ](#)
- [タスク 3. 詳細ビューでのプロファイル結果の表示, 27 ページ](#)
- [デフォルトプロファイルの作成のサマリ, 28 ページ](#)

デフォルトプロファイルの作成の概要

プロファイルは、データの内容と構造に基づいたデータ品質の分析です。デフォルトプロファイルは、デフォルトのオプションで作成するプロファイルです。プロファイルにカラムとオプションを設定せずにプロファイル結果を取得するには、デフォルトプロファイルを使用します。

データ品質プロジェクトの開始時に、デフォルトプロファイルを作成および実行してデータの品質を分析します。デフォルトプロファイルオブジェクトの作成時には、分析するデータオブジェクトとデータオブジェクトカラムを選択します。デフォルトプロファイルでは、プロファイルカラムとオプションの設定がスキップされます。Analyst ツールは、フラットファイルデータオブジェクトの実フラットファイルでプロファイリングを実行します。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。データはデータウェアハウスに組み込む前にクレンジングする必要があります。あなたは、データの品質を評価し、そのデータのクレンジングを担当する開発者に情報を渡すアナリストです。プロファイル結果を迅速に表示し、データの品質についての基本的な考えを得る必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. LA_Customers フラットファイルデータオブジェクトのデフォルトプロファイルを作成および実行します。
2. プロファイル結果を表示します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 2 までを完了している。

時間

このレッスンの完了には、5～10 分見ておいてください。

タスク 1. デフォルトプロファイルの作成と実行

このタスクでは、データオブジェクト内のすべてのカラムのデフォルトプロファイルを作成し、デフォルトのサンプリングオプションとドリルダウンオプションを使用します。

1. **【ライブラリ】** ワークスペースの **【アセット】** パネルでデータオブ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て、**【プロファイルの作成】** 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。
3. **【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。
4. **【全般プロパティの指定】** 画面では、デフォルトで、名前、説明、場所が入力されています。
5. **【次へ】** をクリックします。
6. **【ソースの選択】** 画面では、デフォルトでデータオブジェクトが表示されます。**【カラムを選択】** ペインでカラムを表示できます。
7. **【次へ】** をクリックします。
8. **【設定の指定】** 画面では、次のオプションがデフォルトで選択されています。
 - **カラムプロファイルの実行**
 - **【次に対してプロファイルを実行】** ペインの **【すべての行】**
 - **【ドリルダウン】** ペインの **【ライブ】**
 - **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。**
 - **【ランタイム環境】** ペインの **【ネイティブ】** 接続。
9. **【次へ】** をクリックします。
10. **【ルールとフィルタの指定】** 画面で、**【保存して実行】** をクリックしてプロファイルを作成し、実行します。

Analyst ツールでプロファイルを作成して実行します。サマリビューにプロファイル結果が表示されます。

タスク 2. サマリビューでのプロファイル結果の表示

このタスクでは、LA_Customers プロファイルのサマリビューを使用して、プロファイル結果の簡単な概要を表示します。

1. **【ライブラリ】** > **【アセット】** > **【プロファイル】** ペインで、**LA_Customers** プロファイルをクリックします。
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、**【フィルタ基準】** ペインの **【カラム】** をクリックして、カラムのプロファイル結果を表示します。
3. パーセントで値を表示するには、水平棒グラフ上にマウスのポインタを移動します。
4. **【NULL 重複なし 重複あり %】** セクションに、カラムの Null 値、重複しない値、重複する値をパーセントで表示できます。
5. **【パターン】** セクションでは、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。
6. **【長さ】** セクションでは、カラム内の値の最小長および最大長を表示できます。
7. **【値】** セクションでは、カラム内の最小値および最大値を表示できます。
8. **【データ型】** セクションでは、マウスポインタを値の上に移動すると、カラムのすべての推測されたデータ型と文書化されたデータ型を表示できます。
9. **【データドメイン】** セクションでは、マウスポインタを値の上に移動すると、カラムのすべての推測されるデータドメインを表示できます。
10. 異常値データを表示するには、**【アクション】** > **【異常値の検出】** をクリックして、プロファイル結果で異常値を検出します。
11. プロファイル結果に含まれる異常値を表示するには、**【パターン異常値】** または **【値頻度の異常値】** フィルタをクリックします。
12. 詳細ビューにカラムのプロファイル結果を表示するには、カラム名をクリックします。

タスク 3. 詳細ビューでのプロファイル結果の表示

このタスクでは、LA_Customers プロファイルの詳細ビューを使用して、プロファイル結果に含まれるカラムの簡単な概要を表示します。

1. 詳細ビューでカラムのプロファイル結果を表示するには、サマリビューでカラム名をクリックします。
カラムのプロファイル結果が詳細ビューに表示されます。
2. **【全般】** ペインでは、水平棒グラフとパーセント値で、NULL、重複しない、および重複する値を表示できます。
3. **【値】** ペインでは、上部パネルのすべての値を表示できます。スライダを上部パネルの値上にスライドさせると、そのスライダの範囲内のすべての値が下部パネルに表示されます。
4. **【データプレビュー】** ペインで、**【アクション】** > **【プレビューの表示】** をクリックしてソースデータを表示します。
5. **【パターン】** ペインでは、水平棒グラフとパーセント値で、カラム内の推測されたパターンを表示できます。
6. **【データ型】** ペインでは、水平棒グラフとパーセント値で、カラム内の推測されたデータ型を表示できます。
7. **【データドメイン】** ペインでは、推測されたデータドメインを表示できます。

デフォルトプロファイルの作成のサマリ

このレッスンでは、デフォルトプロファイルを使用すると、プロファイルにカラムと行のサンプリングオプションを設定せずにプロファイル結果が表示されることを学習しました。データ品質プロジェクトの開始時に、デフォルトプロファイルを作成および実行してデータの品質を分析することも学習しました。また、Analyst ツールにより、フラットファイルデータオブジェクトの実フラットファイルでプロファイリングが実行されることを学習しました。

デフォルトプロファイルを作成し、プロファイル結果を分析しました。NULL 値やデータ型など、プロファイル内のカラムに関する詳細な情報を取得しました。また、カラム値とパターンを使用して、データ品質の問題を特定しました。

クイックプロファイルの結果を分析すると、次のタスクを完了できます。

- カスタムプロファイルを作成して、プロファイルからカラムを除外し、関心のあるカラムのみを組み込みます。
- 式ルールを作成して仮想カラムを作成し、それらをプロファイルします。
- カラムの有効な値を含める参照テーブルを作成します。

第 5 章

レッスン 4. カスタムプロファイルの作成

この章では、以下の項目について説明します。

- [カスタムプロファイルの作成の概要, 29 ページ](#)
- [タスク 1. カスタムプロファイルの作成, 30 ページ](#)
- [タスク 2. プロファイルの実行, 31 ページ](#)
- [タスク 3. プロファイル結果でのドリルダウン, 31 ページ](#)
- [カスタムプロファイルの作成のまとめ, 32 ページ](#)

カスタムプロファイルの作成の概要

プロファイルは、データの内容と構造に基づいたデータ品質の分析です。カスタムプロファイルは、プロファイリングを迅速に行うために、カラム、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の設定時に作成するプロファイルです。ソースのサンプル行を選択するには、サンプリングオプションを設定します。ソースデータまたはステージングされたデータでデータ行にドリルダウンするには、ドリルダウンオプションを設定します。プロファイルは、Hive またはネイティブの環境で実行することを選択できます。

データ品質プロジェクトの開始時に、プロファイルを作成および実行してデータの品質を分析します。プロファイルオブジェクトを作成するときは、プロファイルを実行するデータオブジェクトとデータオブジェクトカラムを選択します。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。HypoStores は、LA の顧客データファイル内の顧客層データの品質にアクセスする必要があります。あなたは、データの品質を評価し、そのデータのクレンジングを担当する開発者に情報を渡すアナリストです。

目標

このレッスンでは、以下の作業を完了します。

1. フラットファイルデータオブジェクトのカスタムプロファイルを作成し、NULL 値のカラムを除外します。
2. プロファイルを実行して、CustomerTier カラムの内容と構造を分析します。
3. プロファイル結果の行にドリルダウンします。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1、2、および 3 を完了している。

時間

このレッスンの完了には、5～10 分見ておいてください。

タスク 1. カスタムプロファイルの作成

このタスクでは、カスタムプロファイルを作成します。カスタムプロファイルの作成時には、プロファイルを実行するデータオブジェクトとカラムを選択します。サンプリングオプションとドリルダウンオプションも設定します。

1. **【新規】 > 【プロファイル】** をクリックします。
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、次のオプションを設定します。
 - **【名前】** フィールドに、「**Profile_LA_Customers**」と入力します。
 - **【場所】** フィールドで、**Customers** フォルダを選択します。
4. **【次へ】** をクリックします。
5. **【ソースの選択】** 画面で、**【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。
6. **【データオブジェクトの選択】** ダイアログボックスで、**【LA_Customers】** を選択します。**【OK】** をクリックします。
7. **【ソースの選択】** 画面で、Address2、Address3、および City2 の各カラムをクリアします。
8. **【次へ】** をクリックします。
9. **【設定の指定】** 画面で、次のオプションを設定します。
 - **【カラムプロファイルの実行】**。
 - **【次に対してプロファイルを実行】** ペインで、**【ランダムサンプリング】** オプションを選択します。
 - **【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】** オプションを選択します。
 - **【ドリルダウン】** ペインで **【ステージングされた】** オプションを選択します。
10. **【次へ】** をクリックします。
11. **【ルールとフィルタの指定】** 画面で **【保存して完了】** をクリックして、プロファイルを作成、実行します。
Analyst ツールによってプロファイルが作成され、**【検出】** ワークスペースにプロファイルが表示されます。プロファイルを実行して結果を表示する必要があります。

タスク 2. プロファイルの実行

このタスクでは、プロファイルを実行してデータオブジェクトについてプロファイリングを実行し、プロファイル結果を表示します。Analyst ツールは、フラットファイルデータオブジェクトのステージングされたフラットファイルでプロファイリングを実行します。

1. **【検出】** ワークスペースが表示されていることを確認します。
ワークスペースに **Profile_LA_Customers_Custom** プロファイルが表示されています。
2. ワークスペースで、**Profile_LA_Customers_Custom** をクリックします。
3. プロファイル画面が表示されます。この画面でプロファイルの編集またはプロファイルの実行を選択できます。**【実行】** をクリックします。
4. サマリビューにプロファイル結果が表示されます。

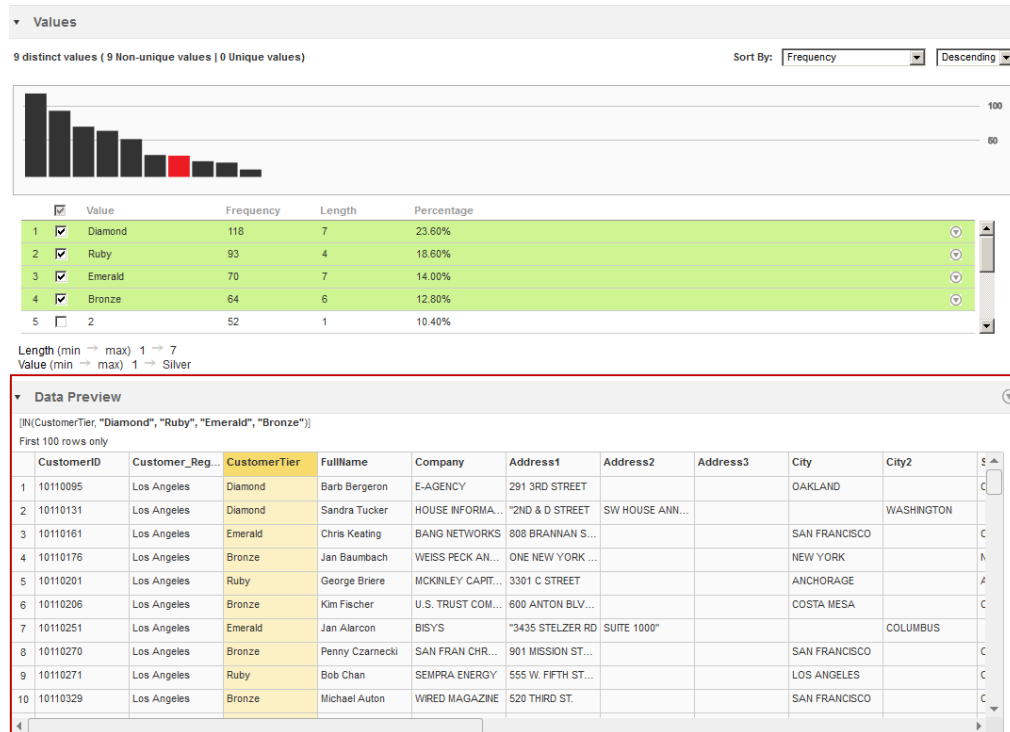
タスク 3. プロファイル結果でのドリルダウン

このタスクでは、CustomerTier カラムの値でドリルダウンして、プロファイルのデータオブジェクト内のソース行を確認します。

1. **Profile_LA_Customers** プロファイルで、プロファイル結果のサマリビューを開いていることを確認します。
2. **【CustomerTier】** カラムをクリックします。
カラムのプロファイル結果が詳細ビューに表示されます。

3. 詳細ビューで、Diamond、Ruby、Emerald、および Bronze の値を選択します。【値】ペインで値を右クリックし、【ドリルダウン】を選択します。

値が Diamond、Ruby、Emerald、Bronze のカラムの行が【データプレビュー】ペインに表示されます。次の画像は、Diamond、Ruby、Emerald、または Bronze の値でドリルダウンするときに【データプレビュー】ペインに表示されるドリルダウン結果を示しています。



【データプレビュー】ペインには、選択したカラムの最初の 100 行が表示されます。【データプレビュー】ペインのタイトルには、ソースカラムで使用されたロジックが表示されます。

カスタムプロファイルの作成のまとめ

このレッスンでは、プロファイルされるカラムを設定できることと、サンプリングオプションとドリルダウンオプションを設定できることを学習しました。ドリルダウンしてカラム値の基礎となる行を確認できることと、カラム値を表示するときに含まれるカラムを設定できることを学習しました。

CustomerTier カラムが含まれるカスタムプロファイルを作成し、結果内の CustomerTier カラムの基礎となる行にドリルダウンしました。

このカスタムプロファイルオブジェクトを使用して、レッスン 5 で式ルールを作成します。

第 6 章

レッスン 5. 式ルールの作成

この章では、以下の項目について説明します。

- [式ルールの作成の概要, 33 ページ](#)
- [タスク 1. 式ルールの作成とプロファイルの実行, 34 ページ](#)
- [タスク 2. 式ルールの出力の表示, 34 ページ](#)
- [タスク 3. 式ルールの編集, 35 ページ](#)
- [式ルールの作成のまとめ, 35 ページ](#)

式ルールの作成の概要

式ルールは、式関数とソースカラムを使用してルールの論理を定義します。式ルールを作成して、Analyst ツールのプロファイルに追加することができます。式ルールは、1 つまたは複数のプロファイルと関連付けることができます。

式ルールの出力は、プロファイル内の仮想カラムです。プロファイルを実行すると、Analyst ツールによって仮想カラムがプロファイルされます。

式ルールを使用して、ソースカラムを検証したり、ソースカラムの値に基づいて追加のソースカラムを作成したりすることができます。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。HypoStores は、顧客名を分析して、名と姓に分離する必要があります。HypoStores は、式ルールを使用して名と姓が含まれるカラムを別々の仮想カラムに解析し、それらのカラムをプロファイルする必要があります。HypoStores は、これらのルールの出力を分析する必要がある他のアナリストがそれらのルールを使用できるようにする必要もあります。

目標

このレッスンでは、以下の作業を完了します。

1. 式ルールを作成して、FullName カラムを名カラムと姓カラムに分離します。氏名から名を分離するルールを作成します。名から姓を分離するもう 1 つのルールを作成します。これらのルールは Profile_LA_Customers プロファイルに対して作成します。
2. プロファイルを実行し、プロファイル内のルールの出力を表示します。
3. ルールを編集して、他の Analyst ツールユーザーが使用できるようにします。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- レッスン 1～4 を完了している。

時間

このレッスンの所要時間は 10～15 分です。

タスク 1. 式ルールを作成とプロファイルの実行

このタスクでは、2 つの正規表現を作成して、FullName カラムを FirstName および LastName という名前の 2 つの仮想カラムに解析します。ルール名は FirstName と LastName です。

1. **【ライブラリ】** ワークスペースで、**Profile_LA_Customers** プロファイルをクリックします。
サマリビューにプロファイル結果が表示されます。
2. **【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】** 画面をクリックします。
4. **【ルール】** ペインで、**【アクション】** > **【ルールの作成】** をクリックします。
5. **【名前】** フィールドに、「**FirstName**」と入力します。
6. **【式】** セクションで次の式を入力して、**【名前】** カラムから名前を分離します。
`SUBSTR(FullName,1,INSTR(FullName,' ', -1,1) - 1)`
7. **【検証】** をクリックします。
8. **【OK】** をクリックします。
9. 手順 4～8 を繰り返して、**LastName** という名前のルールを作成します。次の式を入力して、**【名前】** カラムから姓を分離します。
`SUBSTR(FullName,INSTR(FullName,' ', -1,1),LENGTH(FullName))`
10. **【保存して実行】** をクリックし、プロファイルを保存して実行します。

タスク 2. 式ルールの出力の表示

このタスクでは、プロファイルの実行後に名と姓を分離する式ルールの出力を表示します。

1. サマリビューで **【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
2. プロファイルウィザードで、**【ソースの選択】** をクリックします。
3. **【ソースの選択】** 画面で、ツールバーの **【名前】** の横にあるチェックボックスを選択して、すべてのカラムをクリアします。
【カラム】 セクションでは、少なくとも 1 つのカラムを選択する必要があるため、デフォルトでカラムの 1 つが選択されています。
4. **【FullName】** カラム、**【FirstName】** ルール、および **【LastName】** ルールを選択します。
5. 他のカラムが選択されていれば、それをクリアします。

6. **【保存して実行】** をクリックします。
サマリビューにプロファイル結果が表示されます。
7. **【FirstName】** ルールをクリックすると、ルールのプロファイル結果が詳細ビューに表示されます。
8. **【値】** ペインで値を選択します。その値を右クリックして、**【ドリルダウン】** をクリックします。
【FullName】 カラムの値、および **【FirstName】** ルールと **【LastName】** ルールが、他のカラムの値と一緒に **【データプレビュー】** ペインに表示されます。Analyst ツールにより、**【FullName】** カラムが名と姓に分離されています。

タスク 3. 式ルールの編集

このタスクでは、すべての Analyst ツールユーザーに対して式ルールを再利用可能および使用可能にします。

1. **Profile_LA_Customers** プロファイルのサマリビューで、**【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。
2. **【ルールとフィルタの指定】** 画面をクリックします。
3. **【ルールとフィルタの指定】** 画面で、**【FirstName】** ルールを選択して、**【アクション】** > **【ルールの編集】** をクリックします。
【ルールの編集】 ダイアログボックスが表示されます。
4. **【このルールを再利用可能なルールとして保存しますか?】** オプションを選択してから、ルールを保存する場所を選択します。
5. **【OK】** をクリックします。
6. **【LastName】** ルールを選択し、手順 3~5 を繰り返します。
7. **【保存して完了】** をクリックしてプロファイルを保存します。

Analyst ツールのすべてのユーザーは、FirstName ルールと LastName ルールを使用して、姓名を含むカラムを個別のカラムに分割できます。

式ルールの作成のまとめ

このレッスンでは、式ルールが式関数とソースカラムを使用してルールの論理を定義することを学習しました。式ルールの出力はプロファイル内の仮想カラムであることを学習しました。プロファイルを実行すると、Analyst ツールによって仮想カラムが組み込まれます。

2つの式ルールを作成してプロファイルに追加し、そのプロファイルを実行しました。ルールの出力を表示し、それらのルールをすべての Analyst ツールユーザーが使用できるようにしました。

第 7 章

レッスン 6. スコアカードの作成と実行

この章では、以下の項目について説明します。

- [スコアカードの作成と実行の概要, 36 ページ](#)
- [タスク 1. プロファイル結果からのスコアカードの作成, 37 ページ](#)
- [タスク 2. スコアカードの実行, 38 ページ](#)
- [タスク 3. スコアカードの表示, 38 ページ](#)
- [タスク 4. スコアカードの編集, 39 ページ](#)
- [タスク 5. しきい値の設定, 40 ページ](#)
- [タスク 6. スコアの傾向グラフの表示, 40 ページ](#)
- [スコアカードの作成と実行のまとめ, 40 ページ](#)

スコアカードの作成と実行の概要

スコアカードは、グラフィカル形式の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です。長期間にわたってデータ品質の進捗を測定して監視するには、スコアカードを使用します。

スコアカードを作成するには、プロファイル内のカラムをメトリックとしてスコアカードに追加し、メトリックにウェイトを割り当て、スコアしきい値を設定します。ソースデータに基づいて、フィルタをスコアカードに追加できます。スコアカードを実行するには、メトリックの有効値を選択します。スコアカードを実行して、メトリックのスコアを確認します。

スコアカードには、プロファイル内のカラムの値の頻度がスコアとして表示されます。スコアにはメトリックの有効値の割合が反映されます。

内容

HypoStores は、新たに確保した Los Angeles オフィスからデータウェアハウスにデータを組み込む必要があります。組織は、データをマージする前に、さまざまな顧客層と状態のデータのデータ品質が分析されていることを確認する必要があります。あなたは、データ品質分析の実行の進捗を監視するアナリストです。顧客層プロファイルカラムと状態プロファイルカラムからスコアカードを作成し、データ品質のしきい値を設定し、スコアの傾向グラフを表示して時間の経過とともにスコアがどのように改善するかを確認する必要があります。

目標

このレッスンでは、以下のタスクを完了します。

1. Profile_LA_Customers_Custom プロファイルの結果からスコアカードを作成して、CustomerTier カラムと State カラムのスコアを表示します。
2. スコアカードを実行して、CustomerTier カラムと State カラムのスコアを生成します。
3. スコアカードを表示して、各カラムのスコアを確認します。
4. スコアカードを編集して、スコアのさまざまな有効値を指定します。
5. スコアのしきい値を設定し、スコアカードを実行します。
6. スコアの傾向グラフを表示して、時間の経過とともにスコアがどのように改善するかを確認します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 から 5 までを完了している。

時間

このレッスンのタスクの所要時間は 15 分です。

タスク 1. プロファイル結果からのスコアカードの作成

このタスクでは、Profile_LA_Customers_Custom プロファイルからスコアカードを作成して、CustomerTier カラムと State カラムの値をスコアリングします。

1. **【ライブラリ】** ワークスペースで、**Profile_LA_Customers** プロファイルをクリックします。
プロファイル結果のサマリビューが表示されます。
2. サマリビューで **【CustomerTier】** カラムを選択し、そのカラムを右クリックして、**【次に追加】 > 【スコアカード】** を選択します。
スコアカードに追加ウィザードが表示されます。
3. **スコアカードに追加ウィザード**では、デフォルトで **【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。
4. **【手順 2/8】** 画面で、スコアカード名を「**sc_LA_Customer**」と入力し、スコアカードの場所の **Customers** フォルダに移動します。
5. **【次へ】** をクリックします。
6. **【手順 3/8】** 画面で **【CustomerTier】** カラムと **【State】** カラムを選択し、スコアカードに追加します。
7. **【次へ】** をクリックします。
8. **【手順 4/8】** 画面では、メトリックのフィルタを作成、編集、または削除できます。このチュートリアルでは、スコアカードフィルタを作成しません。**【次へ】** をクリックします。
9. **【手順 5/8】** 画面の **【メトリック】** ペインで、**CustomerTier** メトリックを選択します。
10. **【スコアで使用: 値】** ペインですべての値を選択し、**【すべて追加】** ボタンをクリックして、選択した値を **【有効な値】** セクションに移動します。
Shift キーを使用して複数の値を選択します。

11. **【メトリック】** ペインで、**State** メトリックを選択し、**【スコアで使用: 値】** セクションで 2 文字の州コードを持つ値を選択します。
12. **【追加】** ボタンをクリックして、値を **【有効な値】** セクションに移動します。
セクションの最上部に有効値の総数および有効値の割合が表示されます。
13. **【メトリック】** セクションの各メトリックで、**【メトリックしきい値】** セクションのスコアしきい値のデフォルト設定をそのまま使用します。
14. **【次へ】** をクリックします。
15. **【手順 6/8】** 画面で、必要に応じて、メトリックを追加するメトリックグループを選択できます。デフォルトでは、Analyst ツールはメトリックを **【デフォルト】** メトリックグループに追加します。
16. **【次へ】** をクリックします。
17. **【手順 7/8】** 画面の **【デフォルト - メトリック】** ペインで、**CustomerTier** メトリックの **【ウェイト】** カラムをダブルクリックします。
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。
18. **CustomerTier** および **State** メトリックのウェイトを入力します。
19. **【次へ】** をクリックします。
20. **【手順 8/8】** 画面では、デフォルトで **【ネイティブ】** オプションが選択されています。**【保存】** をクリックして、スコアカードを作成します。
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。

タスク 2. スコアカードの実行

このタスクは、sc_LA_Customer スコアカードを実行して、CustomerTier カラムと State カラムのスコアを生成します。

1. **【ライブラリ】** ワークスペースで、**【アセット】** > **【スコアカード】** をクリックします。
2. **【スコアカード】** ペインでスコアカードを選択します。
3. **【アクション】** > **【開く】** をクリックします。
スコアカードが表示されます。
4. **【アクション】** > **【スコアカードの実行】** をクリックします。
【スコアカードの実行】 ダイアログボックスが表示されます。
5. ダイアログボックスの設定を確認し、**【実行】** をクリックします。
【スコアカード】 ワークスペースに、CustomerTier カラムと State カラムのスコアが表示されます。

タスク 3. スコアカードの表示

このタスクでは、sc_LA_Customer スコアカードを表示して、CustomerTier カラムと State カラムのスコアを確認します。

1. 表示する State スコアが含まれる **State** 行を選択します。

[**sc_LA_Customer - メトリック**] セクションでは、次のスコアカードのプロパティを表示できます。

- スコアカード名。
 - スコアカード内の行の総数。
 - 無効な行数。
 - スコアと水平棒グラフ。
 - スコアの傾向。スコアの傾向をクリックすると、**【傾向グラフの詳細】** 画面にその傾向を視覚的に表示することができます。
 - メトリックのウェイト。
 - 無効なデータのコスト。
 - コスト傾向。
 - データオブジェクト。データオブジェクトのデータプレビューを **【検出】** ワークスペースで表示するには、データオブジェクトをクリックします。
 - カラムまたはルール名。
 - ソースのタイプ。
 - ドリルダウンアイコン。
2. **【状態】** 行でドリルダウンアイコンをクリックします。
【状態】 カラムで無効なスコアは、**【ドリルダウン】** ペインの **【無効な行】** セクションに表示されます。
 3. **【有効な行】** を選択し、State カラムに対して有効なスコアを表示します。
 4. **【CustomerTier】** 行でドリルダウンアイコンをクリックします。
【CustomerTier】 カラムのスコアは、すべて有効です。

タスク 4. スコアカードの編集

このタスクでは、sc_LA_Customer スコアカードを編集して、Ruby の値を CustomerTier スコアに対して有効ではないと指定します。

1. **【スコアカード】** ワークスペースを開いており、sc_LA_Customer スコアカードが開かれていることを確認します。
2. **【アクション】** > **【編集】** > **【メトリック】** 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。
3. **【メトリック】** セクションで **CustomerTier** を選択します。
4. **【スコアで使用: 値】** セクションで、**Ruby** を **【有効な値】** セクションから **【使用可能な値】** セクションに移動します。
【メトリックしきい値】 セクションのデフォルト設定をそのまま使用します。
5. **【保存して実行】** をクリックし、スコアカードの変更内容を保存して、スコアカードを実行します。
6. CustomerTier スコアをもう一度表示します。
CustomerTier のスコアが 81.4 パーセントに変更されています。

タスク 5. しきい値の設定

このタスクでは、sc_LA_Customer スコアカード内の State スコアのしきい値を設定して、State カラムのデータの許容範囲を決定します。CA などの 2 文字のコードの値は許容され、Calif などの 2 文字を超えるコードは許容されません。

1. **【スコアカード】** ワークスペースを開いており、sc_LA_Customer スコアカードが開かれていることを確認します。
2. **【アクション】 > 【編集】 > 【メトリック】** 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。
3. **【メトリック】** セクションで、**State** を選択します。
4. **【メトリックしきい値】** セクションの **【正常】** スコアと **【許可不可】** スコアに次のように入力します。**【正常】** スコアに 90~100%、**【許可不可】** スコアに 0~50%、**【許可】** スコアに 51%~89% の範囲を入力します。
しきい値は許容可能な範囲と正常な範囲の下方境界を表します。
5. **【保存して実行】** をクリックし、スコアカードの変更内容を保存して、スコアカードを実行します。
【スコアカード】 パネルで、スコアの割合の変更およびバーとして表示される State スコアのスコアを表示します。

タスク 6. スコアの傾向グラフの表示

このタスクでは、State スコアの傾向グラフを表示します。傾向グラフを表示して、スコアを時系列で監視することができます。

1. **【スコアカード】** ワークスペースを開いており、sc_LA_Customer スコアカードが開かれていることを確認します。
2. **【状態】** 行を選択します。
3. **【アクション】 > 【傾向グラフの表示】** をクリックするか、または **【スコアの傾向】** カラムの下矢印をクリックします。
【傾向グラフの詳細】 ダイアログボックスが表示されます。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変わります。
4. グラフ内の任意の円をポイントすると、グラフの下部の **【有効な値】** セクションに有効な値が表示されます。
5. **【閉じる】** をクリックしてスコアカードに戻ります。

スコアカードの作成と実行のまとめ

このレッスンでは、プロファイルの結果からスコアカードを作成できることを学習しました。スコアカードには、プロファイルのカラムが含まれます。スコアカードを実行してカラムのスコアを生成できることを学習し

ました。スコアカードを編集して有効値を設定し、スコアのしきい値を設定しました。スコアの傾向グラフを表示する方法も学習しました。

プロファイル内の CustomerTier カラムと State カラムからスコアカードを作成して、CustomerTier カラムと State カラムのデータ品質を分析しました。スコアカードを実行して、各カラムのスコアを生成しました。スコアカードを編集して、スコアのさまざまな有効値を指定しました。スコアのしきい値を設定し、スコアの傾向グラフを表示しました。

第 8 章

レッスン 7. プロファイルカラムからの参照テーブルの作成

この章では、以下の項目について説明します。

- [プロファイルカラムからの参照テーブルの作成の概要, 42 ページ](#)
- [タスク 1. プロファイルカラムからの参照テーブルの作成, 43 ページ](#)
- [タスク 2. 参照テーブルの編集, 44 ページ](#)
- [プロファイルカラムからの参照テーブルの作成のまとめ, 44 ページ](#)

プロファイルカラムからの参照テーブルの作成の概要

参照テーブルには、ソースデータの標準化に使用できる参照データが含まれます。参照データには、有効な標準値を含めることができます。参照テーブルを作成して、ソースデータの値と有効な標準値の間にリレーションを確立します。

参照テーブルは、プロファイルの結果から作成できます。参照テーブルの作成後、参照テーブルを編集してカラムや行を追加したり、有効な標準値を追加または編集したりすることができます。参照テーブルに加えた変更は監査証跡で表示できます。

内容

HypoStores は、データをプロファイルして異常を明らかにしたり、有効値が含まれるデータを標準化したりする必要があります。あなたは、データ内の有効値を標準化するアナリストです。プロファイルカラムの有効値に基づいて、参照テーブルを作成する必要があります。

目標

このレッスンでは、以下の作業を完了します。

1. カラムの有効値を選択することによって、Profile_LA_Customers_Custom プロファイル内の CustomerTier カラムから参照テーブルを作成します。
2. 参照テーブルを編集して、カラムのさまざまな有効値を設定します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 から 6 までを完了している。

時間

このレッスンのタスクの所要時間は 15 分です。

タスク 1. プロファイルカラムからの参照テーブルの作成

このタスクでは、参照テーブルを作成し、Profile_LA_Customers プロファイルからその参照テーブルに CustomerTier カラムを追加します。

1. **【ライブラリ】** ワークスペースで、**【アセット】** > **【プロファイル】** をクリックします。
2. **Profile_LA_Customers** プロファイルをクリックして、プロファイル結果をサマリビューで開きます。
3. サマリビューで、参照テーブルに追加する CustomerTier カラムを選択します。右クリックして **【参照テーブルへの追加】** を選択します。
【参照テーブルへの追加】 ダイアログボックスが表示されます。
4. **【参照テーブルの作成】** を選択します。
5. **【次へ】** をクリックします。
6. **【名前】** フィールドに、「**Reftab_CustTier_HypoStores**」と入力します。
7. 説明を入力し、デフォルト値として 0 を設定します。
値が含まれていないテーブルレコードにはこのデフォルト値が使用されます。
8. **【次へ】** をクリックします。
9. **【カラム属性】** セクションで、CustomerTier カラムに次のカラムプロパティを設定します。

プロパティ	説明
名前	CustomerTier
データ型	String
精度	10
説明	参照顧客層の値

10. 必要に応じて、参照テーブルの行の説明カラムを作成します。カラムの名前および精度を入力します。
11. **【プレビュー】** セクションで、CustomerTier カラムの値を確認します。
12. **【次へ】** をクリックします。
Reftab_CustomerTier_HypoStores という参照テーブル名が表示されます。オプションの説明を入力できます。
13. **【保存先】** セクションで、参照テーブルを作成するチュートリアルプロジェクトを選択します。
参照テーブル: パネルに、選択した場所の参照テーブルが一覧表示されます。
14. オプションの監査のメモを入力します。
15. **【完了】** をクリックします。

タスク 2. 参照テーブルの編集

このタスクでは、Reftab_CustomerTier_HypoStores テーブルを編集して顧客層の代替値を追加します。

1. **【ライブラリ】** ワークスペースで、**【アセット】** > **【参照テーブル】** をクリックします。
2. Reftab_CustomerTier_HypoStores 参照テーブルをクリックします。
参照テーブルが **【設計】** ワークスペースに開きます。
3. 行を編集するには、行を選択し、**【アクション】** > **【編集】** をクリックするか、**【編集】** アイコンをクリックします。
【行の編集】 ダイアログボックスが表示されます。オプションとして、複数の行を選択して同じ代替値を各行に追加します。
4. Diamond、Emerald、Gold、Silver、および Bronze 行に、それぞれ代替値 1、2、3、4、5 を入力します。
オプションの監査のメモを入力します。
5. **【適用】** をクリックして、変更を適用します。
6. **【閉じる】** をクリックします。
変更された参照テーブルの値が **【設計】** ワークスペースに表示されます。

プロファイルカラムからの参照テーブルの作成のまとめ

このレッスンでは、プロファイルの結果から参照テーブルを作成してソースデータの有効値を設定する方法を学習しました。

カラムの有効値を選択することによって、プロファイルカラムから参照テーブルを作成しました。参照テーブルを編集して、カラムのさまざまな有効値を設定しました。

第 9 章

レッスン 8. 参照テーブルの作成

この章では、以下の項目について説明します。

- [参照テーブルの作成の概要, 45 ページ](#)
- [タスク 1. 参照テーブルの作成, 46 ページ](#)
- [参照テーブルの作成のまとめ, 46 ページ](#)

参照テーブルの作成の概要

参照テーブルには、ソースデータの標準化に使用できる参照データが含まれます。参照データには、有効な標準値を含めることができます。参照テーブルを作成して、ソースデータの値と有効な標準値の間にリレーションを確立します。

参照テーブルは、参照テーブルエディタを使用して手動で作成できます。参照テーブルを使用して、ソースデータの定義や標準化を行います。参照テーブルを開発者と共有して、Developer ツールの標準化トランスフォーメーションやルックアップトランスフォーメーションで 사용할 ことができます。

内容

HypoStores は、有効な値が含まれるデータを標準化する必要があります。あなたは、データ内の有効値を標準化するアナリストです。参照テーブルを作成して、LA の顧客データを参照する標準の顧客層コードを定義する必要があります。それにより、開発者と参照テーブルを共有できます。

目標

このレッスンでは、以下のタスクを完了します。

- 参照テーブルエディタを使用して参照テーブルを作成し、LA の顧客データを参照する標準の顧客層コードを定義します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスンの 1 から 2 までを完了している。

時間

このレッスンのタスクの完了には、10 分見ておいてください。

タスク 1. 参照テーブルの作成

このタスクでは、Reftab_CustomerTier_Codes 参照テーブルを作成して、顧客層データの有効値を標準化します。

1. **【新規】** > **【参照テーブル】** をクリックします。
新しい参照テーブルウィザードが表示されます。
2. **【参照テーブルエディタを使用します】** を選択します。
3. **【次へ】** をクリックします。
4. 参照テーブルに組み込むカラムごとに、**【新しいカラムの追加】** アイコンをクリックし、各カラムのカラムプロパティを設定します。
カラム名 CustomerID、CustomerTier、Status を追加します。カラムの順序を変更したりカラムを削除したりすることができます。
5. 必要に応じて説明を入力し、デフォルト値を **0** に設定します。
値が含まれていないテーブルレコードにはこのデフォルト値が使用されます。
6. **【次へ】** をクリックします。
7. **【名前】** フィールドに「**Reftab_CustomerTier_Codes**」と入力します。
8. **【フォルダー】** セクションで、チュートリアルプロジェクトの **Customers** フォルダーを選択します。
9. **【完了】** をクリックします。
参照テーブルが **【設計】** ワークスペースに表示されます。
10. **【アクション】** メニューで **【行の追加】** を選択して、各参照テーブルカラムに次の 4 つの値を埋め込みます。
CustomerID = LA1、LA2、LA3、LA4
CustomerTier = 1、2、3、4
Status = Active、Inactive

参照テーブルの作成のまとめ

このレッスンでは、参照テーブルエディタを使用して参照テーブルを作成し、ソースデータで使用する標準の有効値を作成する方法を学習しました。

参照テーブルエディタを使用して参照テーブルを作成し、LA の顧客データの顧客層の値を標準化しました。

パート II: Informatica Developer の基本操作

この部には、以下の章があります。

- [レッスン 1. Informatica Developer の設定, 48 ページ](#)
- [レッスン 2: 物理データオブジェクトのインポート, 52 ページ](#)
- [レッスン 3. データソースでのプロファイルの実行, 60 ページ](#)

第 10 章

レッスン 1. Informatica Developer の設定

この章では、以下の項目について説明します。

- [Informatica Developer の設定の概要, 48 ページ](#)
- [タスク 1. Informatica Developer の起動, 49 ページ](#)
- [タスク 2. ドメインの追加, 49 ページ](#)
- [タスク 3. モデルリポジトリの追加, 50 ページ](#)
- [タスク 4. プロジェクトの作成, 50 ページ](#)
- [タスク 5. フォルダの作成, 50 ページ](#)
- [タスク 6. デフォルトの Data Integration Service の選択, 51 ページ](#)
- [Informatica Developer の設定のまとめ, 51 ページ](#)

Informatica Developer の設定の概要

このチュートリアルレッスンを開始する前に、Developer ツールを起動して設定する必要があります。Developer ツールを設定するには、ドメインを追加します。ドメイン内にあるモデルリポジトリを追加し、作業を格納するプロジェクトとフォルダを作成します。デフォルトの Data Integration Service も選択します。

Informatica ドメインは、Informatica 環境を定義するノードとサービスの集合です。ドメイン内のサービスには、Model Repository サービスや Data Integration Service などがあります。

Model Repository サービスでは、モデルリポジトリが管理されます。モデルリポジトリは、Developer ツールで作成す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です。プロジェクトには、Developer ツールで作成するオブジェクトが格納されます。プロジェクトには、同じビジネス要件の一部であるオブジェクトなど、関連するオブジェクトを格納するフォルダも含まれます。

Data Integration Service は、Developer ツールでデータ統合タスクを実行します。

目標

このレッスンでは、以下の作業を完了します。

- Developer ツールを起動し、Developer ツールのワークベンチに移動します。
- Developer ツールでドメインを追加します。
- プロジェクトを作成できるように、モデルリポジトリを追加します。

- Developer ツールで作成するオブジェクトを格納するプロジェクトを作成します。
- プロジェクトに、関連するオブジェクトを格納できるフォルダーを作成します。
- デフォルトの Data Integration Service を選択して、データ統合タスクを実行します。

要件

このレッスンを開始する前に、次の要件を確認してください。

- Developer ツールをインストールしている。
- ドメインに接続するためのドメイン名、ホスト名、およびポート番号がある。この情報はドメイン管理者から取得することができます。
- ドメイン管理者が Administrator ツールで Model Repository サービスを設定済みである。
- Model Repository サービスにアクセスするためのユーザー名とパスワードがある。この情報はドメイン管理者から取得することができます。
- ドメイン管理者が Data Integration Service を設定済みである。
- Data Integration Service が実行されている。

時間

このレッスンのタスクの完了には、5～10 分見ておいてください。

タスク 1. Informatica Developer の起動

Developer ツールを起動してチュートリアルを開始します。

1. Developer ツールを起動します。
Developer ツールの【ようこそ】ページが表示されます。
2. 【ワークベンチ】ボタンをクリックします。
Developer ツールのワークベンチが表示されます。

タスク 2. ドメインの追加

このタスクでは、Developer ツールにドメインを追加して、モデルリポジトリにアクセスします。

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. 【Informatica】 > 【ドメイン】を選択します。
3. 【追加】をクリックします。
【新しいドメイン】ダイアログボックスが表示されます。
4. ドメイン名、ホスト名、およびポート番号を入力します。
5. 【完了】をクリックします。
6. 【OK】をクリックします。

タスク 3. モデルリポジトリの追加

このタスクでは、プロジェクトとフォルダの格納に使用するモデルリポジトリを追加します。

1. **【ファイル】** > **【リポジトリへの接続】** をクリックします。
【リポジトリへの接続】 ダイアログボックスが表示されます。
2. **【参照】** をクリックして、Model Repository Service を選択します。
3. **【OK】** をクリックします。
4. **【次へ】** をクリックします。
5. ユーザ名とパスワードを入力します。
6. 名前空間を選択します。
7. **【完了】** をクリックします。
Object Explorer ビューにモデルリポジトリが表示されます。

タスク 4. プロジェクトの作成

このタスクでは、Developer ツールで作成するオブジェクトを格納するプロジェクトを作成します。このガイドのすべてのチュートリアルで使用するプロジェクトを 1 つ作成します。

1. **【Object Explorer】** ビューで、Model Repository Service を選択します。
2. **【ファイル】** > **【新規】** > **【プロジェクト】** をクリックします。
【新規プロジェクト】 ダイアログボックスが表示されます。
3. 「Tutorial_あなたの名前」をプロジェクト名として入力します。
4. **【完了】** をクリックします。
【Object Explorer】 ビューの **【Model Repository Service】** の下にプロジェクトが表示されます。

タスク 5. フォルダの作成

このタスクでは、関連するオブジェクトを格納するフォルダを作成します。このガイドのすべてのチュートリアルで使用するフォルダを 1 つ作成します。

1. **【Object Explorer】** ビューで、フォルダの追加先のプロジェクトを選択します。
2. **【ファイル】** > **【新規】** > **【フォルダ】** をクリックします。
3. フォルダの名前を入力します。
4. **【完了】** をクリックします。

Developer ツールは、**【Object Explorer】** ビューのプロジェクトの下にフォルダを追加します。プロジェクトを展開して、フォルダを確認してください。

タスク 6. デフォルトの Data Integration Service の選択

このタスクでは、マッピングを実行したりデータをプレビューしたりできるように、デフォルトの Data Integration Service を選択します。

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. **【Informatica】** > **【Data Integration Services】** を選択します。
3. ドメインを展開します。
4. Data Integration Service を選択します。
5. **【デフォルトとして設定】** をクリックします。
6. **【OK】** をクリックします。

Informatica Developer の設定のまとめ

このレッスンでは、Informatica ドメインに Model Repository Service と Data Integration Service が含まれることを学習しました。Model Repository Service では、モデルリポジトリが管理されます。モデルリポジトリには、プロジェクトとフォルダが含まれます。Data Integration Service は、データ統合タスクを実行します。

Developer ツールを起動し、Developer ツールを設定しました。Developer ツールにドメインを追加し、モデルリポジトリを追加し、プロジェクトとフォルダを作成しました。デフォルトの Data Integration Service も選択しました。

これで、Developer ツールを使用してこのチュートリアルの中の他のレッスンを完了することができます。

第 11 章

レッスン 2: 物理データオブジェクトのインポート

この章では、以下の項目について説明します。

- [物理データオブジェクトのインポートの概要, 52 ページ](#)
- [タスク 1. Boston_Customers フラットファイルデータオブジェクトのインポート, 53 ページ](#)
- [タスク 2. LA_Customers フラットファイルデータオブジェクトのインポート, 58 ページ](#)
- [タスク 3. All_Customers フラットファイルデータオブジェクトのインポート, 59 ページ](#)
- [物理データオブジェクトのインポートのまとめ, 59 ページ](#)

物理データオブジェクトのインポートの概要

物理データオブジェクトは、フラットファイルまたはリレーショナルデータベーステーブルに基づいたデータの表現です。フラットファイルまたはリレーショナルデータベーステーブルを物理データオブジェクトとしてインポートして、マッピングでソースまたはターゲットとして使用することができます。

内容

HypoStores Corporation では、Los Angeles オフィスと Boston オフィスの顧客データをフラットファイルに格納しています。この顧客データを Developer ツールで使用したいと考えています。そのためには、各フラットファイルを物理データオブジェクトとしてインポートする必要があります。

目標

このレッスンでは、フラットファイルを物理データオブジェクトとしてインポートします。Data Integration Service が正しいディレクトリからソースデータを読み取ることができるようにソースファイルディレクトリを設定することもできます。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 を完了している。

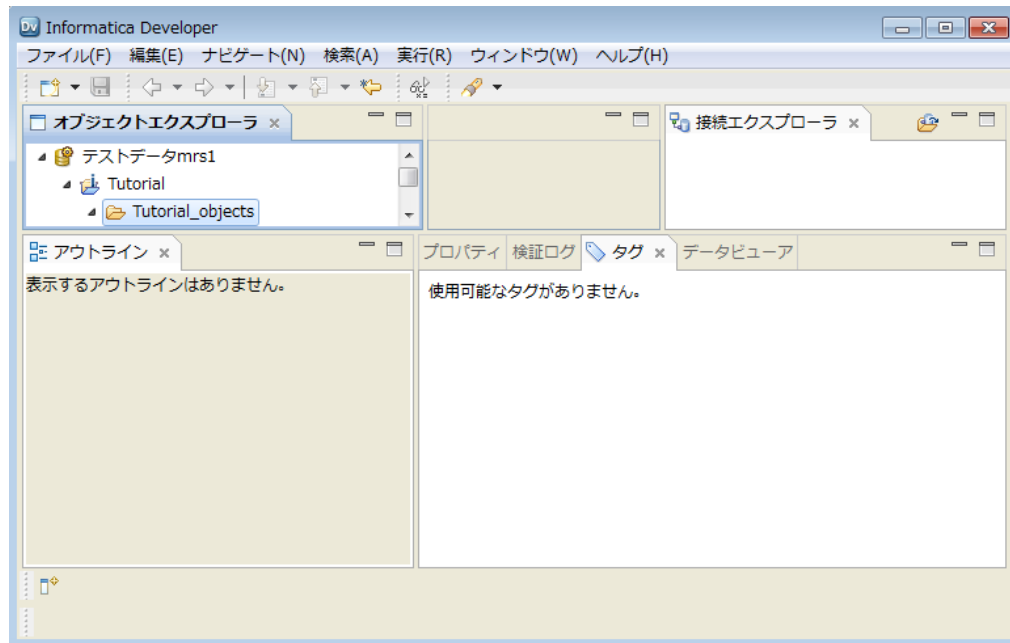
時間

このレッスンのタスクの完了には、10～15 分見ておいてください。

タスク 1. Boston_Customers フラットファイルデータオブジェクトのインポート

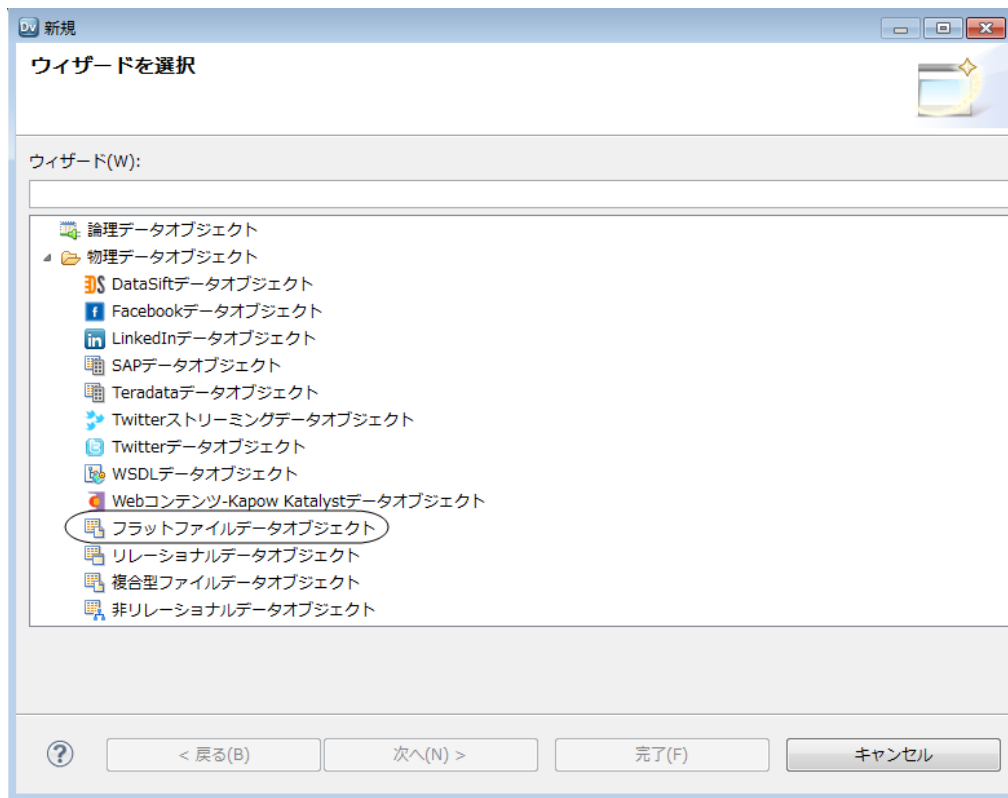
このタスクでは、Boston オフィスの顧客データが含まれるファイルから、物理データオブジェクトをインポートします。

1. **[Object Explorer]** ビューで、Tutorial_Objects フォルダを選択します。



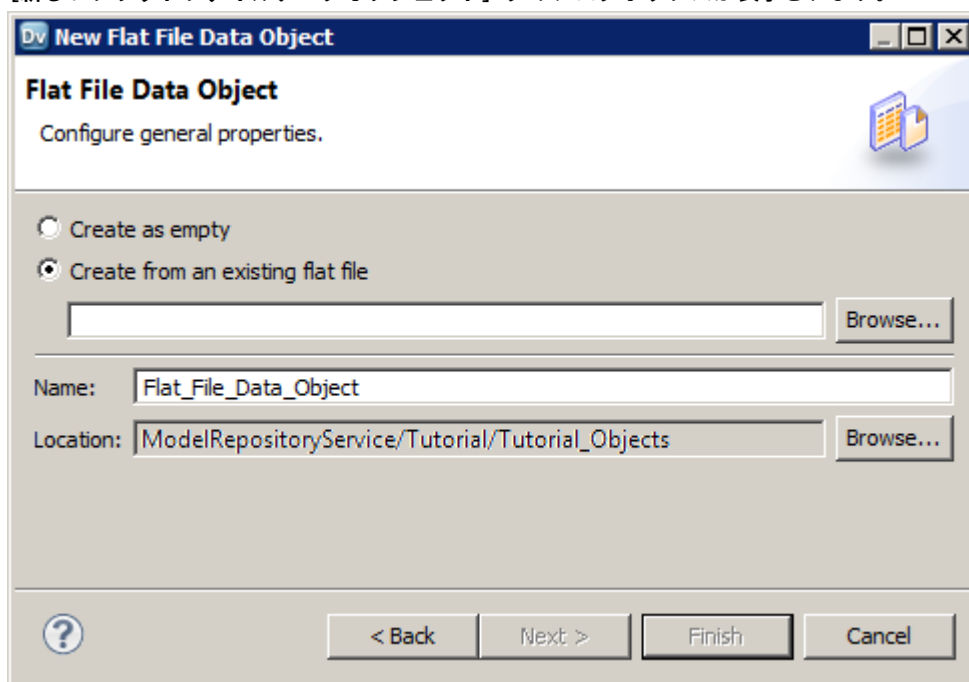
2. Tutorial_Objects フォルダを右クリックし、**[新規]** > **[データオブジェクト]** を選択します。

【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】 > 【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。



4. **【既存のフラットファイルから作成】** を選択します。
5. **【参照】** をクリックし、Developer tool マシンのディレクトリ<Informatica インストールディレクトリ>\clients\DeveloperClient\Tutorials の Boston_Customers.csv に移動します。
6. **【開く】** をクリックします。
ウィザードによって、データオブジェクトに Boston_Customers という名前が付けられます。
7. **【次へ】** をクリックします。
8. コードページが **【MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセット】** に設定され、形式が **【区切り】** に設定されていることを確認します。

【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にはデフォルトコードページ、形式、フラットファイルデータのプレビューが表示されます。

新しいフラットファイル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
コードページと形式を設定します。

コードページ(P): MS Windows Latin 1 (ANSI), superset of Latin1

形式

☒ 区切り(D) (区切り文字で区切られたフィールド)

☐ 固定幅(I) (カラム内でフィールドを整列)

プレビューする最大行数(X): 500

	フィールド1	フィールド2	フィールド3	フィールド4	
1	CustomerID	CustomerTier	Lastname	Firstname	Company
2	10110095	Diamond	Bergeron	Barb	E-AGENCY
3	10110102	Bronze	Brosseau	Derrick	BANK ONE
4	10110105		Anderson	Lauren	KPMG PEAT MARWICK

? < 戻る(B) 次へ(N) > 完了(F) キャンセル

9. **【次へ】** をクリックします。
10. **【1 行目からカラム名をインポート】** を選択します。

【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のフラットファイルデータのプレビューにカラム名が表示されます。

新しいフラットファイル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

区切り文字で区切ったプロパティを設定します。

区切り文字

☐ タブ(T) ☐ セミコロン(M) ☒ カンマ(C)

☐ スペース(P) ☐ その他(H): ...

テキスト修飾子

☒ 引用符なし(O) ☐ 一重引用符(S) ☐ 二重引用符(Q)

プレビューのオプション

☒ 1行目からカラム名をインポート(I) 次の行からインポートを開始(R):

行の区切り文字(L): ☐ 連続した区切り文字を1文字として扱う(U)

エスケープ文字(E): ☐ データ内のエスケープ文字を維持(A)

プレビューする最大行数(X):

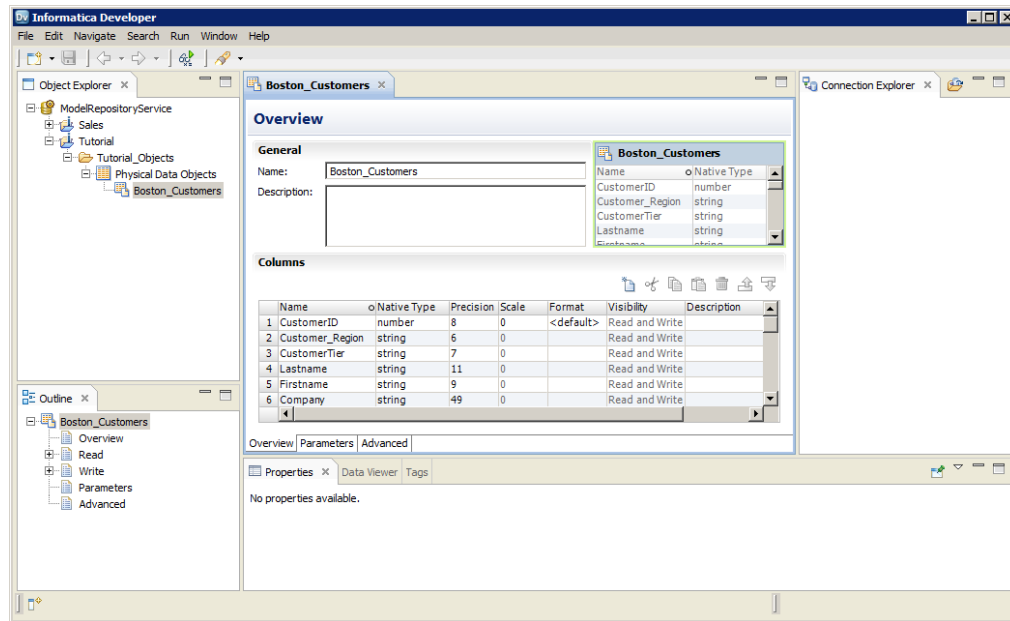
	CustomerID	CustomerTier	Lastname	Firstname	
1	10110095	Diamond	Bergeron	Barb	E-AGENCY
2	10110102	Bronze	Brosseau	Derrick	BANK ONE
3	10110105		Anderson	Lauren	KPMG PEAT MARWICK

< 戻る(B) 次へ(N) > 完了(F) キャンセル

1. 【1行目からカラム名をインポート】オプション
2. カラム名

11. 【完了】をクリックします。

Boston_Customers 物理データオブジェクトが [Tutorial_Objects] フォルダの物理データオブジェクトフォルダの下に表示されます。【概要】ビューにファイルコンテンツが表示されます。ファイルコンテンツはエディタで開いています。



12. 【詳細】ビューをクリックします。

【詳細】ビューには、物理データオブジェクトのプロパティが表示されます。

13. 【詳細】ビューで、[ランタイム: 読み取り] セクションまでスクロールします。

14. 【ランタイム: 読み取り】セクションで、[ソースファイルのディレクトリ] をデータ統合サービスマシンのディレクトリ<Informatica installation directory>\server\Tutorials に設定します。

データ統合サービスは、データ統合サービスを実行するマシンのサーバーディレクトリでソースファイルを検索します。サーバーのインストールには、チュートリアルファイルのコピーが含まれています。ソースファイルおよびディレクトリに対するアクセス権限を変更しなければ、データ統合サービスはクライアントインストールディレクトリからファイルを読み取ることはできません。

以下の図に、ソースファイルディレクトリの例を示します。

Advanced	
Name	Value
Runtime : Read	
Input type	File
Source type	Direct
Source file name	Boston_Customers.csv
Source file directory	\\MyMachine\Informatica\10.0\server\Tutorials

注: Developer tool マシンは、データ統合サービスを実行するマシンのソースファイルディレクトリにアクセスできる必要があります。Developer tool がソースファイルディレクトリにアクセスできない場合、Developer tool はソースファイルのデータをプレビューできなかつたり、ソースファイルのデータにアクセスするマッピングを実行できなかつたりします。複数のデータ統合サービスを実行する場合、データ統合サービス別に個々のソースファイルディレクトリが存在します。

15. 【データビューア】ビューをクリックします。

16. **【データビューア】** ビューで、**【実行】** をクリックします。
データ統合サービスが、Boston_Customers ファイルからデータを読み取り、結果を出力ウィンドウに表示します。
17. **【ファイル】** > **【保存】** をクリックし、Boston_Customers 物理データオブジェクトを保存します。

タスク 2. LA_Customers フラットファイルデータオブジェクトのインポート

このタスクでは、Los Angeles オフィスの顧客データが含まれるフラットファイルから、物理データオブジェクトをインポートします。

1. **【Object Explorer】** ビューで、チュートリアル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】** > **【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。
4. **【既存のフラットファイルから作成】** を選択します。
5. **【参照】** をクリックし、次のディレクトリの LA_Customers.csv に移動します。 <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
6. **【開く】** をクリックします。
ウィザードでデータオブジェクトに LA_Customers という名前が付けられます。
7. **【次へ】** をクリックします。
8. コードページが MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセットであることを確認します。
9. 区切り形式であることを確認します。
10. **【次へ】** をクリックします。
11. 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。
12. **【1 行目からカラム名をインポート】** を選択します。
13. **【完了】** をクリックします。
LA_Customers 物理データオブジェクトが、チュートリアルプロジェクトの **【物理データオブジェクト】** の下に表示されます。
14. **【読み取り】** ビューをクリックし、出力トランスフォーメーションを選択します。
15. **【プロパティ】** ビューで **【ランタイム】** タブをクリックします。
16. **【ソースファイルのディレクトリ】** を Data Integration Service マシンの次のディレクトリに設定します。
<Informatica Installation Directory>\server\Tutorials
17. **【ファイル】** > **【保存】** をクリックします。

タスク 3. All_Customers フラットファイルデータオブジェクトのインポート

このタスクでは、Los Angeles オフィスと Boston オフィスの顧客注文データを組み合わせたフラットファイルから、物理データオブジェクトをインポートします。

1. **[Object Explorer]** ビューで、チュートリアル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] > [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ソース] ダイアログボックスが表示されます。
4. **[既存のフラットファイルから作成]** を選択します。
5. **[参照]** をクリックし、次のディレクトリの All_Customers.csv に移動します。 <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
6. **[開く]** をクリックします。
ウィザードでデータオブジェクトに All_Customers という名前が付けられます。
7. **[次へ]** をクリックします。
8. コードページが MS Windows Latin 1 (ANSI)、Latin 1 のスーパーセットであることを確認します。
9. 区切り形式であることを確認します。
10. **[次へ]** をクリックします。
11. 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。
12. **[1 行目からカラム名をインポート]** を選択します。
13. **[完了]** をクリックします。
All_Customers 物理データオブジェクトが、チュートリアルプロジェクトの **[物理データオブジェクト]** の下に表示されます。
14. **[読み取り]** ビューをクリックし、出力トランスフォーメーションを選択します。
15. **[プロパティ]** ビューで **[ランタイム]** タブをクリックします。
16. **[ソースファイルのディレクトリ]** を Data Integration Service マシンの次のディレクトリに設定します。
<Informatica Installation Directory>\server\Tutorials
17. **[ファイル] > [保存]** をクリックします。

物理データオブジェクトのインポートのまとめ

このレッスンでは、物理データオブジェクトがフラットファイルまたはリレーショナルデータベーステーブルに基づいたデータの表現であることを学習しました。

フラットファイルから物理データオブジェクトを作成しました。Data Integration Service が正しいディレクトリからソースデータを読み取ることができるようにソースファイルディレクトリを設定することもできます。

作成したデータオブジェクトは、データ品質のレッスンでマッピングソースとして使用します。

第 12 章

レッスン 3. データソースでのプロファイルの実行

この章では、以下の項目について説明します。

- [データのプロファイリングの概要, 60 ページ](#)
- [タスク 1. 2つのデータソースでの結合分析の実行, 61 ページ](#)
- [タスク 2. 結合分析結果の表示, 62 ページ](#)
- [タスク 3. データソースでのプロファイルの実行, 62 ページ](#)
- [タスク 4. カラムプロファイル結果の表示, 63 ページ](#)
- [データのプロファイリングのまとめ, 63 ページ](#)

データのプロファイリングの概要

プロファイルは、データセットの内容と構造を記述する一連のメタデータです。

プロファイリングとデータ検出は、多くの場合、プロジェクトで最初に行われるステップです。プロファイルを実行してデータの構造を評価し、想定する情報のタイプがデータカラムに入力されていることを確認することができます。プロファイルによってデータの問題が明らかになった場合は、それらの問題を修正する手順をプロジェクト内で定義できます。例えば、想定された長さを超える値がカラムに含まれていることがプロファイルによって明らかになった場合は、問題の値を削除または修正するようにデータ品質プロセスを設計できます。

選択したカラムのデータ品質を分析するプロファイルは、カラムプロファイルと呼ばれます。

注: Developer tool を使用して、プライマリキー、外部キー、および機能依存性リレーションを検出することや、データカラムの結合条件を分析することもできます。

カラムプロファイルは、データに関する次のファクトを提供します。

- カラムごとの重複しない値と NULL 値の数。数および割合として表されます。
- カラムごとのデータのパターン、およびこ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。
- 結合分析プロファイルでの、2つのデータカラム間の重複の度合い。ベン図および割合値として表示されます。結合分析プロファイルを使用して、カラムの結合条件で発生する可能性のある問題を特定します。

カラムプロファイルは、プロジェクトのあらゆる段階で実行して、データ品質を測定したり、データの変更がプロジェクトの目標になっていることを確認したりすることができます。カラムプロファイルをマッピング

のトランスフォーメーションで実行して、そのトランスフォーメーションがデータに与える影響を示すことができます。

内容

HypoStores は、顧客データにエラー、不整合、および重複した情報がないことを確認する必要があります。HypoStores は、データ品質の目標を達成するためのプロセスを設計する前に、ソースデータファイルの品質を測定し、データを処理する準備ができていることを確認する必要があります。

目標

このレッスンでは、以下の作業を完了します。

- Boston_Customers データソースと LA_Customers データソースで、結合分析を実行します。
- 結合分析の結果を表示して、2 つのオフィスのデータを正常にマージできるかどうかを判断します。
- All_Customers データソースで、カラムプロファイルを実行します。
- カラムプロファイリングの結果を表示して、データに含まれる値とパターンを確認します。

前提条件

このレッスンを開始する前に、次の要件を確認してください。

- このチュートリアルレッスン 1 から 2 までを完了している。

所要時間

- このレッスンの完了には、20 分見ておいてください。

タスク 1.2 つのデータソースでの結合分析の実行

このタスクでは、Boston_Customers および LA_Customers データソースについて結合分析を実行し、結合条件を表示します。

1. チュートリアルフォルダーを選択し、[ファイル] > [新規] > [プロファイル] をクリックします。
2. [エンタープライズ検出プロファイル] を選択します。
3. [次へ] をクリックします。
4. [名前] フィールドに「Tutorial_Profile」と入力します。
5. [完了] をクリックします。
Object Explorer に Tutorial_Profile プロファイルが表示されます。
6. **Boston_Customers** データソースと **LA_Customers** データソースを右側のエディタにドラッグします。
ヒント: 複数のデータオブジェクトを選択するには、Shift キーを押しながら選択します。
7. データオブジェクト名を右クリックし、[結合プロファイル] を選択します。
新しい結合プロファイルウィザードが表示されます。
8. [名前] フィールドに「JoinAnalysis」と入力します。
9. Boston_Customers と LA_Customers がデータオブジェクトとして表示されていることを確認し、[次へ] をクリックします。
10. 両方のデータソースで **CustomerID** カラムが選択されていることを確認します。
ウィザードのペインをスクロールダウンして、両方のデータセットのカラムを表示します。
[次へ] をクリックします。

11. **【追加】** をクリックして、結合条件を追加します。
【結合条件】 ダイアログボックスが表示されます。
12. **【カラム】** セクションで、**【行の追加】** をクリックします。
13. 左側のカラムの最初の行をダブルクリックし、**CustomerID** を選択します。
14. 右側のカラムの最初の行をダブルクリックし、**CustomerID** を選択します。
15. **【OK】** をクリックし、**【完了】** をクリックします。
16. Developer ツールにより、変更内容を保存するように求められたら、**【はい】** をクリックします。
Developer ツールがプロファイルを実行します。

注: プロファイルは閉じないでください。次のタスクでプロファイル結果を表示します。

タスク 2. 結合分析結果の表示

このタスクでは、JoinAnalysis プロファイルの**【結合結果】** ビューに結合分析結果を表示します。

1. エディタの **【JoinAnalysis】** タブをクリックします。
2. **【結合結果】** セクションで、最初の行をクリックします。
【詳細】 セクションに、ベン図と結合分析結果の詳細を表すカラーキーが表示されます。
3. **【行の結合】** カラムで、結合が含まれる行の数としてゼロが表示されていることを確認します。
この値は、CustomerID フィールドに重複がないことを示します。2つのデータソースを正常に結合できます。
4. LA_Customers データオブジェクトの CustomerID 値を表示するには、ベン図内で **LA_Customers** という名前の円をダブルクリックします。

ヒント: ベン図内の円をダブルクリックしてデータ行を表示します。ベン図内で円が交差する場合は、交差をダブルクリックして両方のデータセットに共通のデータ値を表示します。

データビューアに、LA_Customers データオブジェクトに含まれる CustomerID 値が表示されます。

タスク 3. データソースでのプロファイルの実行

このタスクでは、All_Customers データソースでプロファイルを実行して、データの内容と構造を表示します。

1. **【Object Explorer】** ビューで、チュートリアルプロジェクト内のデータオブジェクトを参照します。
2. **All_Customers** データソースを選択します。
3. **【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。
4. **【プロファイル】** を選択します。
5. **【次へ】** をクリックします。
6. **【名前】** フィールドに **All_Customers** と入力します。
7. **【完了】** をクリックします。

エディタに All_Customers プロファイルが開き、プロファイルが実行されます。

タスク 4. カラムプロファイル結果の表示

このタスクでは、All_Customers データオブジェクトのカラムプロファイリング結果を表示し、データに含まれる値とパターンを確認します。

1. **【ウィンドウ】 > 【ビューの表示】 > 【進行状況】** をクリックして、All_Customers プロファイルの進行状況を表示します。
進行状況ビューが開きます。
2. **【進行状況】** ビューで All_Customers プロファイルの実行完了が報告されたら、エディタで **【結果】** ビューをクリックします。
3. **【カラムプロファイリング】** セクションで、**CustomerTier** カラムをクリックします。
【詳細】 セクションに、CustomerTier カラムに含まれているすべての値が表示され、データセットに値が出現する頻度に関する情報が表示されます。
4. **【詳細】** セクションで、**Ruby** をダブルクリックします。
データビューアにより、CustomerTier カラムに値 **Ruby** が含まれるレコードが実行および表示されます。
5. **【カラムプロファイリング】** セクションで、**OrderAmount** カラムをクリックします。
6. **【詳細】** セクションで、**【リストを表示】** をクリックし、**【パターン】** を選択します。
【詳細】 セクションに、OrderAmount カラムで見つかったパターンが表示されます。Pattern カラム内の文字列 **9(5)** は、5 桁の注文金額が含まれるレコードを参照します。文字列 **9(4)** は、4 桁の注文金額が含まれるレコードを参照します。
7. **Pattern** カラムで、文字列 **9(4)** をダブルクリックします。
データビューアにより、OrderAmount カラムに 4 桁の注文金額が含まれるレコードが実行および表示されます。
8. **【詳細】** セクションで、**【リストを表示】** をクリックし、**【統計】** を選択します。
【詳細】 セクションには、平均値、標準偏差、最大および最小長さ、一般的な順に 5 つの値、一般的なでない順に 5 つの値など、OrderAmount カラムの統計が表示されます。

データのプロファイリングのまとめ

このレッスンでは、プロファイルがデータの内容と構造に関する情報を提供することを学習しました。

2 つのデータオブジェクトについて結合分析を実行し、データオブジェクト間の重なり度の合いを表示できることを学習しました。データオブジェクトについてカラムプロファイルを実行し、データオブジェクト内の各カラムに関連する値、パターン、および統計を表示できることも学習しました。

Boston_Customers データオブジェクトのデータを LA_Customers データオブジェクトのデータとマージできるかどうかを判断するために、JoinAnalysis プロファイルを作成しました。このプロファイルの結果を表示し、CustomerID カラム内のすべての値が一意であることと、データオブジェクトを正常にマージできることを確認しました。

All_Customers プロファイルを作成し、All_Customers データオブジェクトについてカラムプロファイルを実行しました。All_Customers データオブジェクト内のカラムの値、パターン、および統計を検出するために、このプロファイルの結果を表示しました。最後に、データビューアを実行して選択した値とパターンが含まれる行を表示し、データの品質を確認できるようになりました。

索引

A

Analyst ツールの設定
概要 [19](#)

D

Developer tool の設定
概要 [48](#)

か

カスタムプロファイルの作成
概要 [29](#)
カラムからの参照テーブルの作成
概要 [42](#)

さ

参照テーブル
概要 [45](#)

し

式ルールの作成
概要 [33](#)

す

スコアカードの作成
概要 [36](#)

て

データオブジェクトの作成
概要 [22](#)
データのプロファイリング
概要 [60](#)
デフォルトプロファイルの作成
概要 [25](#)

ふ

物理データオブジェクトのインポート
概要 [52](#)
プロファイリング
概要 [9](#)