



Informatica® Data Quality
10.4.0

시작 가이드

Informatica Data Quality 시작 가이드

10.4.0

2019년 12월

© 저작권 Informatica LLC 2011, 2020

이 소프트웨어와 설명서는 사용 및 공개에 대한 제한 사항이 포함되어 있는 별도의 사용권 계약에 따라서만 제공됩니다. 본 문서의 어떤 부분도 Informatica LLC의 사전 통지 없이 어떠한 형태나 수단(전자적, 사진 복사, 녹음 등)으로 복제되거나 전송될 수 없습니다.

미국 정부 권한. 미국 정부 고객에게 제공되는 프로그램, 소프트웨어, 데이터베이스, 관련 문서 및 기술 데이터는 해당하는 연방 입수 규정 및 기관별 보안 규정에 따라 "상용 컴퓨터 소프트웨어" 또는 "상용 기술 데이터"입니다. 따라서 사용, 복제, 공개, 수정 및 조정은 해당하는 정부 계약에 규정된 제한 사항 및 라이선스 조건을 따르며, 정부 계약 조건에 의해 적용 가능한 한도 내에서, FAR 52.227-19, 상용 소프트웨어 라이선스에 규정된 추가 권한이 적용됩니다.

Informatica, PowerCenter, PowerExchange 및 Informatica 로고는 미국과 전 세계 여러 관할 국가에서 Informatica LLC의 상표 또는 등록 상표입니다. Informatica 상표의 현재 목록은 <https://www.informatica.com/trademarks.html> 웹에서 확인할 수 있습니다. 다른 회사 및 제품명은 해당 소유자의 상표 또는 등록 상표일 수 있습니다.

이 소프트웨어 및/또는 설명서의 일부에는 타사의 저작권이 적용될 수 있습니다. 필요한 타사 고지 사항은 제품에 포함되어 있습니다.

이 설명서의 정보는 예고 없이 변경될 수 있습니다. 이 문서에서 문제가 발견되는 경우 infa_documentation@informatica.com으로 보고해 주십시오.

Informatica 제품은 제품이 제공될 당시의 계약 조건에 따라 보증됩니다. Informatica는 상품성과 특정 목적에의 적합성에 대한 보증 그리고 비침해에 대한 보증 또는 조건을 포함하여 어떠한 종류의 명시적이거나 묵시적인 보증 없이 이 문서의 정보를 "있는 그대로" 제공합니다.

발행 날짜: 2020-02-07

목차

서문	7
Informatica 리소스	7
Informatica 네트워크	7
Informatica 기술 자료	7
Informatica 설명서	7
Informatica Product Availability Matrix	8
Informatica Velocity	8
Informatica Marketplace	8
Informatica 글로벌 고객 지원 센터	8
장 1: 시작 개요	9
Informatica 도메인 개요	9
피처 가용성	11
Informatica Analyst 소개	12
Informatica Developer 개요	12
Informatica Developer 시작 페이지	13
치트 시트	13
데이터 품질 및 프로파일링	13
자습서 사례	14
자습서 구조	15
Informatica Analyst 자습서	15
Informatica Developer Tool	16
자습서 선행 조건	16
파트 I: Informatica Analyst 시작	17
장 2: 1과. Informatica Analyst 설정	18
Informatica Analyst 설정 개요	18
태스크 1. Informatica Analyst에 로그인	19
작업 2. 프로젝트 작성	19
작업 3. 폴더 작성	19
Informatica Analyst 설정 요약	20
장 3: 2과. 데이터 개체 작성	21
데이터 개체 작성 개요	21
태스크 1. 플랫폼 파일 데이터 개체 생성	22
태스크 2. 데이터 개체 속성 보기	22
데이터 개체 작성 요약	23

장 4: 3과. 기본 프로필 작성.....	24
기본 프로필 작성 개요.....	24
태스크 1. 기본 프로필 작성 및 실행.....	25
태스크 2. 요약 보기에서 프로필 결과 보기.....	25
기본 프로필 작성 요약.....	26
장 5: 4과. 사용자 지정 프로필 생성.....	27
사용자 지정 프로필 작성 개요.....	27
태스크 1. 사용자 지정 프로필 작성.....	28
태스크 2. 프로필 실행.....	28
태스크 3. 프로필 결과 드릴다운.....	29
사용자 지정 프로필 생성 요약.....	29
장 6: 5과. 식 규칙 생성.....	30
식 규칙 작성 개요.....	30
태스크 1. 식 규칙 작성 및 프로필 실행.....	31
태스크 2. 식 규칙 출력 보기.....	31
태스크 3. 식 규칙 편집.....	32
식 규칙 생성 요약.....	32
장 7: 6과. 성과 기록표 생성 및 실행.....	33
성과 기록표 작성 및 실행 개요.....	33
태스크 1. 프로필 결과에서 성과 기록표 작성.....	34
태스크 2. 성과 기록표 실행.....	35
태스크 3. 성과 기록표 보기.....	35
태스크 4. 성과 기록표 편집.....	36
태스크 5. 임계값 구성.....	36
태스크 6. 점수 경향 차트 보기.....	37
성과 기록표 생성 및 실행 요약.....	37
장 8: 7과. 프로필 열에서 참조 테이블 생성.....	38
프로필 열에서 참조 테이블 작성 개요.....	38
태스크 1. 프로필 열에서 참조 테이블 작성.....	39
태스크 2. 참조 테이블 편집.....	39
프로필 열에서 참조 테이블 생성 요약.....	40
장 9: 8과. 참조 테이블 생성.....	41
참조 테이블 작성 개요.....	41
작업 1. 참조 테이블 생성.....	42
참조 테이블 생성 요약.....	42

파트 II: Informatica Developer 시작.....	43
장 10: 1과. Informatica Developer 설정.....	44
Informatica Developer 설정 개요.....	44
작업 1. Informatica Developer 시작.....	45
작업 2. 도메인 추가.....	45
작업 3. 모델 리포지토리 추가.....	46
작업 4. 프로젝트 작성.....	46
작업 5. 폴더 작성.....	46
작업 6. 기본 데이터 통합 서비스 선택.....	47
Informatica Developer 설정 요약.....	47
장 11: 단원 2: 실제 데이터 개체 가져오기.....	48
실제 데이터 개체 가져오기 개요.....	48
태스크 1. Boston_Customers 플랫폼 파일 데이터 개체 가져오기.....	49
작업 2. LA_Customers 플랫폼 파일 데이터 개체 가져오기.....	55
태스크 3. All_Customers 플랫폼 파일 데이터 개체 가져오기.....	55
실제 데이터 개체 가져오기 요약.....	56
장 12: 3과. 소스 데이터에서 프로필 실행.....	57
데이터 프로파일링 개요.....	57
작업 1. 2개의 데이터 소스에서 조인 분석 수행.....	58
작업 2. 조인 분석 결과 보기.....	59
작업 3. 데이터 소스에서 프로필 실행.....	59
작업 4. 열 프로파일링 결과 보기.....	60
데이터 프로파일링 요약.....	60
장 13: 4과. 데이터를 구문 분석하는 중.....	61
데이터 구문 분석 개요.....	61
태스크 1. 대상 데이터 개체 생성.....	62
1단계. LA_Customers_tgt 데이터 개체 생성.....	62
2단계. 읽기 및 쓰기 옵션 구성.....	63
3단계. 데이터 개체에 열 추가.....	63
태스크 2. 데이터 구문 분석용 매핑 생성.....	63
1단계. 매핑 생성.....	64
2단계. 데이터 개체를 매핑에 추가.....	64
3단계. 파서 변환을 매핑에 추가.....	64
4단계. 파서 변환 구성.....	65
태스크 3. 파서 변환에서 프로필 실행.....	65
태스크 4. 매핑 실행.....	66
태스크 5. 매핑 출력 보기.....	66

데이터 구문 분석 요약.	66
--------------------	----

장 14: 5과. 데이터 표준화 67

데이터 표준화 개요.	67
태스크 1. 대상 데이터 개체 생성.	68
1단계. All_Customers_Stdz_tgt 데이터 개체 생성.	68
2단계. 읽기 및 쓰기 옵션 구성.	69
태스크 2. 데이터를 표준화하는 매핑 생성.	69
1단계. 매핑 생성.	69
2단계. 데이터 개체를 매핑에 추가.	70
3단계. 매핑에 표준화 변환 추가.	70
4단계. 표준화 변환 구성.	70
태스크 3. 매핑 실행.	71
태스크 4. 매핑 출력 보기.	71
데이터 표준화 요약.	72

장 15: 6과. 주소 데이터의 유효성 검사 73

주소 데이터 유효성 검사 개요.	73
태스크 1. 대상 데이터 개체 생성.	74
1단계. All_Customers_av_tgt 데이터 개체 생성.	74
2단계. 읽기 및 쓰기 옵션 구성.	75
3단계. 데이터 개체에 포트 추가.	75
태스크 2. 주소의 유효성을 검사하는 매핑 생성.	76
1단계. 매핑 생성.	76
2단계. 데이터 개체를 매핑에 추가.	76
3단계. 매핑에 주소 유효성 검사기 변환을 추가.	76
태스크 3. 주소 유효성 검사기 변환 구성.	77
1단계. 주소 유효성 검사를 위한 기본 국가 설정.	77
2단계. 주소 유효성 검사기 변환 입력 포트 구성.	77
3단계. 주소 유효성 검사기 변환 출력 포트 구성.	78
4단계. 데이터 대상에 사용되지 않은 데이터 소스 포트 연결.	79
태스크 4. 매핑 실행.	80
태스크 5. 매핑 출력 보기.	80
주소 데이터 유효성 검사 요약.	82

부록 A: FAQ 83

Informatica Analyst FAQ.	83
Informatica Developer FAQ.	83

인덱스 85

서문

데이터 품질 시작 가이드에서 Data Quality의 기본 기능을 탐색하고 Informatica Developer 및 Informatica Analyst에서 데이터 품질 태스크를 수행하는 방법에 대해 알아보십시오.

Informatica 리소스

Informatica는 Informatica Network 및 기타 온라인 포털을 통해 다양한 범위의 제품 리소스를 제공합니다. 리소스를 통해 Informatica 제품 및 솔루션을 최대한 활용하고 다른 Informatica 사용자 및 주제별 전문가로부터 배울 수 있습니다.

Informatica 네트워크

Informatica Network는 Informatica 기술 자료, Informatica 글로벌 고객 지원 센터 등 여러 리소스로 연결되는 관문입니다. Informatica Network를 시작하려면 <https://network.informatica.com>을 방문하십시오.

Informatica Network 멤버인 경우 다음 옵션이 가능합니다.

- 기술 자료에서 제품 리소스를 검색할 수 있습니다.
- 제품 사용 가능 여부에 대한 정보를 봅니다.
- 지원 사례를 생성하고 검토할 수 있습니다.
- 거주 지역의 Informatica 사용자 그룹 네트워크를 검색하고 동료와 협업 관계 유지

Informatica 기술 자료

Informatica 기술 자료를 사용하여 사용 방법 문서, 모범 사례, 비디오 자습서, 자주 묻는 질문에 대한 답변 등 제품 리소스를 확인할 수 있습니다.

기술 자료를 검색하려면 <https://search.informatica.com>을 방문하십시오. 기술 자료에 대한 질문, 의견 또는 아이디어가 있는 경우 KB_Feedback@informatica.com을 통해 Informatica 기술 자료 팀에 문의해 주시기 바랍니다.

Informatica 설명서

Informatica 설명서 포털에서 확장된 설명서 라이브러리를 탐색하여 현재 및 최근 제품 릴리스를 확인할 수 있습니다. 설명서 포털을 탐색하려면 <https://docs.informatica.com>을 방문하십시오.

제품 설명서에 대한 질문, 의견 또는 아이디어가 있는 경우 infa_documentation@informatica.com에서 Informatica 설명서 팀에 문의해 주시기 바랍니다.

Informatica Product Availability Matrix

PAM(Product Availability Matrix)은 제품 릴리스에서 지원하는 운영 체제 버전, 데이터베이스 및 데이터 소스 유형과 대상을 나타냅니다.

<https://network.informatica.com/community/informatica-network/product-availability-matrices>에서 Informatica PAM을 찾을 수 있습니다.

Informatica Velocity

Informatica Velocity는 수백 가지 데이터 관리 프로젝트의 실제 경험을 토대로 Informatica 전문 서비스업에서 개발한 팁과 모범 사례 모음입니다. Informatica Velocity는 전 세계의 조직과 협력하여 성공적인 데이터 관리 솔루션을 계획, 개발, 배포 및 유지 관리하는 Informatica 컨설턴트의 포괄적인 지식을 보여줍니다.

Informatica Velocity 리소스는 <http://velocity.informatica.com>에서 확인할 수 있습니다. Informatica Velocity에 대한 질문, 주석 또는 아이디어가 있으시면 Informatica 전문 서비스업(ips@informatica.com)에 문의하십시오.

Informatica Marketplace

Informatica Marketplace는 Informatica 구현을 확대 및 개선하기 위한 솔루션을 찾을 수 있는 포럼입니다. Marketplace에서 Informatica 개발자와 파트너가 제공하는 수백 개의 솔루션을 활용하여 생산성을 향상시키고 프로젝트의 구현에 걸리는 시간을 줄일 수 있습니다. <https://marketplace.informatica.com>에서 Informatica Marketplace를 찾을 수 있습니다.

Informatica 글로벌 고객 지원 센터

전화 또는 Informatica Network를 통해 글로벌 지원 센터에 문의할 수 있습니다.

해당 지역의 Informatica 글로벌 고객 지원 전화 번호는 Informatica 웹 사이트 (<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>)를 방문하여 찾을 수 있습니다.

Informatica Network에서 온라인 지원 리소스를 찾으려면 <https://network.informatica.com>을 방문하고 eSupport 옵션을 선택하십시오.

제 1 장

시작 개요

이 장에 포함된 항목:

- [Informatica 도메인 개요, 9](#)
- [Informatica Analyst 소개, 12](#)
- [Informatica Developer 개요, 12](#)
- [자습서 사례, 14](#)
- [자습서 구조, 15](#)

Informatica 도메인 개요

Informatica는 서비스를 확장하고 여러 시스템 간에 리소스를 공유하는 기능을 제공하는 서비스 중심 아키텍처입니다. Informatica 도메인은 서비스를 관리하기 위한 기본 단위입니다.

Informatica를 설치한 후 Informatica Administrator에 로그인할 수 있습니다. Administrator 도구를 사용하여 도메인을 관리할 수 있으며 필수 응용 프로그램 서비스를 구성하면 나머지 응용 프로그램 클라이언트에 액세스할 수 있습니다.

Informatica 도메인에는 다음과 같은 구성 요소가 포함됩니다.

- 응용 프로그램 클라이언트. 기본 Informatica 기능에 액세스하는 데 사용되는 클라이언트 그룹입니다. 응용 프로그램 클라이언트는 서비스 관리자 또는 응용 프로그램 서비스에 대해 요청합니다.
- 응용 프로그램 서비스. 서버 기반 기능을 나타내는 서비스 그룹입니다. Informatica 도메인에 응용 프로그램 서비스 중 일부가 포함될 수 있습니다. 응용 프로그램 클라이언트가 요구하는 응용 프로그램 서비스를 작성하고 구성합니다.

응용 프로그램 서비스에는 도메인에서 단일 인스턴스를 가질 수 있는 시스템 서비스가 포함됩니다. 도메인을 작성할 때 사용자를 위한 시스템 서비스가 작성됩니다. 시스템 서비스가 제공하는 기능을 사용하도록 해당 서비스를 구성하고 활성화할 수 있습니다.

- 프로필 웨어하우스. 데이터 통합 서비스가 프로필 결과를 저장하기 위해 사용하는 관계형 데이터베이스입니다.
- 참조 데이터 웨어하우스. 모델 리포지토리에서 참조 테이블 개체에 대한 참조 데이터 값을 저장하는 관계형 데이터베이스입니다.
- 리포지토리. 응용 프로그램 클라이언트의 요청을 처리하는 데 필요한 개체 및 프로세스에 대한 메타데이터를 저장하는 관계형 데이터베이스 그룹입니다.
- 서비스 관리자. 모든 도메인 작업을 관리하기 위해 도메인에 기본 제공되는 서비스입니다. 서비스 관리자는 응용 프로그램 서비스를 실행하고 인증, 권한 부여 및 로깅 같은 도메인 기능을 수행합니다.

- 워크플로우 데이터베이스. 워크플로우에 대한 런타임 메타데이터를 저장하는 관계형 데이터베이스입니다.

다음 테이블에는 **Administrator** 도구를 제외한 응용 프로그램 클라이언트, 응용 프로그램 서비스 및 클라이언트에서 필요한 리포지토리가 나열되어 있습니다.

응용 프로그램 클라이언트	응용 프로그램 서비스	리포지토리
Informatica Analyst	<ul style="list-style-type: none"> - 분석 서비스 - 콘텐츠 관리 서비스 - 데이터 통합 서비스 - 모델 리포지토리 서비스 - 검색 서비스 	모델 리포지토리
Informatica Developer	<ul style="list-style-type: none"> - 분석 서비스 - 콘텐츠 관리 서비스 - 데이터 통합 서비스 - 모델 리포지토리 서비스 	모델 리포지토리
Metadata Manager	<ul style="list-style-type: none"> - Metadata Manager 서비스 - PowerCenter 통합 서비스 - PowerCenter 리포지토리 서비스 	<ul style="list-style-type: none"> - Metadata Manager 리포지토리 - PowerCenter 리포지토리
PowerCenter® 클라이언트	<ul style="list-style-type: none"> - PowerCenter 통합 서비스 - PowerCenter 리포지토리 서비스 	PowerCenter 리포지토리
웹 서비스 협 콘솔	<ul style="list-style-type: none"> - PowerCenter 통합 서비스 - PowerCenter 리포지토리 서비스 - 웹 서비스 협 	PowerCenter 리포지토리

다음 응용 프로그램 서비스는 **Informatica** 응용 프로그램 클라이언트에서 액세스하지 않습니다.

- **PowerExchange® Listener** 서비스. 대량 데이터 이동 및 변경 데이터 캡처를 위해 **PowerExchange Listener**를 관리합니다. **PowerCenter** 통합 서비스는 수신기 서비스를 통해 **PowerExchange Listener**에 연결합니다.
- **PowerExchange Logger** 서비스. Linux, UNIX 및 Windows가 변경 데이터를 캡처하고 **PowerExchange Logger** 로그 파일에 쓸 수 있도록 **PowerExchange Logger**를 관리합니다. 변경 데이터의 출처는 DB2 복구 로그, Oracle 다시 실행 로그, Microsoft SQL Server 분산 데이터베이스 또는 i5/OS 시스템이나 z/OS 시스템의 데이터 소스일 수 있습니다.
- **SAP BW** 서비스. SAP BI의 RFC 요청을 수신하고 **PowerCenter** 통합 서비스가 SAP BI에서 추출하거나 SAP BI로 로드하는 워크플로우를 실행하도록 요청합니다.

피처 가용성

Informatica 제품은 응용 프로그램의 공통 집합을 사용합니다. 사용할 수 있는 제품 기능은 사용 중인 제품 라이선스에 따라 달라집니다.

다음 테이블은 라이선스가 허가된 옵션과 각 옵션에 사용할 수 있는 응용 프로그램 피처를 설명합니다.

라이선스가 허가된 옵션	Informatica Developer 피처	Informatica Analyst 피처
데이터 품질	<ul style="list-style-type: none"> - 매핑 생성 및 실행 - 맵렛 및 규칙 생성 및 실행 - 프로필 생성 및 실행(엔터프라이즈 검색, 기본 키 및 외래 키 검색과 함수 종속성 검색을 위한 프로필 포함) - 유추된 프로필 결과 관리 - 성과 기록표 생성 및 실행 - 참조 테이블 관리 - 데이터 소스에서 예외 레코드 식별 - PowerCenter로 개체 내보내기 	<ul style="list-style-type: none"> - 엔터프라이즈 검색을 포함한 프로파일링 - 검색을 사용하여 프로파일링 리포지토리에서 데이터 및 메타데이터의 위치 찾기 - 성과 기록표 생성 및 실행 - 유추된 프로필 결과 관리 - 프로파일링 규칙 생성 및 실행 - 참조 테이블 관리 - 맵렛에 규칙 사양과 컴파일 규칙 사양 생성 - 예외 레코드 검토 및 편집
Data Services	<ul style="list-style-type: none"> - 논리적 데이터 개체 모델 생성 - Data Services 변환을 통해 매핑 생성 및 실행 - SQL 데이터 서비스 생성 - 웹 서비스 생성 - PowerCenter로 개체 내보내기 	<ul style="list-style-type: none"> - 참조 테이블 관리
Data Services 및 프로파일링 옵션	<ul style="list-style-type: none"> - 논리적 데이터 개체 모델 생성 - Data Services 변환을 통해 매핑 생성 및 실행 - SQL 데이터 서비스 생성 - 웹 서비스 생성 - PowerCenter로 개체 내보내기 - Data Services 변환을 통해 규칙 생성 및 실행 - 프로파일링 수행 	<ul style="list-style-type: none"> - 참조 테이블 관리

참고: Data Engineering 라이선스로 Informatica 제품을 사용하는 경우 기능 집합이 달라질 수 있습니다. 예를 들어 Data Engineering 응용 프로그램은 PowerCenter와 통합되지 않습니다. Data Engineering Quality에서 예외 레코드 관리를 수행할 수 없습니다.

Informatica Analyst 소개

Informatica Analyst는 엔터프라이즈 데이터의 분석, 정리, 표준화, 프로파일링 및 성과 기록에 사용할 수 있는 웹 기반 응용 프로그램 클라이언트입니다.

비즈니스 분석가 및 개발자는 보유한 라이선스에 따라 **Analyst** 도구를 사용하여 데이터 중심 공동 작업을 수행할 수 있습니다. 열 및 규칙 프로파일링과 성과 기록을 수행하고 잘못된 레코드 및 중복 레코드를 관리할 수 있습니다. 또한 참조 데이터를 관리하고 데이터 품질 솔루션 개발자에게 데이터를 제공할 수 있습니다.

Informatica Developer 개요

Developer tool은 Data Integration, Data Quality, Data Profiling, Data Services 및 Data Engineering 솔루션을 디자인하고 구현하는 데 사용되는 응용 프로그램입니다.

Developer tool을 사용하여 메타데이터를 가져오고 연결을 작성하고 데이터 개체를 작성할 수 있습니다. 또한 Developer tool을 사용하여 프로필, 매핑 및 워크플로를 작성하고 실행할 수 있습니다.

Informatica Developer의 보기

Developer tool 작업 영역은 편집기와 보기로 구성됩니다. 편집기에서는 매핑 같은 개체를 편집합니다.

Developer tool은 편집기에서 선택된 개체에 기반하여 보기를 표시합니다.

추가적인 보기를 선택하고 보기를 숨길 수 있으며 Developer tool 작업 영역의 다른 위치로 보기를 이동할 수 있습니다.

표시할 보기를 선택하려면 **창 > 보기 표시**를 클릭하십시오.

기본적으로 Developer tool은 다음과 같은 보기를 표시합니다.

연결 탐색기 보기

관계형 데이터베이스에 대한 연결을 표시합니다.

데이터 뷰어 보기

소스 데이터, 프로필 결과를 표시하고 변환의 출력을 미리 봅니다.

Object Explorer 보기

도메인과 도메인의 디자인 타임 개체 및 런타임 개체를 표시합니다. 디자인 타임 개체는 모델 리포지토리의 프로젝트와 폴더에 저장됩니다. 런타임 개체는 런타임 응용 프로그램의 일부로 데이터 통합 서비스에 저장됩니다.

아웃라인 보기

Object Explorer 보기에서 선택된 개체에 종속된 개체를 표시합니다.

진행률 보기

Developer tool에서 작업(예: 매핑 실행)의 진행률을 표시합니다.

속성 보기

편집기에서 선택된 개체의 속성을 표시합니다.

보기 표시 메뉴를 사용하여 다음 보기를 표시할 수도 있습니다.

알림 보기

연결 상태 알림을 표시합니다.

체크 아웃된 개체 보기

체크 아웃한 모든 개체를 표시합니다.

[알림] 보기

휴먼 태스크의 모든 작업이 완료되면 사용자나 그룹에 알리는 옵션을 표시합니다.

개체 종속성 보기

개체를 보거나 수정하거나 삭제할 때 개체 종속성을 표시합니다.

검색 보기

검색 결과를 표시합니다. 또한 검색 옵션 대화 상자를 실행할 수 있습니다.

태그 보기

비즈니스 용도에 따라 모델 리포지토리의 개체를 정의하는 태그를 표시합니다.

Informatica Developer 시작 페이지

Developer tool을 처음으로 열면 시작 페이지가 나타납니다. 시작 페이지에서 Developer tool을 설정하고 시작하는 방법을 익힐 수 있습니다.

시작 페이지에는 다음 옵션이 표시됩니다.

- 개요. 개요 단추를 클릭하면 데이터 품질 및 데이터 서비스 솔루션에 대한 개요가 표시됩니다.
- 첫 번째 단계. 첫 번째 단계 단추를 클릭하면 Developer tool을 설정하고 Informatica Data Quality 및 Informatica Data Services 단원에 액세스하는 방법을 확인할 수 있습니다.
- 자습서. 자습서 단추를 클릭하면 데이터 품질 및 데이터 서비스 솔루션에 대한 자습서 단원이 표시됩니다.
- 웹 리소스. 웹 리소스 단추를 클릭하면 Informatica 기술 자료에 연결됩니다. Informatica 방법 라이브러리에 액세스할 수 있습니다. Informatica 방법 라이브러리에는 Informatica Data Quality, Informatica Data Services 및 기타 Informatica 제품에 대한 문서가 포함되어 있습니다.
- 작업 영역. 작업 영역 단추를 클릭하면 Developer 도구에서 작업을 시작할 수 있습니다.

시작 페이지를 닫은 후 시작 페이지에 액세스하려면 **도움말 > 시작**을 클릭합니다.

치트 시트

Developer 도구에는 온라인 도움말의 일부로 치트 시트가 포함됩니다. 치트 시트는 Developer 도구에서 하나 이상의 태스크를 완료하는 데 도움이 되는 단계별 가이드입니다.

치트 시트를 완료하면 태스크가 완료되고 결과를 볼 수 있습니다. 예를 들어 치트 시트를 완료하여 관계형 데이터 개체를 가져오고 미리 보면 Developer 도구에서 관계형 데이터베이스 테이블을 가져오고 미리 보는 태스크를 완료한 것이 됩니다.

치트 시트에 액세스하려면 **도움말 > 치트 시트**를 클릭합니다.

데이터 품질 및 프로파일링

데이터 콘텐츠 및 구조를 분석하려면 Developer tool에서 데이터 품질 기능을 사용하십시오. 비즈니스 요구 사항을 충족하는 방식으로 데이터를 강화시킬 수 있습니다.

Developer tool을 사용하여 다음 목표를 달성할 수 있는 프로세스를 설계 및 실행하십시오.

- 프로파일 데이터. 프로파일링은 데이터의 콘텐츠와 구조를 나타냅니다. 프로파일링은 데이터의 강점과 단점을 파악하여 프로젝트 계획을 정립하는 데 도움이 될 수 있으므로 어떤 데이터 프로젝트에서도 핵심 단계로 꼽힙니다.

- 데이터 품질 검토를 위해 성과 기록표를 작성하십시오. 성과 기록표는 프로필의 품질 측정을 그래픽으로 표현합니다.
- 데이터 값 표준화. 프로필을 실행할 때 찾는 오류와 불일치를 제거하기 위해 데이터를 표준화하십시오. 구두점, 형식 및 맞춤법에서의 변형을 표준화할 수 있습니다. 예를 들어 도시, 도 및 우편 번호 값이 일관하도록 보장할 수 있습니다.
- 레코드 구문 분석. 레코드 구조를 개선하고 데이터에서 추가적인 정보를 파생하기 위해 데이터 레코드를 구문 분석하십시오. 자유 형식 데이터의 단일 필드를 다양한 정보 유형을 포함한 필드로 분할할 수 있습니다. 또한 레코드에 정보를 추가할 수도 있습니다. 예를 들어 고객 레코드를 개인 또는 비즈니스 고객으로 플래그를 설정할 수 있습니다.
- 우편 주소 유효성 검사. 주소 유효성 검사는 우편 주소 데이터의 정확도와 전달 가능성을 평가 및 향상시킵니다. 주소 유효성 검사는 주소 레코드를 국립 우편 전송 서비스의 참조 데이터와 비교함으로써 주소의 오류를 수정하고 부분 주소를 완성할 수 있게 합니다. 주소 유효성 검사를 통해 메일 전송을 가속화하고 메일 비용을 감소시키는 우편 정보를 추가할 수도 있습니다.
- 중복 레코드 찾기. 중복 레코드 분석은 선택된 데이터 열에서 유사하거나 일치하는 값을 찾기 위해 일련의 레코드를 서로 비교합니다. 필드 값 사이의 높은 일치율을 나타내는 유사성 수준을 설정합니다. 또한 일치 계산에서 각 열에 고정된 상대 가중치도 설정할 수 있습니다. 예를 들어 이름 정보보다 성에 관한 정보가 우선하도록 설정할 수 있습니다.
- 데이터 품질 규칙 생성 및 실행. Informatica는 프로젝트 목표에 맞게 실행하거나 편집할 수 있는 사전 구축 규칙을 제공합니다. Developer tool에서 규칙을 생성할 수 있습니다.
- Informatica 사용자와 공동 작업. 모델 리포지토리에 추가하는 규칙과 참조 데이터 테이블은 개발자 도구와 분석 도구에서 사용자가 사용할 수 있습니다. 사용자는 프로젝트에 대한 공동 작업을 할 수 있고 다른 사용자는 다른 단계의 프로젝트에서 개체 소유권을 보유할 수 있습니다.
- PowerCenter에 매핑 내보내기. 물리적 데이터 통합을 위해 메타데이터를 다시 사용하거나 웹 서비스를 생성하기 위해 PowerCenter에 매핑을 내보낼 수 있습니다.

자습서 사례

HypoStores Corporation은 보스턴에 본사가 있고 여러 주에 걸쳐 상점을 보유한 미국 소매업체입니다. 이 회사는 미국 전역에서 퍼져 있는 여러 상점의 운영 데이터를 본사 데이터 저장소와 정기적으로 통합합니다. 최근 HypoStores Corporation은 LA에 상점을 열었습니다.

본사에는 중앙 ICC 팀이 있는데, 이 팀은 모든 복합 및 BI 응용 프로그램을 위해 일반 데이터 서비스 계층 제공을 책임지는 설계자, 개발자 및 관리자로 구성됩니다. BI 응용 프로그램에는 청구 및 마케팅에 사용되는 마스터 고객 데이터 파일이 포함된 CRM 시스템이 있습니다.

HypoStores Corporation은 보스턴 본사의 데이터와 로스앤젤레스 사업부의 데이터를 통합하기 위해 다음 태스크를 수행해야 합니다.

- 데이터 품질 문제에 대해 보스턴과 로스앤젤레스의 데이터를 검사합니다.
- 로스앤젤레스 데이터의 정보를 구문 분석합니다.
- 보스턴과 로스앤젤레스 전만의 주소 정보를 표준화합니다.
- CRM 목적으로 데이터의 우편 주소 정보 정확도의 유효성을 검사합니다.

자습서 구조

시작 가이드에는 단원과 태스크로 구성되는 자습서가 포함됩니다.

단원

각 단원은 단원에서 완료해야 하는 태스크를 이해하는 데 도움이 되는 개념을 소개합니다. 단원은 전체 사례의 비즈니스 요구 사항을 제공합니다. 단원의 목표에는 비즈니스 요구 사항을 충족하기 위해 완료해야 하는 태스크가 간략히 설명되어 있습니다. 각 단원에는 태스크의 예상 완료 시간이 제공됩니다. 단원의 태스크를 완료한 후에는 단원 요약을 검토할 수 있습니다.

도구 내의 환경이 구성되지 않은 경우 각 자습서의 첫 번째 단원을 바탕으로 환경을 구성하시기 바랍니다.

작업

태스크는 단계별 지침을 제공합니다. 모든 태스크를 나열된 순서로 완료하면 단원이 완료됩니다.

Informatica Analyst 자습서

이 자습서에서는 분석가가 Analyst 도구에 로그인하고, 프로젝트와 폴더를 생성하고, 프로필과 규칙을 생성하고, 데이터를 평가하고 참조 테이블을 생성하는 내용을 다룹니다.

사용자가 수행할 수 있는 강좌는 Informatica Data Quality나 Informatica Data Services 제품을 보유하고 있는지 여부에 달려 있습니다.

다음 테이블은 보유하고 있는 제품에 따라 수행할 수 있는 강좌를 설명합니다.

학습 목표	설명	제품
1과. Informatica Analyst 설정	Analyst 도구에 로그인한 후 자습서 학습 목표를 달성하는 데 필요한 프로젝트 및 폴더를 작성합니다.	데이터 품질 Data Services
2과. 데이터 개체 작성	플랫 파일을 데이터 개체로 가져오고 데이터를 미리 보십시오.	데이터 품질
3과. 빠른 프로필 생성	데이터 품질의 아이디어를 빠르게 얻으려면 빠른 프로필을 생성합니다.	데이터 품질
4과. 사용자 지정 프로필 생성	열, 샘플링 및 드릴다운 옵션을 구성하는 사용자 지정 프로필을 생성합니다.	데이터 품질
5과. 식 규칙 생성	열 값을 수정 및 프로파일링할 식 규칙을 생성합니다.	데이터 품질
6과. 성과 기록표 생성 및 실행	시간의 경과에 따른 데이터 품질의 진행 상황을 측정하려면 성과 기록표를 생성 및 실행합니다.	데이터 품질
7과. 프로필 결과에서 참조 테이블 생성	소스 데이터를 표준화하는 데 사용할 수 있는 참조 테이블을 생성합니다.	데이터 품질 Data Services
8과. 참조 테이블 생성	소스 데이터와 유효한 표준 값 사이의 관계를 설정하는 참조 테이블을 생성합니다.	데이터 품질 Data Services

Informatica Developer Tool

이 자습서에서는 여러 데이터 품질 작업을 수행하는 데 **Developer tool**을 사용합니다.

Informatica Data Quality 사용자는 데이터 품질을 향상시키는 프로세스를 디자인 및 실행하는 데 **Developer tool**을 사용합니다. 또한 **Informatica Data Quality** 사용자는 데이터 콘텐츠와 구조를 분석하는 프로필을 생성 및 실행하는 데 **Developer tool**을 사용합니다.

데이터 품질 자습서에서 다음 강좌를 완료하십시오.

1과. Informatica Developer 설정

도메인에서 모델 리포지토리 서비스를 통해 관리되는 모델 리포지토리에 대한 연결을 생성하십시오. 자습서에서 강좌를 위한 작업을 저장할 프로젝트와 폴더를 생성하십시오. 도메인에 둘 이상의 데이터 통합 서비스가 포함되어 있는 경우 서비스를 선택합니다.

2과. 실제 데이터 개체 가져오기

이러한 개체와 관련된 사용자 지정 데이터 파일에 대한 데이터 품질 프로세스를 정의합니다.

3과. 데이터 프로파일링

프로파일링은 데이터의 콘텐츠와 구조를 나타냅니다.

프로파일링에는 두 개의 데이터 열 사이에 유효한 조인이 가능한지 여부를 결정하는 분석 형태의 조인 분석이 포함됩니다.

4과. 데이터를 구문 분석하는 중

구문 분석은 데이터 레코드를 강화하고 레코드 구조를 개선합니다. 이는 데이터의 유용한 정보를 찾고 현재 데이터의 새로운 정보를 파생시킬 수도 있습니다.

5과. 데이터 표준화

표준화는 프로파일링 중에 발견되는 데이터 오류와 불일치를 제거합니다.

6과. 주소 데이터의 유효성 검사

주소 유효성 검사는 우편 주소의 정확도 및 배달 가능성을 평가하고 주소에서 발견되는 오류와 누락 사항을 수정합니다.

자습서 선행 조건

자습서 강좌를 시작하려면 최소한 하나의 노드가 설정된 상태에서 **Informatica** 도메인이 실행 중이어야 합니다.

설치 프로그램에는 이 강좌를 완료하는 데 사용할 자습서 파일이 포함되어 있습니다. 클라이언트 및 서버 설치 모두에서 전체 파일을 확인할 수 있습니다.

- 개발자 도구 설치 경로의 다음 위치에서 자습서 파일을 찾을 수 있습니다.

<Informatica Installation Directory>\clients\DeveloperClient\Tutorials

- 서비스 설치 경로의 다음 위치에서 자습서 파일을 찾을 수 있습니다.

<Informatica Installation Directory>\server\Tutorials

자습서 강좌에는 다음 파일이 필요합니다.

- All_Customers.csv
- Boston_Customers.csv
- LA_customers.csv

파트 I: Informatica Analyst 시작

이 파트에 포함된 장:

- [1과. Informatica Analyst 설정, 18](#)
- [2과. 데이터 개체 작성, 21](#)
- [3과. 기본 프로필 작성, 24](#)
- [4과. 사용자 지정 프로필 생성, 27](#)
- [5과. 식 규칙 생성, 30](#)
- [6과. 성과 기록표 생성 및 실행, 33](#)
- [7과. 프로필 열에서 참조 테이블 생성, 38](#)
- [8과. 참조 테이블 생성, 41](#)

제 2 장

1과. Informatica Analyst 설정

이 장에 포함된 항목:

- [Informatica Analyst 설정 개요, 18](#)
- [태스크 1. Informatica Analyst에 로그인, 19](#)
- [작업 2. 프로젝트 작성, 19](#)
- [작업 3. 폴더 작성, 19](#)
- [Informatica Analyst 설정 요약, 20](#)

Informatica Analyst 설정 개요

이 자습서의 단원을 시작하기 전에 **Analyst** 도구를 설정해야 합니다. **Analyst** 도구를 설정하려면 **Analyst** 도구에 로그인하고 작업을 저장할 프로젝트와 폴더를 작성합니다.

Informatica 도메인은 **Informatica** 환경을 정의하는 노드 및 서비스의 컬렉션입니다. 도메인의 서비스에는 분석 서비스와 모델 리포지토리 서비스가 포함됩니다. 분석 서비스는 **Analyst** 도구를 실행하고 모델 리포지토리 서비스는 모델 리포지토리를 관리합니다. **Analyst** 도구에서 작업할 때 자산을 작성하면 **Analyst** 도구가 이 자산을 모델 리포지토리에 저장합니다.

Analyst 도구에서 자산을 작성하려면 먼저 프로젝트를 작성해야 합니다. 프로젝트에는 **Analyst** 도구의 자산이 포함됩니다. 프로젝트에는 비즈니스 요구 사항이 동일한 데이터 개체 등의 관련 자산을 저장하는 폴더도 포함될 수 있습니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

- **Analyst** 도구에 로그인합니다.
- **Analyst** 도구에서 작성하는 자산을 저장할 프로젝트를 작성합니다.
- 프로젝트에서 관련 자산을 저장할 폴더를 작성합니다.

선행 조건

이 학습 과정을 시작하기 전에 다음 선행 조건이 충족되었는지 확인하십시오.

- 관리자가 **Administrator** 도구에서 모델 리포지토리 서비스 및 분석 서비스를 구성했습니다.
- **Analyst** 도구의 호스트 이름과 포트 번호를 알고 있습니다.
- 분석 서비스에 액세스할 사용자 이름과 암호가 있습니다. 이 정보는 관리자로부터 얻을 수 있습니다.

소요 시간

이 단원을 완료하는 데 5-10분 걸립니다.

태스크 1. Informatica Analyst에 로그인

자습서를 시작하려면 Analyst 도구에 로그인해야 합니다.

1. Microsoft Internet Explorer 또는 Google Chrome 브라우저를 시작합니다.
2. 주소 필드에 Informatica Analyst의 URL을 입력합니다.
`http[s]://<fully qualified host name>:<port number>/analyst`
3. 도메인에서 LDAP 또는 원시 인증을 사용하는 경우 로그인 페이지에서 사용자 이름과 암호를 입력합니다.
4. 원시를 선택하거나 특정 보안 도메인의 이름을 선택합니다.
Informatica 도메인에서 LDAP 또는 Kerberos 인증을 사용하는 경우 보안 도메인 필드가 나타납니다. 자신의 사용자 계정이 속한 보안 도메인을 알지 못하는 경우 Informatica 도메인 관리자에게 문의하십시오.
5. 로그인을 클릭합니다.
Analyst 도구가 시작 작업 공간에서 열립니다.

작업 2. 프로젝트 작성

이 태스크에서는 Analyst 도구에서 작성하는 자산을 포함할 프로젝트를 작성하고 프로젝트의 폴더를 포함할 자습서 프로젝트를 작성합니다.

1. 관리 헤더에서 **프로젝트**를 클릭합니다.
프로젝트 작업 공간이 표시됩니다.
2. 작업 메뉴에서 **새로 만들기 > 프로젝트**를 클릭합니다.
새 프로젝트 창이 표시됩니다.
3. 사용자의 이름에 "Tutorial_"를 접두사로 추가한 이름을 프로젝트 이름으로 입력합니다.
4. 확인을 클릭합니다.

작업 3. 폴더 작성

이 태스크에서는 관련 자산을 저장할 폴더를 작성합니다. 폴더를 프로젝트에 작성하거나 다른 폴더에 작성할 수 있습니다. 데이터 품질 프로젝트에 관련된 자산을 저장할 Customers라는 이름의 폴더를 작성합니다.

1. 프로젝트 패널에서 자습서 프로젝트를 선택합니다.
2. 작업 메뉴에서 **새로 만들기 > 폴더**를 클릭합니다.
새 폴더 창이 표시됩니다.
3. 폴더 이름으로 Customers를 입력합니다.
4. 확인을 클릭합니다.
폴더가 자습서 프로젝트 아래에 표시됩니다.

Informatica Analyst 설정 요약

이 단원에서는 **Analyst** 도구가 자산을 프로젝트와 폴더에 저장한다는 것을 배웠습니다. 프로젝트 및 폴더는 모델 리포지토리에 포함됩니다. 분석 서비스는 **Analyst** 도구를 실행합니다. 모델 리포지토리 서비스에서 모델 리포지토리를 관리합니다. 분석 서비스와 모델 리포지토리 서비스는 **Informatica** 도메인의 응용 프로그램 서비스입니다.

이 단원에서 **Analyst** 도구에 로그인하여 프로젝트 및 폴더를 작성했습니다.

이제 **Analyst** 도구를 사용하여 이 자습서의 다른 단원을 완료할 수 있습니다.

제 3 장

2과. 데이터 개체 작성

이 장에 포함된 항목:

- [데이터 개체 작성 개요, 21](#)
- [태스크 1. 플랫폼 파일 데이터 개체 생성, 22](#)
- [태스크 2. 데이터 개체 속성 보기, 22](#)
- [데이터 개체 작성 요약, 23](#)

데이터 개체 작성 개요

Analyst 도구에서 데이터 개체는 플랫폼 파일 또는 관계형 데이터베이스 테이블을 기반으로 하는 데이터 표현입니다. 플랫폼 파일 또는 테이블 개체를 작성하고 이러한 플랫폼 파일 또는 관계형 데이터베이스 테이블의 데이터에 대해 프로필을 실행할 수 있습니다. **Analyst** 도구에서 플랫폼 파일 데이터 개체를 작성하는 경우 **Analyst** 도구를 실행하는 시스템의 플랫폼 파일 캐시에 파일을 업로드하거나 플랫폼 파일이 저장된 네트워크 위치를 지정할 수 있습니다.

사례

HypoStores는 로스앤젤레스 고객 데이터를 플랫폼 파일로 유지합니다. **HypoStores**는 데이터를 프로파일링 및 분석하여 데이터 품질 태스크를 수행해야 합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

1. 플랫폼 파일을 플랫폼 파일 캐시 위치에 업로드하고 데이터 개체를 작성합니다.
2. 플랫폼 파일 데이터 개체의 데이터를 미리 봅니다.

선행 조건

이 학습 과정을 시작하기 전에 다음 선행 조건이 충족되었는지 확인하십시오.

- 자습서의 1과를 완료해야 합니다.
- **LA_Customers.csv** 플랫폼 파일이 있습니다. 이 파일은 <설치 루트 디렉터리>\<릴리스 버전>\clients\DeveloperClient\Tutorials 폴더에서 찾을 수 있습니다.

소요 시간

이 작업을 완료하는 데 5-10분 걸립니다.

태스크 1. 플랫폼 파일 데이터 개체 생성

이 태스크에서는 LA_Customers 파일에서 플랫폼 파일 데이터 개체를 생성합니다.

1. Analyst 도구에서 **새로 만들기 > 플랫폼 파일 데이터 개체**를 클릭하십시오.
플랫폼 파일 추가 마법사가 나타납니다.
2. **찾아보기 및 업로드**를 선택하고 **찾아보기**를 클릭하십시오.
3. LA_Customers.csv의 위치를 찾고 **열기**를 클릭하십시오.
4. **다음**을 클릭합니다.
가져오기 유형 선택 패널에 **구분자로 분리됨** 및 **고정 너비** 옵션이 표시됩니다. **구분자로 분리됨** 옵션을 선택하십시오. 기본 옵션은 **구분자로 분리됨**입니다.
5. **다음**을 클릭합니다.
6. 데이터에 사용되는 구분자 및 텍스트 한정자 지정 아래에서 큰 따옴표를 텍스트 한정자로 선택하십시오.
7. 비어 있지 않은 첫 번째 행에서 열 이름을 가져오려면 **가져올 행 지정** 아래에서 **첫 번째 행에서 가져오기**를 선택하십시오.
미리보기 패널은 첫 번째 행에서 열 제목을 표시하도록 업데이트됩니다.
8. **다음**을 클릭합니다.
열 특성 패널에 각 열에 대한 형식, 배열, 정밀도 및 데이터 유형이 표시됩니다.
9. **다음**을 클릭합니다.
이름 필드에는 LA_Customers가 표시됩니다.
10. **Tutorial_ 프로젝트와 Customers** 폴더를 선택하십시오.
11. **마침**을 클릭합니다.
고객 폴더의 폴더 콘텐츠에 데이터 개체가 표시됩니다.

태스크 2. 데이터 개체 속성 보기

이 태스크에서는 LA_Customers 데이터 개체의 속성을 볼 수 있습니다.

1. **열기**를 클릭하여 라이브러리 작업 공간을 엽니다.
2. 라이브러리 작업 공간의 **자산** 패널에서 **데이터 개체**를 클릭합니다.
데이터 개체의 목록이 **데이터 개체** 패널에 표시됩니다.
3. **LA_Customers** 플랫폼 파일을 클릭합니다.
데이터 미리보기 패널이 LA_Customers 데이터 개체에서 검색된 데이터와 함께 표시됩니다.
4. 데이터 미리보기 패널에서 LA_Customers 데이터 개체의 구조 및 콘텐츠를 검토합니다.
플랫폼 파일 데이터 개체의 처음 100개 행이 Analyst 도구에 표시됩니다.
5. **속성**을 클릭합니다.
속성 패널에 데이터 개체의 이름, 유형, 설명 및 위치가 표시됩니다. 데이터 개체의 열 이름과 열 속성도 볼 수 있습니다.

데이터 개체 작성 요약

이 단원에서는 데이터 개체가 플랫폼 파일 또는 관계형 데이터베이스 소스에 기반하는 데이터를 나타낸다는 것을 배웠습니다. 플랫폼 파일 데이터 개체를 작성하고 포함된 데이터를 미리 볼 수 있다는 사실도 배웠습니다.

이 단원에서는 플랫폼 파일을 업로드하여 플랫폼 파일 데이터 개체를 작성한 다음 데이터 개체의 데이터를 미리 보고 데이터 개체의 속성을 확인했습니다.

이 단원에서 작성한 데이터 개체는 단원 3에서 기본 프로필을 작성하고 단원 4에서 사용자 지정 프로필을 작성하는 데 사용됩니다.

제 4 장

3과. 기본 프로필 작성

이 장에 포함된 항목:

- [기본 프로필 작성 개요, 24](#)
- [태스크 1. 기본 프로필 작성 및 실행, 25](#)
- [태스크 2. 요약 보기에서 프로필 결과 보기, 25](#)
- [기본 프로필 작성 요약, 26](#)

기본 프로필 작성 개요

프로필은 데이터의 콘텐츠 및 구조를 바탕으로 데이터 품질을 분석하는 것입니다. 기본 프로필은 기본 옵션을 사용하여 작성하는 프로필입니다. 기본 프로필을 사용하면 프로필의 모든 열과 옵션을 구성하지 않고 프로필 결과를 얻을 수 있습니다.

데이터 품질 프로젝트를 시작할 때 기본 프로필을 작성하고 실행하여 데이터 품질을 분석합니다. 기본 프로필 개체를 작성할 때는 분석할 데이터 개체 및 데이터 개체 열을 선택합니다. 기본 프로필에서는 프로필 열 및 옵션 구성이 생략됩니다. **Analyst** 도구가 플랫폼 파일 데이터 개체의 라이브 플랫폼 파일에 프로파일링을 수행합니다.

사례

HypoStores는 새로 인수한 로스앤젤레스 사무실의 데이터를 데이터 웨어하우스에 통합해야 합니다. 데이터를 데이터 웨어하우스에 통합하려면 먼저 데이터를 정리해야 합니다. 회사의 분석가는 데이터의 품질을 평가한 후에 데이터 정리를 담당하는 개발자에게 평가 정보를 전달해야 합니다. 분석가는 프로필 결과를 빠르게 확인하여 데이터 품질에 대한 기본적인 아이디어를 얻으려고 합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

1. **LA_Customers** 플랫폼 파일 데이터 개체에 대한 기본 프로필을 작성하고 실행합니다.
2. 프로필 결과를 봅니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 본 자습서의 강좌 1과 2를 완료했습니다.

소요 시간

이 단원을 완료하는 데 5-10분 걸립니다.

태스크 1. 기본 프로필 작성 및 실행

이 태스크에서는 데이터 개체의 모든 열에 대한 기본 프로필을 작성하고 기본 샘플링 및 드릴다운 옵션을 사용합니다.

1. 라이브러리 작업 공간의 **자산** 패널에서 데이터 개체를 선택합니다.
2. 데이터 개체를 마우스 오른쪽 단추로 클릭하고 **프로필 작성**을 선택합니다.
새 **프로필** 마법사가 표시됩니다.
3. **일반 속성 지정** 화면에 이름, 설명 및 위치가 기본적으로 채워집니다.
4. **다음**을 클릭합니다.
5. **소스 선택** 화면에 데이터 개체가 기본적으로 나타납니다. **열 선택** 창에서 열을 볼 수 있습니다.
6. **다음**을 클릭합니다.
7. **설정 지정** 화면에서 다음 옵션이 기본적으로 선택됩니다.
 - **열 프로필 실행**
 - **프로필 실행 대상** 창의 **모든 행**
 - **드릴다운** 창의 **라이브**
 - **후속 프로필 실행의 데이터 유형 및 데이터 도메인 유추에서 승인된 데이터 유형 및 데이터 도메인을 제외합니다.**
 - **런타임 환경** 창의 **원시** 옵션.
8. **다음**을 클릭합니다.
9. **규칙 및 필터 지정** 화면에서 **저장 및 실행**을 클릭하여 프로필을 작성 및 실행합니다.
Analyst 도구에서 프로필을 작성하고 실행합니다. 프로필 결과가 요약 보기에 표시됩니다.

태스크 2. 요약 보기에서 프로필 결과 보기

이 태스크에서는 LA_Customers 프로필의 요약 보기를 사용하여 프로필 결과의 빠른 개요를 확인합니다.

1. 라이브러리 > 자산 > **프로필** 창에서 **LA_Customers** 프로필을 클릭합니다.
프로필 결과가 요약 보기에 표시됩니다.
2. 요약 보기에서 **필터 기준** 창의 **열**을 클릭하여 열에 대한 프로필 결과를 봅니다.
3. 가로 막대 차트 위로 포인터를 이동하면 값을 백분율로 볼 수 있습니다.
4. **Null 고유 비고유 %** 섹션에서는 열의 Null 값, 고유한 값 및 고유하지 않은 값을 백분율로 볼 수 있습니다.
5. **패턴** 섹션에서는 열의 여러 패턴을 가로 막대 차트로 볼 수 있습니다. 막대 차트 위로 포인터를 가져가면 열의 유사한 패턴 수와 패턴 문자를 백분율로 볼 수 있습니다.
6. **길이** 섹션에서는 열의 값에 대한 최소 길이와 최대 길이를 볼 수 있습니다.
7. **값** 섹션에서는 열의 최소값 및 최대값을 볼 수 있습니다.
8. **데이터 유형** 섹션에서는 값 위로 포인터를 이동하여 열의 모든 유추된 데이터 유형 및 문서화된 데이터 유형을 볼 수 있습니다.
9. **데이터 도메인** 섹션에서는 값 위로 포인터를 이동하여 열의 모든 유추된 데이터 도메인을 볼 수 있습니다.
10. 이상값 데이터를 보려면 **작업 > 이상값 검색**을 클릭하여 프로필 결과에서 이상값을 검색할 수 있습니다.
11. **패턴 이상값** 또는 **값 빈도 이상값** 필터를 클릭하여 프로필 결과의 이상값을 봅니다.

12. 열 이름을 클릭하여 상세 보기에서 열에 대한 프로필 결과를 봅니다.

기본 프로필 작성 요약

이 단원에서는 프로필에 대한 모든 열 및 행 샘플링 옵션을 구성하지 않고 프로필 결과를 볼 수 있는 기본 프로필에 대해 배웠습니다. 데이터 품질 프로젝트를 시작할 때 기본 프로필을 작성하고 실행하여 데이터의 품질을 분석할 수 있다는 것을 배웠습니다. 또한 **Analyst** 도구가 플랫폼 파일 데이터 개체의 라이브 플랫폼 파일에 프로파일링을 수행한다는 것도 배웠습니다.

이 단원에서는 기본 프로필을 작성하고 프로필 결과를 분석했습니다. **null** 값과 데이터 유형 등 프로필의 열에 대한 추가 정보를 확인했습니다. 또한 열 값과 패턴을 사용하여 데이터 품질 문제를 식별했습니다.

빠른 프로필의 결과를 분석한 후에는 다음 태스크를 완료할 수 있습니다.

- 프로필에서 열을 제외하고 관심 있는 열만 포함하는 사용자 지정 프로필을 작성합니다.
- 식 규칙을 작성하여 가상 열을 작성하고 프로파일링합니다.
- 유효한 열 값을 포함하는 참조 테이블을 작성합니다.

제 5 장

4과. 사용자 지정 프로필 생성

이 장에 포함된 항목:

- [사용자 지정 프로필 작성 개요, 27](#)
- [태스크 1. 사용자 지정 프로필 작성, 28](#)
- [태스크 2. 프로필 실행, 28](#)
- [태스크 3. 프로필 결과 드릴다운, 29](#)
- [사용자 지정 프로필 생성 요약, 29](#)

사용자 지정 프로필 작성 개요

프로필은 데이터의 콘텐츠 및 구조를 바탕으로 데이터 품질을 분석하는 것입니다. 사용자 지정 프로필은 더 빠른 프로파일링을 위해 열, 샘플링 옵션 및 드릴다운 옵션을 구성할 때 작성하는 프로파일입니다. 소스에서 샘플 행을 선택하려면 샘플링 옵션을 구성하십시오. 소스 데이터 또는 준비된 데이터의 데이터 행으로 드릴다운하려면 드릴다운 옵션을 구성하십시오. Hadoop 또는 원시 환경에서 프로필을 실행하도록 선택할 수 있습니다.

데이터 품질 프로젝트를 시작할 때 프로필을 작성하고 실행하여 데이터 품질을 분석할 수 있습니다. 프로필 개체를 작성할 때는 프로필을 실행할 데이터 개체와 데이터 개체 열을 선택합니다.

사례

HypoStores는 새로 인수한 로스앤젤레스 사무실의 데이터를 회사의 데이터 웨어하우스에 통합해야 합니다.

HypoStores는 LA 고객 데이터 파일에 있는 고객 계층 데이터의 품질을 평가하려고 합니다. 회사의 분석가는 데이터의 품질을 평가한 후에 데이터 정리를 담당하는 개발자에게 평가 정보를 전달해야 합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

1. 플랫폼 파일 데이터 개체에 대한 사용자 지정 프로필을 작성하여 null 값을 가진 열을 제외합니다.
2. 프로필을 실행하여 CustomerTier 열의 콘텐츠 및 구조를 분석합니다.
3. 프로필 결과의 행을 드릴다운합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 본 자습서의 단원 1, 2 및 3을 완료했습니다.

소요 시간

이 단원을 완료하는 데 5-10분 걸립니다.

태스크 1. 사용자 지정 프로필 작성

이 태스크에서는 사용자 지정 프로필을 작성합니다. 사용자 지정 프로필을 작성할 때는 프로필을 실행할 데이터 개체와 열을 선택합니다. 또한 샘플링 및 드릴다운 옵션도 구성합니다.

1. **새로 만들기 > 프로필**을 클릭합니다.
새 **프로필** 마법사가 표시됩니다.
2. **단일 소스** 옵션이 기본적으로 선택됩니다. **다음**을 클릭합니다.
3. **일반 속성 지정** 화면에서 다음 옵션을 설정합니다.
 - 이름 필드에 **Profile_LA_Customers**를 입력합니다.
 - 위치 필드에서 **Customers** 폴더를 선택합니다.
4. **다음**을 클릭합니다.
5. **소스 선택** 화면에서 **선택**을 클릭합니다.
데이터 개체 선택 대화 상자가 표시됩니다.
6. 데이터 개체 선택 대화 상자에서 **LA_Customers**를 선택합니다. **확인**을 클릭합니다.
7. **소스 선택** 화면에서 Address2, Address3 및 City2 열을 선택 취소합니다.
8. **다음**을 클릭합니다.
9. **설정 지정** 화면에서 다음 옵션을 설정합니다.
 - **열 프로필 실행**.
 - **프로필 실행 대상** 창에서 **무작위 샘플링** 옵션을 선택합니다.
 - **후속 프로필 실행 시 데이터 유형 및 데이터 도메인 유추에서 승인된 데이터 유형 및 데이터 도메인 제외** 옵션을 선택합니다.
 - **드릴다운** 창에서 **준비됨** 옵션을 선택합니다.
10. **다음**을 클릭합니다.
11. **규칙 및 필터 지정** 화면에서 **저장 및 마침**을 클릭하여 프로필을 작성 및 실행합니다.
Analyst 도구가 프로필을 작성하고 **검색** 작업 공간에 프로필을 표시합니다. 결과를 보려면 프로필을 실행해야 합니다.

태스크 2. 프로필 실행

이 태스크에서는 프로필을 실행하여 데이터 개체를 프로파일링하고 프로필 결과를 표시합니다. Analyst 도구가 플랫폼 데이터 개체의 준비된 플랫폼 파일에 프로파일링을 수행합니다.

1. **검색** 작업 공간에 있는지 확인합니다.
프로필 **Profile_LA_Customers_Custom**이 작업 공간에 표시됩니다.
2. 작업 공간에서 **Profile_LA_Customers_Custom**을 클릭합니다.
3. 프로필을 편집하거나 프로필을 실행하도록 선택할 수 있는 프로필 화면이 표시됩니다. **실행**을 클릭합니다.
4. 프로필 결과가 요약 보기에 표시됩니다.

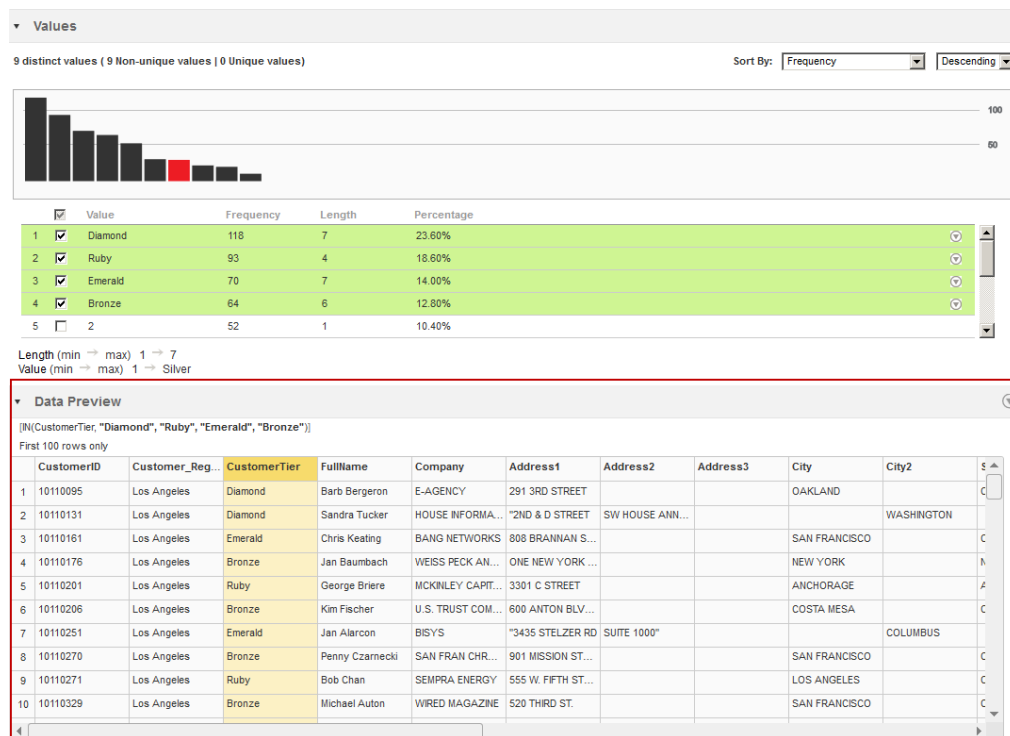
태스크 3. 프로필 결과 드릴다운

이 태스크에서는 CustomerTier 열 값을 드릴다운하여 프로필에 대한 데이터 개체의 소스 행을 확인합니다.

1. **Profile_LA_Customers** 프로필에 대한 프로필 결과의 요약 보기에 있는지 확인합니다.
2. **CustomerTier** 열을 클릭합니다.
열에 대한 프로필 결과가 상세 보기에 표시됩니다.
3. 상세 보기에서 **Diamond**, **Ruby**, **Emerald** 및 **Bronze** 값을 선택합니다. 값 창에서 값을 마우스 오른쪽 단추로 클릭하고 **드릴다운**을 선택합니다.

Diamond, Ruby, Emerald 또는 Bronze 값이 있는 열의 행이 **데이터 미리보기** 창에 표시됩니다.

다음 이미지는 Diamond, Ruby, Emerald 또는 Bronze 값을 드릴다운한 경우 **데이터 미리보기** 창의 드릴다운 결과를 보여 줍니다.



데이터 미리보기 창에는 선택한 열의 처음 10개 행이 표시됩니다. **데이터 미리보기** 창의 제목에는 소스 열에 사용된 논리가 표시됩니다.

사용자 지정 프로필 생성 요약

이 단원에서는 프로파일링할 열을 구성하고 샘플링 및 드릴다운 옵션을 구성할 수 있다는 것을 배웠습니다. 또한 드릴다운을 통해 열 값에 대한 기본 행을 보고 열 값을 볼 때 포함할 열을 구성할 수 있다는 것도 배웠습니다.

이 단원에서는 **CustomerTier** 열을 포함하는 사용자 지정 프로필을 작성하고 프로필을 실행하고 결과에서 **CustomerTier** 열에 대한 기본 행을 드릴다운했습니다.

이 사용자 지정 프로필 개체는 단원 5에서 식 규칙을 작성할 때 사용합니다.

제 6 장

5과. 식 규칙 생성

이 장에 포함된 항목:

- [식 규칙 작성 개요, 30](#)
- [태스크 1. 식 규칙 작성 및 프로필 실행, 31](#)
- [태스크 2. 식 규칙 출력 보기, 31](#)
- [태스크 3. 식 규칙 편집, 32](#)
- [식 규칙 생성 요약, 32](#)

식 규칙 작성 개요

식 규칙은 식 함수와 소스 열을 사용하여 규칙 논리를 정의합니다. **Analyst** 도구에서 식 규칙을 작성하고 프로필에 추가할 수 있습니다. 식 규칙을 하나 이상의 프로필에 연결할 수 있습니다.

식 규칙은 프로필의 가상 열로 출력됩니다. 프로필을 실행하면 **Analyst** 도구가 이 가상 열을 프로파일링합니다.

식 규칙을 사용하여 소스 열의 유효성을 검사하거나 소스 열의 값에 따라 추가 소스 열을 작성할 수 있습니다.

사례

HypoStores는 새로 인수한 로스앤젤레스 사무실의 데이터를 데이터 웨어하우스에 통합해야 합니다.

HypoStores는 고객 이름을 분석하고 고객 이름을 이름과 성으로 구분하려고 합니다. **HypoStores**는 식 규칙을 사용하여 이름과 성을 포함하는 열을 개별 가상 열로 구분 분석한 다음 열을 프로파일링할 것입니다. 또한 이러한 규칙의 출력을 분석해야 하는 다른 분석가들이 규칙을 사용할 수 있도록 할 것입니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

1. **FullName** 열을 이름과 성 열로 구분하는 식 규칙을 작성합니다. 전체 이름에서 이름을 구분하는 규칙을 작성합니다. 전체 이름에서 성을 구분하는 또 다른 규칙을 작성합니다. **Profile_LA_Customers** 프로필에 대해 이러한 규칙을 작성합니다.
2. 프로필을 실행하고 프로필에서 규칙의 출력을 봅니다.
3. 다른 **Analyst** 도구 사용자가 재사용할 수 있도록 규칙을 편집합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 단원 1부터 단원 4까지 완료했습니다.

소요 시간

이 단원을 완료하는 데 10-15분 걸립니다.

태스크 1. 식 규칙 작성 및 프로필 실행

이 태스크에서는 **FullName** 열을 **FirstName** 및 **LastName**의 가상 열 2개로 구문 분석하는 식 규칙 2개를 작성합니다. 규칙 이름은 **FirstName**과 **LastName**입니다.

1. 라이브러리 작업 공간에서 **Profile_LA_Customers** 프로필을 클릭합니다.
프로필 결과가 요약 보기에 나타납니다.
2. 편집을 클릭합니다.
프로필 마법사가 표시됩니다.
3. 규칙 및 필터 지정 화면을 클릭합니다.
4. 규칙 창에서 작업 > 규칙 작성을 클릭합니다.
5. 이름 필드에 **FirstName**을 입력합니다.
6. 식 섹션에서 다음 식을 입력하여 **Name** 열에서 이름을 구분합니다.
`SUBSTR(FullName,1,INSTR(FullName,' ', -1,1) - 1)`
7. 유효성 검사를 클릭합니다.
8. 확인을 클릭합니다.
9. 4~8단계를 반복하여 이름이 **LastName**인 규칙을 작성합니다. 다음 식을 입력하여 **이름** 열에서 성을 분리합니다.
`SUBSTR(FullName,INSTR(FullName,' ', -1,1),LENGTH(FullName))`
10. 저장 및 실행을 클릭하여 프로필을 저장하고 실행합니다.

태스크 2. 식 규칙 출력 보기

이 태스크에서는 프로필을 실행한 후에 이름과 성을 구분하는 식 규칙의 출력을 봅니다.

1. 요약 보기에서 편집을 클릭합니다.
프로필 마법사가 표시됩니다.
2. 프로필 마법사에서 소스 선택을 클릭합니다.
3. 소스 선택 화면에서 도구 모음의 이름 옆에 있는 확인란을 선택하여 모든 열의 선택을 취소합니다.
열 섹션에서 열을 하나 이상 선택해야 하므로 열 하나가 기본적으로 선택되어 있습니다.
4. **FullName** 열을 선택하고 **FirstName** 및 **LastName** 규칙을 선택합니다.
5. 선택된 다른 열은 선택을 취소합니다.
6. 저장 및 실행을 클릭합니다.
프로필 결과가 요약 보기에 표시됩니다.
7. **FirstName** 규칙을 클릭하면 규칙에 대한 프로필 결과가 상세 보기에 표시됩니다.
8. 값 창에서 값을 선택합니다. 값을 마우스 오른쪽 단추로 클릭하고 드릴다운을 클릭합니다.
FullName 열과 **FirstName** 및 **LastName** 규칙에 대한 값이 데이터 미리보기 창에 다른 열 값과 함께 표시됩니다. Analyst 도구에서 **FullName** 열을 이름과 성으로 구분한 것을 볼 수 있습니다.

태스크 3. 식 규칙 편집

이 태스크에서는 모든 **Analyst** 도구 사용자가 사용할 수 있는 재사용 가능한 식 규칙을 만듭니다.

1. **Profile_LA_Customers** 프로파일의 요약 보기에서 **편집**을 클릭합니다.
프로필 마법사가 표시됩니다.
2. **규칙 및 필터 지정** 화면을 클릭합니다.
3. **규칙 및 필터 지정** 화면에서 **FirstName** 규칙을 선택하고 **작업 > 규칙 편집**을 클릭합니다.
규칙 편집 대화 상자가 표시됩니다.
4. **이 규칙을 재사용 가능 규칙으로 저장하시겠습니까?** 옵션을 선택하고 규칙을 저장할 위치를 선택합니다.
5. **확인**을 클릭합니다.
6. **LastName** 규칙을 선택하고 3단계부터 5단계까지 반복합니다.
7. **저장 및 마침**을 클릭하여 프로필을 저장합니다.
이제 모든 **Analyst** 도구 사용자가 **FirstName** 및 **LastName** 규칙을 사용하여 이름과 성이 있는 열을 개별 열로 분리할 수 있습니다.

식 규칙 생성 요약

이 단원에서는 식 함수 및 소스 열을 사용하여 규칙 논리를 정의하는 식 규칙에 대해 배웠습니다. 식 규칙의 출력은 프로파일의 가상 열이라는 것도 배웠습니다. 프로필을 실행하면 **Analyst** 도구에 가상 열이 포함됩니다.

이 단원에서는 식 규칙 2개를 작성하고 프로필에 추가한 다음 프로필을 실행했습니다. 규칙의 출력을 보고 모든 **Analyst** 도구 사용자가 이 규칙을 사용할 수 있도록 편집했습니다.

제 7 장

6과. 성과 기록표 생성 및 실행

이 장에 포함된 항목:

- [성과 기록표 작성 및 실행 개요, 33](#)
- [태스크 1. 프로필 결과에서 성과 기록표 작성, 34](#)
- [태스크 2. 성과 기록표 실행, 35](#)
- [태스크 3. 성과 기록표 보기, 35](#)
- [태스크 4. 성과 기록표 편집, 36](#)
- [태스크 5. 임계값 구성, 36](#)
- [태스크 6. 점수 경향 차트 보기, 37](#)
- [성과 기록표 생성 및 실행 요약, 37](#)

성과 기록표 작성 및 실행 개요

성과 기록표는 프로필 결과의 유효한 열 값 또는 규칙의 출력을 그래픽으로 나타낸 것입니다. 성과 기록표를 사용하여 시간에 따른 데이터 품질 수준을 측정하고 모니터링할 수 있습니다.

성과 기록표를 작성하려면 프로필의 열을 성과 기록표에 메트릭으로 추가하고 메트릭에 가중치를 할당하고 점수 임계값을 구성해야 합니다. 소스 데이터를 기반으로 성과 기록표에 필터를 추가할 수 있습니다. 성과 기록표를 실행하려면 메트릭에 대해 유효한 값을 선택하고 성과 기록표를 실행하여 해당 메트릭에 대한 점수를 봅니다.

성과 기록표에는 프로필의 열에 대한 값 빈도가 점수로 표시됩니다. 점수는 메트릭에 대해 유효한 값의 백분율을 반영합니다.

사례

HypoStores는 새로 인수한 로스앤젤레스 사무실의 데이터를 데이터 웨어하우스에 통합해야 합니다. 이 조직은 데이터를 병합하기 전에 다양한 고객 계층 및 상태의 데이터에서 데이터 품질을 분석했는지 확인하려고 합니다. 회사에서 데이터 품질 분석의 진행 상태를 모니터링해야 하는 분석가는 고객 계층 및 상태 프로필 열에서 성과 기록표를 작성하고 데이터 품질에 대한 임계값을 구성하고 점수 경향 차트를 확인하여 시간에 따른 점수의 개선 정도를 파악하려고 합니다.

학습 목표

이 단원에서는 다음 태스크를 완료합니다.

1. **Profile_LA_Customers_Custom** 프로필의 결과에서 성과 기록표를 작성하여 **CustomerTier** 및 **State** 열에 대한 점수를 봅니다.
2. 성과 기록표를 실행하여 **CustomerTier** 및 **State** 열에 대한 점수를 생성합니다.

3. 성과 기록표를 보고 각 열의 점수를 확인합니다.
4. 성과 기록표를 편집하여 여러 개의 유효한 점수 값을 지정합니다.
5. 점수 임계값을 구성하고 성과 기록표를 실행합니다.
6. 점수 경향 차트를 보고 시간에 따른 점수의 개선 정도를 파악합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 이 자습서의 단원 1부터 단원 5까지 완료했습니다.

소요 시간

이 단원의 태스크를 완료하는 데 15분이 소요됩니다.

태스크 1. 프로필 결과에서 성과 기록표 작성

이 태스크에서는 **Profile_LA_Customers_Custom** 프로필에서 성과 기록표를 작성하여 **CustomerTier** 및 **State** 열 값의 점수를 기록합니다.

1. **라이브러리** 작업 공간에서 **Profile_LA_Customers** 프로필을 클릭합니다.
프로필 결과의 요약 보기가 나타납니다.
2. 요약 보기에서 **CustomerTier** 열을 선택하고 열에서 마우스 오른쪽 단추를 클릭한 다음 **다음에 추가 > 성과 기록표**를 선택합니다.
성과 기록표에 추가 마법사가 나타납니다.
3. 성과 기록표에 추가 마법사에서 새 성과 기록표 옵션은 기본적으로 선택되어 있습니다. **다음**을 클릭합니다.
4. **2/8단계** 화면에서 성과 기록표 이름으로 **sc_LA_Customer**를 입력하고 성과 기록표 위치로는 **Customers** 폴더로 이동합니다.
5. **다음**을 클릭합니다.
6. **3/8단계** 화면에서 성과 기록표에 추가하기 위해 **CustomerTier** 및 **State** 열을 선택합니다.
7. **다음**을 클릭합니다.
8. **4/8단계** 화면에서는 메트릭스에 대한 필터를 작성, 편집 또는 삭제할 수 있습니다. 이 자습서에서는 성과 기록표 필터를 작성하지 않습니다. **다음**을 클릭합니다.
9. **5/8단계** 화면에서 메트릭스 창의 **CustomerTier** 메트릭을 선택합니다.
10. **사용 중인 점수: 값** 창에서 모든 값을 선택하고 **모두 추가** 단추를 클릭하여 값을 **유효한 값** 섹션으로 이동합니다.
여러 값을 선택하려면 **Shift** 키를 사용합니다.
11. 메트릭스 창에서 **State** 메트릭을 선택하고 **사용 중인 점수: 값** 섹션에서 2자리 문자의 상태 코드를 가진 값을 선택합니다.
12. 추가 단추를 클릭하여 값을 **유효한 값** 섹션으로 이동합니다.
유효한 값의 전체 수와 백분율이 화면 위쪽에 표시됩니다.
13. 메트릭스 섹션의 각 메트릭에 대해 **메트릭 임계값** 섹션의 점수 임계값 기본 설정을 허용합니다.
14. **다음**을 클릭합니다.
15. **6/8단계** 화면에서는 필요에 따라 메트릭스를 추가할 메트릭 그룹을 선택할 수 있습니다. 기본적으로 Analyst 도구는 메트릭을 기본 메트릭 그룹에 추가합니다.

16. 다음을 클릭합니다.
17. **7/8단계** 화면에서 **기본 - 메트릭스** 창의 **CustomerTier** 메트릭에 대한 **Weight** 열을 두 번 클릭합니다.
성과 기록표를 실행하면 **Analyst** 도구가 메트릭 점수와 각 메트릭에 할당된 가중치를 바탕으로 각 메트릭 그룹에 대한 가중치 평균을 계산합니다.
18. **CustomerTier** 및 **State** 메트릭스에 대한 가중치를 입력합니다.
19. 다음을 클릭합니다.
20. **8/8단계** 화면에서는 **원시** 옵션이 기본적으로 선택되어 있습니다. **저장**을 클릭하여 성과 기록표를 작성합니다.
성과 기록표가 **성과 기록표** 작업 공간에 표시됩니다.

태스크 2. 성과 기록표 실행

이 태스크에서는 **sc_LA_Customer** 성과 기록표를 실행하여 **CustomerTier** 및 **State** 열에 대한 점수를 생성합니다.

1. **라이브러리** 작업 공간에서 **자산 > 성과 기록표**를 클릭합니다.
2. **성과 기록표** 패널에서 성과 기록표를 클릭합니다.
3. **작업 > 열기**를 클릭합니다.
성과 기록표가 표시됩니다.
4. **작업 > 성과 기록표 실행**을 클릭합니다
성과 기록표 실행 대화 상자가 표시됩니다.
5. 대화 상자의 설정을 확인하고 **실행**을 클릭합니다.
성과 기록표 작업 공간에 **CustomerTier** 및 **State** 열에 대한 점수가 표시됩니다.

태스크 3. 성과 기록표 보기

이 태스크에서는 **sc_LA_Customer** 성과 기록표에서 **CustomerTier** 및 **State** 열에 대한 점수를 봅니다.

1. 보려는 **State** 점수를 포함하는 **State** 행을 선택합니다.
sc_LA_Customer - metrics 섹션에서 성과 기록표의 다음 속성을 볼 수 있습니다.
 - 성과 기록표 이름.
 - 성과 기록표의 전체 행 수.
 - 유효하지 않은 행 수.
 - 가로 막대 차트와 점수.
 - 점수 경향. 점수 경향을 클릭하여 **경향 차트 세부 정보** 화면의 그래픽 표현을 볼 수 있습니다.
 - 메트릭의 가중치.
 - 올바르게 않은 데이터의 비용.
 - 비용 경향.

- 데이터 개체. **검색** 작업 공간에서 데이터 개체의 데이터 미리보기를 보려면 데이터 개체를 클릭합니다.
 - 열 또는 규칙 이름.
 - 소스 유형.
 - 드릴다운 아이콘.
2. 상태 행의 드릴다운 아이콘을 클릭합니다.
상태 열에 대해 유효하지 않은 점수가 **드릴다운** 창의 **올바르지 않은 행 수** 섹션에 표시됩니다.
 3. **올바른 행**을 선택하여 **State** 열의 유효한 점수를 봅니다.
 4. **CustomerTier** 행의 드릴다운 아이콘을 클릭합니다.
CustomerTier 열의 모든 점수가 유효합니다.

태스크 4. 성과 기록표 편집

이 태스크에서는 **sc_LA_Customer** 성과 기록표를 편집하여 **Ruby** 값을 **CustomerTier** 점수의 유효하지 않은 값으로 지정합니다.

1. **성과 기록표** 작업 공간에 있으며 **sc_LA_Customer** 성과 기록표가 열려 있는지 확인합니다.
2. **작업 > 편집 > 메트릭스**를 선택합니다.
성과 기록표 편집 대화 상자가 표시됩니다.
3. **메트릭** 섹션에서 **CustomerTier**를 선택합니다.
4. **사용 중인 점수: 값** 섹션에서 **Ruby**를 **유효한 값** 섹션에서 **사용 가능한 값** 섹션으로 이동합니다.
메트릭 임계값 섹션의 기본 설정을 허용합니다.
5. **저장 및 실행**을 클릭하여 성과 기록표에 대한 변경 사항을 저장하고 실행합니다.
6. **CustomerTier** 점수를 다시 봅니다.
CustomerTier 점수가 81.4%로 변경됩니다.

태스크 5. 임계값 구성

이 태스크에서는 **sc_LA_Customer** 성과 기록표의 **State** 점수에 대한 임계값을 구성하여 **State** 열의 데이터에 대한 허용 가능한 범위를 결정합니다. **CA**와 같이 2자리 문자 코드가 있는 값은 허용 가능하며 **Calif**와 같이 2자리 이상인 문자 코드는 허용되지 않습니다.

1. **성과 기록표** 작업 공간에 있으며 **sc_LA_Customer** 성과 기록표가 열려 있는지 확인합니다.
2. **작업 > 편집 > 메트릭스**를 선택합니다.
성과 기록표 편집 대화 상자가 표시됩니다.
3. **메트릭** 섹션에서 **State**를 선택합니다.
4. **메트릭 임계값** 섹션에서 양호 및 허용되지 않은 점수에 대한 다음 범위를 입력합니다. 90-100% 양호, 0-50% 허용되지 않음, 51-89% 허용.
임계값은 허용 및 양호 범위의 하한을 나타냅니다.
5. **저장 및 실행**을 클릭하여 성과 기록표에 대한 변경 사항을 저장하고 실행합니다.
성과 기록표 패널에서 점수 백분율에 대한 변경 사항 및 막대로 표시된 **State** 점수를 봅니다.

태스크 6. 점수 경향 차트 보기

이 태스크에서는 **State** 점수에 대한 경향 차트를 봅니다. 경향 차트를 확인하여 시간에 따른 점수의 추이를 모니터링할 수 있습니다.

1. **성과 기록표** 작업 공간에 있으며 **sc_LA_Customer** 성과 기록표가 열려 있는지 확인합니다.
2. **상태** 행을 선택합니다.
3. **작업 > 경향 차트 표시**를 클릭하거나 **점수 경향** 열 아래의 화살표를 클릭합니다.

경향 차트 세부 정보 대화 상자가 표시됩니다. 점수에 대한 **양호**, **허용** 및 **허용되지 않음** 임계값이 표시됩니다. 성과 기록표의 점수 값을 편집한 후에 성과 기록표를 실행할 때마다 이 임계값이 변경됩니다.

4. 차트의 원을 가리키면 차트의 아래쪽에 있는 **유효한 값** 섹션에 유효한 값이 표시됩니다.
5. **닫기**를 클릭하여 성과 기록표로 돌아갑니다.

성과 기록표 생성 및 실행 요약

이 단원에서는 프로필 결과에서 성과 기록표를 작성할 수 있다는 것을 배웠습니다. 성과 기록표에는 프로필의 열이 포함됩니다. 성과 기록표를 실행하여 열에 대한 점수를 생성하는 방법을 배웠습니다. 성과 기록표를 편집하여 유효한 값을 구성하고 점수의 임계값을 설정했으며 점수 경향 차트를 보는 방법도 배웠습니다.

이 단원에서는 프로필의 **CustomerTier** 및 **State** 열에서 성과 기록표를 작성하여 고객 계층 및 상태 열의 데이터 품질을 분석했습니다. 성과 기록표를 실행하여 각 열에 대한 점수를 생성했습니다. 성과 기록표를 편집하여 여러 개의 유효한 점수 값을 지정했습니다. 점수에 대한 임계값을 구성하고 점수 경향 차트를 확인했습니다.

제 8 장

7과. 프로필 열에서 참조 테이블 생성

이 장에 포함된 항목:

- [프로필 열에서 참조 테이블 작성 개요, 38](#)
- [태스크 1. 프로필 열에서 참조 테이블 작성, 39](#)
- [태스크 2. 참조 테이블 편집, 39](#)
- [프로필 열에서 참조 테이블 생성 요약, 40](#)

프로필 열에서 참조 테이블 작성 개요

참조 테이블에는 소스 데이터를 표준화하는 데 사용할 수 있는 참조 데이터가 포함됩니다. 참조 데이터에는 유효한 표준 값이 포함됩니다. 참조 테이블을 작성하면 소스 데이터 값과 유효한 표준 값 사이의 관계를 설정할 수 있습니다.

참조 테이블은 프로필 결과에서 작성할 수 있습니다. 참조 테이블을 작성한 후에 참조 테이블을 편집하여 열 또는 행을 추가하고 유효한 표준 값을 추가하거나 편집할 수 있습니다. 참조 테이블의 변경 사항은 감사 내역에서 볼 수 있습니다.

사례

HypoStores는 데이터를 프로파일링하여 잘못된 데이터를 검색하고 해당 데이터를 유효한 값으로 표준화하려고 합니다. 회사의 분석가는 데이터의 유효한 값을 표준화해야 합니다. 이 분석가는 프로필 열의 유효한 값을 바탕으로 참조 테이블을 작성하려고 합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

1. **Profile_LA_Customers_Custom** 프로필의 **CustomerTier** 열에서 열의 유효한 값을 선택하여 참조 테이블을 작성합니다.
2. 참조 테이블을 편집하여 여러 개의 유효한 열 값을 구성합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 이 자습서의 1~6 단원을 완료했습니다.

소요 시간

이 단원의 태스크를 완료하는 데 15분이 소요됩니다.

태스크 1. 프로필 열에서 참조 테이블 작성

이 태스크에서는 참조 테이블을 작성하고 Profile_LA_Customers 프로필의 CustomerTier 열을 참조 테이블에 추가합니다.

1. 라이브러리 작업 공간에서 **자산 > 프로필**을 클릭합니다.
2. **Profile_LA_Customers** 프로필을 클릭하여 요약 보기에서 프로필 결과를 엽니다.
3. 요약 보기에서 참조 테이블에 추가할 CustomerTier 열을 선택합니다. **참조 테이블에 추가**를 마우스 오른쪽 단추로 클릭하고 선택합니다.
참조 테이블에 추가 대화 상자가 표시됩니다.
4. **참조 테이블 작성**을 선택합니다.
5. **다음**을 클릭합니다.
6. **이름** 필드에 **Reftab_CustTier_HypoStores**를 입력합니다.
7. 설명을 입력하고 기본값으로 0을 설정합니다.
Analyst 도구는 값이 없는 테이블 레코드에 기본값을 사용합니다.
8. **다음**을 클릭합니다.
9. **열 특성** 섹션에서 CustomerTier 열에 대한 열 속성을 다음과 같이 구성합니다.

속성	설명
이름	CustomerTier
데이터 유형	문자열
전체 자릿수	10
설명	참조 고객 계층 값

10. 필요한 경우 참조 테이블의 행에 대한 설명 열을 작성합니다. 열의 이름과 전체 자릿수를 입력합니다.
11. **미리보기** 섹션에서 CustomerTier 열 값을 확인합니다.
12. **다음**을 클릭합니다.
Reftab_CustomerTier_HypoStores 참조 테이블 이름이 표시됩니다. 필요한 경우 설명을 입력할 수 있습니다.
13. **저장 위치** 섹션에서 참조 테이블을 작성할 자습서 프로젝트를 선택합니다.
참조 테이블: 패널에서 선택한 위치에 참조 테이블이 나열됩니다.
14. 필요한 경우 감사 참고 내용을 입력합니다.
15. **마침**을 클릭합니다.

태스크 2. 참조 테이블 편집

이 태스크에서는 Reftab_CustomerTier_HypoStores 테이블을 편집하여 고객 계층에 대한 대체 값을 추가합니다.

1. 라이브러리 작업 공간에서 **자산 > 참조 테이블**을 클릭합니다.

2. Reftab_CustomerTier_HypoStores 참조 테이블을 클릭합니다.
참조 테이블이 **디자인** 작업 공간에 열립니다.
3. 행을 편집하려면 행을 선택하고 **작업 > 편집**을 클릭하거나 **편집** 아이콘을 클릭합니다.
행 편집 대화 상자가 표시됩니다. 필요한 경우 여러 행을 선택하여 각 행에 동일한 대체 값을 추가합니다.
4. Diamond, Emerald, Gold, Silver 및 Bronze 행에 대한 다음 대체 값을 입력합니다. 1, 2, 3, 4, 5.
필요한 경우 감사 참고 내용을 입력합니다.
5. **적용**을 클릭하여 변경 내용을 적용합니다.
6. **닫기**를 클릭합니다.
변경된 참조 테이블 값이 **디자인** 작업 공간에 표시됩니다.

프로필 열에서 참조 테이블 생성 요약

이 단원에서는 프로필 결과에서 참조 테이블을 작성하여 소스 데이터의 유효한 값을 구성하는 방법을 배웠습니다.

프로필 열에서 열의 유효한 값을 선택하여 참조 테이블을 작성하고 참조 테이블을 편집하여 여러 개의 유효한 열 값을 구성했습니다.

제 9 장

8과. 참조 테이블 생성

이 장에 포함된 항목:

- [참조 테이블 작성 개요, 41](#)
- [작업 1. 참조 테이블 생성, 42](#)
- [참조 테이블 생성 요약, 42](#)

참조 테이블 작성 개요

참조 테이블에는 소스 데이터를 표준화하는 데 사용할 수 있는 참조 데이터가 포함됩니다. 참조 데이터에는 유효한 표준 값이 포함됩니다. 참조 테이블을 작성하면 소스 데이터 값과 유효한 표준 값 사이의 관계를 설정할 수 있습니다.

참조 테이블 편집기를 사용하면 참조 테이블을 수동으로 작성할 수 있습니다. 참조 테이블은 소스 데이터를 정의하고 표준화하는 데 사용됩니다. **Developer tool**의 표준화 및 조회 변환에서 사용하도록 다른 개발자와 참조 테이블을 공유할 수 있습니다.

사례

HypoStores는 데이터를 유효한 값으로 표준화해야 합니다. 회사의 분석가는 데이터의 유효한 값을 표준화해야 합니다. 이 분석가는 참조 테이블을 작성하여 LA 고객 데이터를 참조하는 표준 고객 계층 코드를 정의하려고 합니다. 그런 다음 이 참조 테이블을 개발자와 공유할 것입니다.

학습 목표

이 단원에서는 다음 태스크를 완료합니다.

- 참조 테이블 편집기를 사용하여 참조 테이블을 작성하고 LA 고객 데이터를 참조하는 표준 고객 계층 코드를 정의합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 본 자습서의 강좌 1과 2를 완료했습니다.

소요 시간

이 단원의 태스크를 완료하는 데 10분이 소요됩니다.

작업 1. 참조 테이블 생성

이 태스크에서는 `Reftab_CustomerTier_Codes` 참조 테이블을 작성하여 고객 계층 데이터의 유효한 값을 표준화합니다.

1. **새로 만들기 > 참조 테이블**을 클릭합니다.
새 **참조 테이블** 마법사가 나타납니다.
2. **참조 테이블 편집기 사용**을 선택합니다.
3. **다음**을 클릭합니다.
4. 참조 테이블에 포함할 각 열에 대해 **새 열 추가** 아이콘을 클릭하고 각 열에 대한 열 속성을 구성합니다.
다음 열 이름을 추가합니다. `CustomerID`, `CustomerTier` 및 `Status`. 열 순서를 다시 지정하거나 열을 삭제할 수 있습니다.
5. 필요한 경우 설명을 입력하고 기본값을 **0**으로 설정합니다.
`Analyst` 도구는 값이 없는 테이블 레코드에 기본값을 사용합니다.
6. **다음**을 클릭합니다.
7. 이름 필드에 **`Reftab_CustomerTier_Codes`**를 입력합니다.
8. **폴더** 섹션에서 자습서 프로젝트의 **Customers** 폴더를 선택합니다.
9. **마침**을 클릭합니다.
참조 테이블이 **디자인** 작업 공간에 표시됩니다.
10. **작업** 메뉴에서 **행 추가**를 선택하여 각 참조 테이블 열을 다음 4개 값으로 채웁니다.
`CustomerID` = LA1, LA2, LA3, LA4
`CustomerTier` = 1, 2, 3, 4.
상태= 활성, 비활성

참조 테이블 생성 요약

이 단원에서는 참조 테이블 편집기를 사용하여 소스 데이터에 사용할 유효한 표준 값을 작성하는 참조 테이블을 작성하는 방법을 배웠습니다.

참조 테이블 편집기에서 LA 고객 데이터에 대한 고객 계층 값을 표준화하는 참조 테이블을 작성했습니다.

파트 II: Informatica Developer 시작

이 파트에 포함된 장:

- [1과. Informatica Developer 설정, 44](#)
- [단원 2: 실제 데이터 개체 가져오기, 48](#)
- [3과. 소스 데이터에서 프로필 실행, 57](#)
- [4과. 데이터를 구문 분석하는 중, 61](#)
- [5과. 데이터 표준화, 67](#)
- [6과. 주소 데이터의 유효성 검사, 73](#)

제 10 장

1과. Informatica Developer 설정

이 장에 포함된 항목:

- [Informatica Developer 설정 개요, 44](#)
- [작업 1. Informatica Developer 시작, 45](#)
- [작업 2. 도메인 추가, 45](#)
- [작업 3. 모델 리포지토리 추가, 46](#)
- [작업 4. 프로젝트 작성, 46](#)
- [작업 5. 폴더 작성, 46](#)
- [작업 6. 기본 데이터 통합 서비스 선택, 47](#)
- [Informatica Developer 설정 요약, 47](#)

Informatica Developer 설정 개요

이 자습서의 단원을 시작하기 전에 **Developer tool**을 시작하고 설정해야 합니다. **Developer tool**을 설정하려면 도메인을 추가합니다. 해당 도메인에 있는 모델 리포지토리를 추가하고 작업을 저장할 프로젝트 및 폴더를 작성합니다. 또한 도메인에 둘 이상의 서비스가 포함된 경우 기본 데이터 통합 서비스를 선택해야 합니다.

Informatica 도메인은 **Informatica** 환경을 정의하는 노드 및 서비스의 컬렉션입니다. 도메인의 서비스에는 모델 리포지토리 서비스와 데이터 통합 서비스가 포함됩니다.

모델 리포지토리 서비스에서 모델 리포지토리를 관리합니다. 모델 리포지토리는 **Developer tool**에서 작성하는 프로젝트의 메타데이터를 저장하는 관계형 데이터베이스입니다. 프로젝트는 사용자가 **Developer tool**에서 작성하는 개체를 저장합니다. 프로젝트에는 비즈니스 요구 사항이 동일한 개체 등의 관련 개체를 저장하는 폴더도 포함될 수 있습니다.

데이터 통합 서비스는 **Developer tool**의 데이터 통합 태스크를 수행합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

- **Developer tool**을 시작하고 **Developer tool** 작업 영역으로 이동합니다.
- **Developer tool**에 도메인을 추가합니다.
- 프로젝트를 작성할 수 있도록 모델 리포지토리를 추가합니다.
- **Developer tool**에서 작성하는 개체를 저장할 프로젝트를 작성합니다.
- 프로젝트에서 관련 개체를 저장할 수 있는 폴더를 작성합니다.

- 데이터 통합 태스크를 수행할 기본 데이터 통합 서비스를 선택합니다.

선행 조건

이 학습 과정을 시작하기 전에 다음 선행 조건이 충족되었는지 확인하십시오.

- Developer tool이 설치되어 있습니다.
- 도메인 이름, 호스트 이름 및 도메인에 연결할 포트 번호가 있습니다. 이 정보는 도메인 관리자로부터 얻을 수 있습니다.
- 도메인 관리자가 Administrator 도구에서 모델 리포지토리 서비스를 구성했습니다.
- 모델 리포지토리 서비스에 액세스하기 위한 사용자 이름과 암호가 있습니다. 이 정보는 도메인 관리자로부터 얻을 수 있습니다.
- 도메인 관리자가 데이터 통합 서비스를 구성했습니다.
- 데이터 통합 서비스가 실행 중입니다.

소요 시간

이 단원의 작업을 완료하는 데 5-10분 걸립니다.

작업 1. Informatica Developer 시작

Developer 도구에서 자습서를 시작합니다.

1. Developer 도구를 시작합니다.
Developer 도구의 **시작** 페이지가 나타납니다.
2. **작업 영역** 단추를 클릭합니다.
Developer 도구 작업 영역이 표시됩니다.

작업 2. 도메인 추가

이 태스크에서는 Developer 도구에 도메인을 추가하여 모델 리포지토리에 액세스합니다.

1. **창 > 기본 설정**을 클릭합니다.
기본 설정 대화 상자가 표시됩니다.
2. **Informatica > 도메인**을 선택합니다.
3. **추가**를 클릭합니다.
새 도메인 대화 상자가 표시됩니다.
4. 도메인 이름, 호스트 이름 및 포트 번호를 입력합니다.
5. **마침**을 클릭합니다.
6. **확인**을 클릭합니다.

작업 3. 모델 리포지토리 추가

이 태스크에서는 프로젝트 및 폴더를 저장하는 데 사용할 모델 리포지토리를 추가합니다.

1. **파일 > 리포지토리에 연결**을 클릭합니다.
리포지토리에 연결 대화 상자가 표시됩니다.
2. **찾아보기**를 클릭하여 모델 리포지토리 서비스를 선택합니다.
3. **확인**을 클릭합니다.
4. **다음**을 클릭합니다.
5. 사용자 이름과 암호를 입력합니다.
6. 네임스페이스를 선택합니다.
7. **마침**을 클릭합니다.
모델 리포지토리가 **개체 탐색기** 보기에 표시됩니다.

작업 4. 프로젝트 작성

이 태스크에서는 **Developer** 도구에서 작성하는 개체를 저장할 프로젝트를 작성합니다. 이 가이드의 모든 자습서를 위한 하나의 프로젝트를 작성할 수 있습니다.

1. **개체 탐색기** 보기에서 모델 리포지토리 서비스를 선택합니다.
2. **파일 > 새로 만들기 > 프로젝트**를 클릭합니다.
새 프로젝트 대화 상자가 나타납니다.
3. 사용자의 이름에 "Tutorial_"를 접두사로 추가한 이름을 프로젝트 이름으로 입력합니다.
4. **마침**을 클릭합니다.
개체 탐색기 보기의 모델 리포지토리 서비스 아래에 프로젝트가 표시됩니다.

작업 5. 폴더 작성

이 태스크에서는 관련 개체를 저장할 폴더를 작성합니다. 이 가이드의 모든 자습서를 위한 하나의 폴더를 작성할 수 있습니다.

1. **개체 탐색기** 보기에서 폴더를 추가할 프로젝트를 선택합니다.
2. **파일 > 새로 만들기 > 폴더**를 클릭합니다.
3. 폴더 이름을 입력합니다.
4. **마침**을 클릭합니다.

Developer 도구가 **개체 탐색기** 보기의 프로젝트 아래에 폴더를 추가합니다. 프로젝트를 확장하면 폴더가 표시됩니다.

작업 6. 기본 데이터 통합 서비스 선택

이 태스크에서는 기본 데이터 통합 서비스를 선택하여 매핑을 실행하고 데이터를 미리 봅니다. 이 단계는 도메인에 둘 이상의 데이터 통합 서비스가 있는 경우에 필요합니다. 도메인에 하나의 데이터 통합 서비스가 포함되어 있으면 해당 데이터 통합 서비스가 기본값으로 설정됩니다.

1. **창 > 기본 설정**을 클릭합니다.
기본 설정 대화 상자가 표시됩니다.
2. **Informatica > 데이터 통합 서비스**를 선택합니다.
3. 도메인을 확장합니다.
4. 데이터 통합 서비스를 선택합니다.
5. **기본값으로 설정**을 클릭합니다.
6. **확인**을 클릭합니다.

Informatica Developer 설정 요약

이 단원에서는 Informatica 도메인에 모델 리포지토리 서비스와 데이터 통합 서비스가 포함된다는 것을 배웠습니다. 모델 리포지토리 서비스에서 모델 리포지토리를 관리합니다. 모델 리포지토리에는 프로젝트와 폴더가 포함됩니다. 데이터 통합 서비스는 데이터 통합 태스크를 수행합니다.

이 단원에서는 Developer 도구를 시작하고 설정했습니다. 도메인을 Developer 도구에 추가하고 모델 리포지토리를 추가하고 프로젝트 및 폴더를 작성했습니다. 또한 기본 데이터 통합 서비스도 선택했습니다.

이제 Developer 도구를 사용하여 이 자습서의 다른 단원을 완료할 수 있습니다.

제 11 장

단원 2: 실제 데이터 개체 가져오기

이 장에 포함된 항목:

- [실제 데이터 개체 가져오기 개요, 48](#)
- [태스크 1. Boston_Customers 플랫폼 파일 데이터 개체 가져오기, 49](#)
- [작업 2. LA_Customers 플랫폼 파일 데이터 개체 가져오기, 55](#)
- [태스크 3. All_Customers 플랫폼 파일 데이터 개체 가져오기, 55](#)
- [실제 데이터 개체 가져오기 요약, 56](#)

실제 데이터 개체 가져오기 개요

실제 데이터 개체는 플랫폼 파일 또는 관계형 데이터베이스 테이블에 기반하는 데이터를 나타냅니다. 플랫폼 파일 또는 관계형 데이터베이스 테이블을 실제 데이터 개체로 가져와서 매핑의 소스 또는 대상으로 사용할 수 있습니다.

사례

HypoStores Corporation은 로스앤젤레스 사무실과 보스턴 사무실의 고객 데이터를 플랫폼 파일에 저장합니다. 이 고객 데이터를 Developer tool에서 작업하려면 각 플랫폼 파일을 실제 데이터 개체로 가져와야 합니다.

학습 목표

이 단원에서는 플랫폼 파일을 실제 데이터 개체로 가져옵니다. 또한 데이터 통합 서비스가 올바른 디렉터리의 소스 데이터를 읽을 수 있도록 소스 파일 디렉터리를 설정합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 자습서의 1과를 완료해야 합니다.

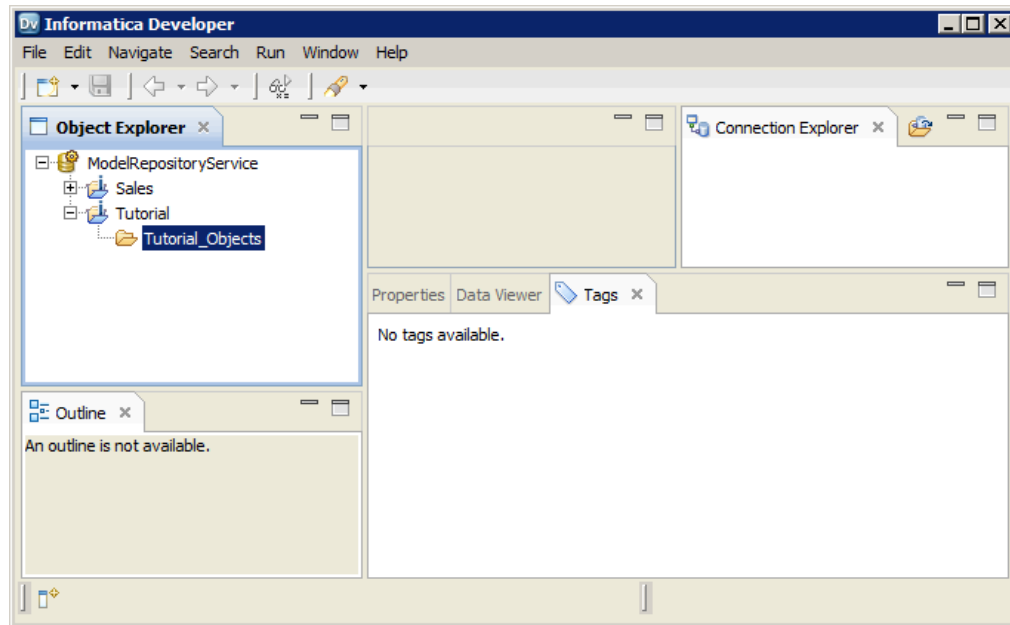
소요 시간

이 단원의 작업을 완료하는 데 10-15분 걸립니다.

태스크 1. Boston_Customers 플랫폼 파일 데이터 개체 가져오기

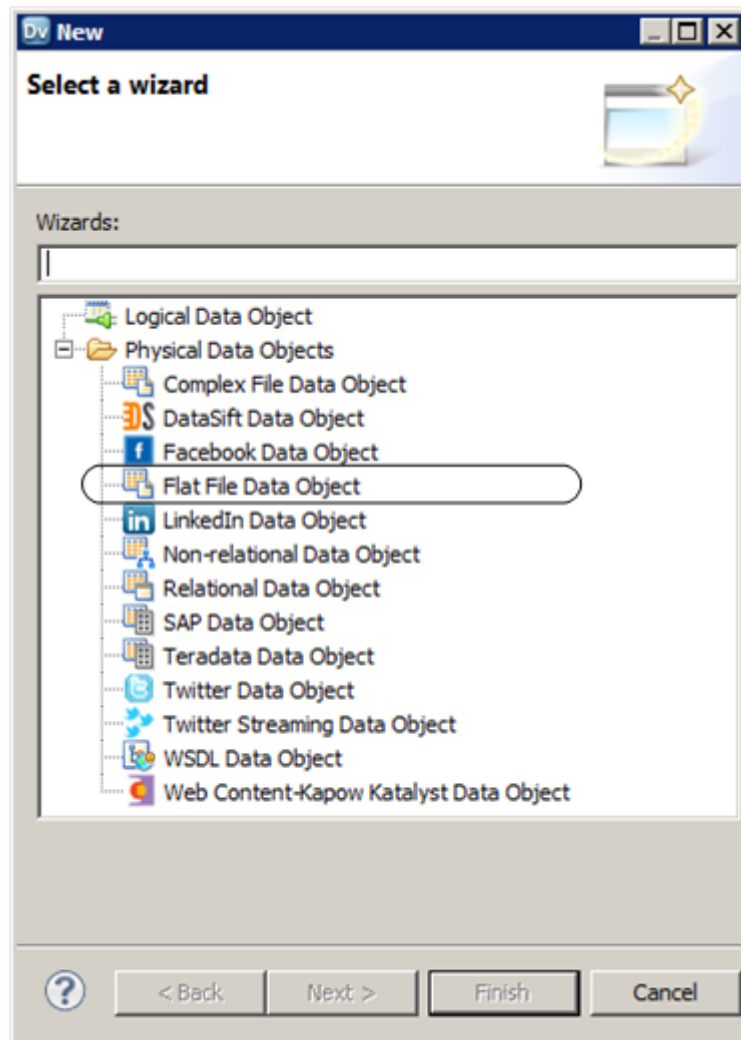
이 태스크에서는 보스턴 사무실의 고객 데이터가 포함된 파일에서 실제 데이터 개체를 가져옵니다.

1. 개체 탐색기 보기에서 Tutorial_Objects 폴더를 선택합니다.



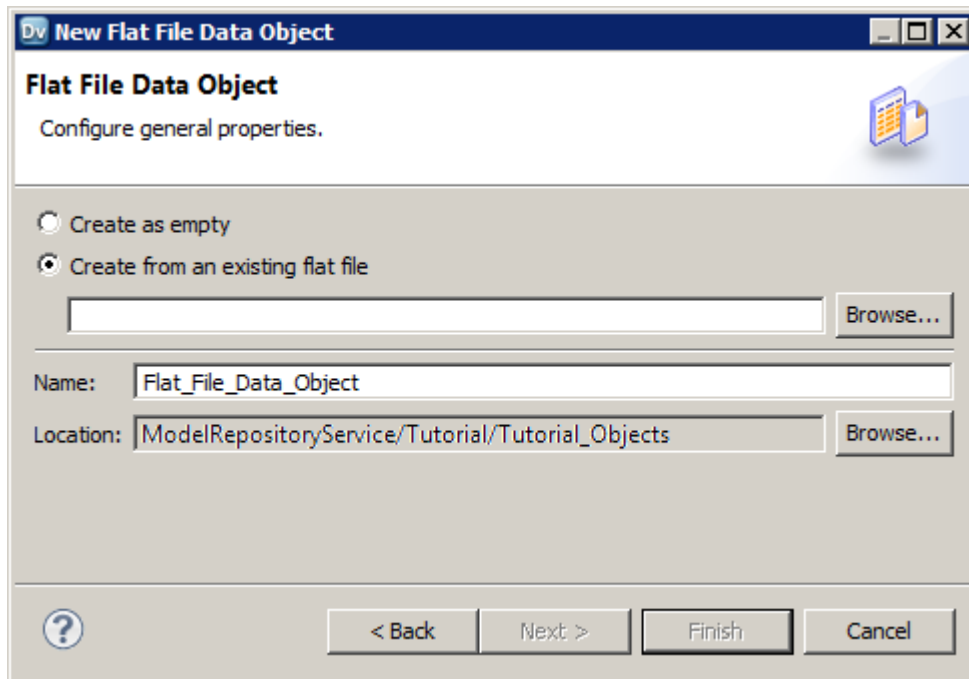
2. Tutorial_Objects 폴더를 마우스 오른쪽 단추로 클릭하고 새로 만들기 > 데이터 개체를 선택합니다.

새로 만들기 대화 상자가 나타납니다.



3. 실제 데이터 개체 > 플랫폼 파일 데이터 개체를 선택하고 다음을 클릭합니다.

새 플랫 파일 데이터 개체 대화 상자가 나타납니다.



4. 기존 플랫 파일에서 작성을 선택합니다.
5. 찾아보기를 클릭하고 Developer tool 시스템의 다음 디렉터리에서 Boston_Customers.csv를 찾습니다.
<Informatica installation directory>\clients\DeveloperClient\Tutorials
6. 열기를 클릭합니다.
마법사가 데이터 개체 이름을 "Boston_Customers"로 지정합니다.
7. 다음을 클릭합니다.
8. 코드 페이지가 **MS Windows Latin 1 (ANSI), superset of Latin 1**으로 설정되고 형식이 **구분자로 분리됨**으로 설정되었는지 확인합니다.

새 플랫 파일 데이터 개체 대화 상자에서 기본 코드 페이지, 형식 및 플랫 파일 데이터의 미리보기가 표시됩니다.

New Flat File Data Object

Flat File Data Object
Configure code page and format.

Code page: MS Windows Latin 1 (ANSI), superset of Latin 1

Format

☒ Delimited (fields separated by delimiters)
☐ Fixed-width (fields aligned in columns)

Maximum rows to preview: 500

	Field1	Field2	Field3	
1	CustomerID	Customer Region	CustomerTier	La
2	10110102	Boston	Bronze	Br
3	10110105	Boston		An
4	10110106	Boston	Emerald	Bc
5	10110107	Boston	Ruby	Ch
6	10110109	Boston	Gold	Ac
7	10110111	Boston	2	Ke
8	10110117	Boston	2	Cc

Navigation buttons: ? < Back Next > Finish Cancel

9. 다음을 클릭합니다.
10. 첫 줄에서 열 이름 가져오기를 선택합니다.

새 플랫폼 파일 데이터 개체 대화 상자의 플랫폼 파일 데이터 미리보기에 열 이름이 표시됩니다.

New Flat File Data Object

Flat File Data Object
Configure delimited format properties.

Delimiters

☐ Tab ☐ Semicolon ☒ Comma

☐ Space ☐ Other: ...

Text qualifier

☒ No quotes ☐ Single quotes ☐ Double quotes

Preview options

☒ Import column names from first line Start import at line: 1

Row delimiter: \012 LF (\n) ☐ Treat consecutive delimiters as one

Escape character: ☐ Retain escape character in data

Maximum rows to preview: 500

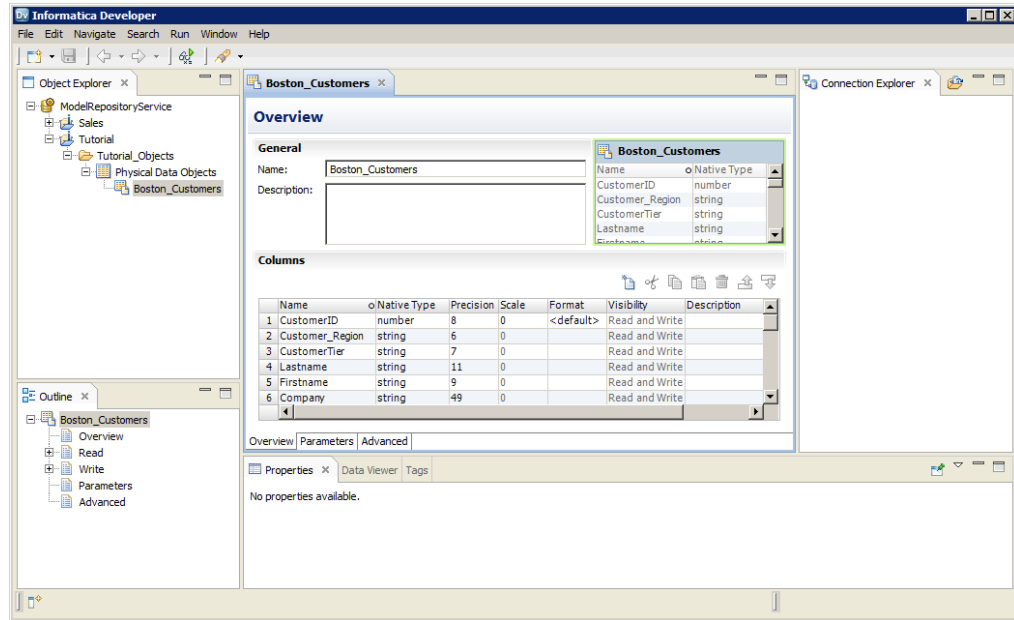
	CustomerID	Customer_Region	CustomerTier
1	10110102	Boston	Bronze
2	10110105	Boston	
3	10110106	Boston	Emerald
4	10110107	Boston	Ruby

? < Back Next > Finish Cancel

1. 첫 줄에서 열 이름 가져오기 옵션
2. 열 이름

11. 마침을 클릭합니다.

Boston_Customers 실제 데이터 개체가 Tutorial_Objects 폴더의 실제 데이터 개체 폴더 아래 나타납니다. 개요 보기가 편집기에서 열리면서 파일 콘텐츠가 표시됩니다.



12. 고급 보기를 클릭합니다.

고급 보기에는 실제 데이터 개체의 속성이 표시됩니다.

13. 고급 보기에서 런타임: 읽기 섹션으로 스크롤합니다.

14. 런타임: 읽기 섹션에서 소스 파일 디렉터리를 데이터 통합 서비스 시스템의 다음 디렉터리로 설정합니다.
<Informatica 설치 디렉터리>\server\Tutorials

데이터 통합 서비스에서 데이터 통합 서비스를 실행하는 시스템의 서버 디렉터리에서 소스 파일을 검색합니다. 서버 설치에는 자습서 파일의 복사본이 포함되어 있습니다. 소스 파일 및 디렉터리에 대한 액세스 권한을 변경하지 않으면 데이터 통합 서비스가 클라이언트 설치 디렉터리에서 파일을 읽을 수 없습니다. 다음 그림은 샘플 소스 파일 디렉터리를 보여 줍니다.

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

참고: Developer tool 시스템은 데이터 통합 서비스를 실행하는 시스템의 소스 파일 디렉터리를 액세스할 수 있어야 합니다. Developer tool이 소스 파일 디렉터리를 액세스할 수 없을 경우 Developer tool은 소스 파일의 데이터를 미리 볼 수 없거나 소스 파일의 데이터에 액세스하는 매핑을 실행할 수 없습니다. 데이터 통합 서비스를 여러 개 실행하는 경우 데이터 통합 서비스마다 별개의 소스 파일 디렉터리가 있습니다.

15. 데이터 뷰어 보기를 클릭합니다.

16. 데이터 뷰어 보기에서 실행을 클릭합니다.

데이터 통합 서비스가 Boston_Customers 파일의 데이터를 읽고 출력 창에 결과를 표시합니다.

17. 파일 > 저장을 클릭하여 Boston_Customers 실제 데이터 개체를 저장합니다.

작업 2. LA_Customers 플랫폼 파일 데이터 개체 가져오기

이 태스크에서는 로스앤젤레스 사무실의 고객 데이터가 포함된 플랫폼 파일에서 실제 데이터 개체를 가져옵니다.

1. **개체 탐색기** 보기에서 자습서 프로젝트를 선택합니다.
2. **파일 > 새로 만들기 > 데이터 개체**를 클릭하십시오.
새로 만들기 대화 상자가 나타납니다.
3. **실제 데이터 개체 > 플랫폼 파일 데이터 개체**를 선택하고 **다음**을 클릭합니다.
새 플랫폼 파일 데이터 개체 대화 상자가 나타납니다.
4. **기존 플랫폼 파일에서 작성**을 선택합니다.
5. **찾아보기**를 클릭하고 다음 디렉터리의 LA_Customers.csv로 이동합니다. <Informatica 설치 디렉터리>\clients\DeveloperClient\Tutorials
6. **열기**를 클릭합니다.
마법사가 데이터 개체 이름을 LA_Customers로 지정합니다.
7. **다음**을 클릭합니다.
8. 코드 페이지가 MS Windows Latin 1 (ANSI), superset of Latin 1인지 확인합니다.
9. 형식이 구분자로 분리되었는지 확인합니다.
10. **다음**을 클릭합니다.
11. 구분자가 쉼표로 설정되어 있는지 확인합니다.
12. **첫 줄에서 열 이름 가져오기**를 선택합니다.
13. **마침**을 클릭합니다.
LA_Customers 실제 데이터 개체가 자습서 프로젝트의 **실제 데이터 개체** 아래에 표시됩니다.
14. **읽기 보기**를 클릭하고 **출력 변환**을 선택합니다.
15. **속성 보기**에서 **런타임** 탭을 클릭합니다.
16. 소스 파일 디렉터리를 데이터 통합 서비스 시스템의 다음 디렉터리로 설정합니다. <Informatica 설치 디렉터리>\server\Tutorials
17. **파일 > 저장**을 클릭합니다.

태스크 3. All_Customers 플랫폼 파일 데이터 개체 가져오기

이 태스크에서는 로스앤젤레스 사무실과 보스턴 사무실의 고객 주문 데이터를 결합한 플랫폼 파일에서 실제 데이터 개체를 가져옵니다.

1. **Object Explorer** 보기에서 자습서 프로젝트를 선택합니다.
2. **파일 > 새로 만들기 > 데이터 개체**를 클릭하십시오.
새로 만들기 대화 상자가 나타납니다.
3. **실제 데이터 개체 > 플랫폼 파일 데이터 개체**를 선택하고 **다음**을 클릭합니다.
새 플랫폼 파일 데이터 소스 대화 상자가 표시됩니다.
4. **기존 플랫폼 파일에서 작성**을 선택합니다.

5. **찾아보기**를 클릭하고 다음 디렉터리의 All_Customers.csv로 이동합니다. <Informatica 설치 디렉터리>\clients\DeveloperClient\Tutorials.
6. **열기**를 클릭합니다.
마법사가 데이터 개체 이름을 All_Customers로 지정합니다.
7. **다음**을 클릭합니다.
8. 코드 페이지가 MS Windows Latin 1 (ANSI), superset of Latin 1인지 확인합니다.
9. 형식이 구분자로 분리되었는지 확인합니다.
10. **다음**을 클릭합니다.
11. 구분자가 쉼표로 설정되어 있는지 확인합니다.
12. **첫 줄에서 열 이름 가져오기**를 선택합니다.
13. **마침**을 클릭합니다.
All_Customers 실제 데이터 개체가 자습서 프로젝트의 **실제 데이터 개체** 아래에 표시됩니다.
14. **읽기 보기**를 클릭하고 **출력 변환**을 선택합니다.
15. **속성 보기**에서 **런타임** 탭을 클릭합니다.
16. 소스 파일 디렉터리를 데이터 통합 서비스 시스템의 다음 디렉터리로 설정합니다. <Informatica 설치 디렉터리>\server\Tutorials
17. **파일 > 저장**을 클릭합니다.

실제 데이터 개체 가져오기 요약

이 단원에서는 실제 데이터 개체가 플랫폼 파일 또는 관계형 데이터베이스 테이블에 기반하는 데이터를 나타내는 것을 배웠습니다.

이 단원에서는 플랫폼 파일에서 실제 데이터 개체를 작성했습니다. 또한 데이터 통합 서비스가 올바른 디렉터리의 소스 데이터를 읽을 수 있도록 소스 파일 디렉터리를 설정합니다.

이 데이터 개체는 데이터 품질 단원에서 매핑 소스로 사용됩니다.

제 12 장

3과. 소스 데이터에서 프로필 실행

이 장에 포함된 항목:

- [데이터 프로파일링 개요, 57](#)
- [작업 1. 2개의 데이터 소스에서 조인 분석 수행, 58](#)
- [작업 2. 조인 분석 결과 보기, 59](#)
- [작업 3. 데이터 소스에서 프로필 실행, 59](#)
- [작업 4. 열 프로파일링 결과 보기, 60](#)
- [데이터 프로파일링 요약, 60](#)

데이터 프로파일링 개요

프로필은 데이터 집합의 콘텐츠와 구조를 설명하는 메타데이터의 집합입니다.

프로젝트의 첫 번째 단계는 주로 프로파일링과 데이터 검색입니다. 프로필을 실행하면 데이터의 구조를 평가하고 데이터 열이 원하는 유형의 정보로 채워져 있는지 확인할 수 있습니다. 프로필을 실행한 후에 데이터 문제가 발견될 경우 이러한 문제를 해결하는 단계를 프로젝트에 정의할 수 있습니다. 예를 들어 프로필을 실행한 후 예상한 길이보다 큰 값이 열에 포함된 것을 알게 된 경우 데이터 품질 프로세스를 디자인하여 문제가 된 값을 제거하거나 수정할 수 있습니다.

선택한 열의 데이터 품질을 분석하는 프로필을 열 프로파일이라고 합니다.

참고: 또한 **Developer tool**을 사용하여 기본 키, 외래 키 및 함수 종속성 관계를 검색하고 데이터 열에 대한 조인 조건을 분석할 수 있습니다.

열 프로파일은 데이터에 대한 다음과 같은 정보를 제공합니다.

- 숫자 및 백분율로 표시되는 각 열의 고유한 값 및 **Null** 값을 제공합니다.
- 각 열에 있는 데이터의 패턴과 이러한 값이 나타나는 빈도를 제공합니다.
- 각 열에 있는 값의 최대 길이 및 최소 길이와 각 열의 첫 번째 값과 마지막 값 등 열 값에 대한 통계를 제공합니다.
- 조인 분석 프로파일의 경우 두 데이터 열 간의 겹침 수준이 벤다이어그램 및 백분율 값으로 표시됩니다. 조인 분석 프로필을 사용하면 열 조인 조건에서 발생할 수 있는 문제를 식별할 수 있습니다.

프로젝트의 원하는 단계에서 열 프로필을 실행하여 데이터 품질을 측정하고 데이터에 대한 변경 사항이 프로젝트 목표에 부합되는지 확인할 수 있습니다. 매핑의 변환에 열 프로필을 실행하면 변환이 데이터에 미치는 영향을 나타낼 수 있습니다.

사례

HypoStores는 고객 데이터에 오류, 불일치 및 중복 정보가 없는지 확인하려고 합니다. HypoStores는 데이터 품질 목표를 달성하는 프로세스를 디자인하기 전에 소스 데이터 파일의 품질을 측정하여 데이터가 처리 가능한 상태인지 확인해야 합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 작업을 완료합니다.

- Boston_Customers 데이터 소스 및 LA_Customers 데이터 소스에 조인 분석을 수행합니다.
- 조인 분석의 결과를 보고 두 사무실의 데이터를 병합할 수 있는지 여부를 결정합니다.
- All_Customers 데이터 소스에 열 프로필을 실행합니다.
- 열 프로파일링 결과를 보고 데이터에 포함된 값 및 패턴을 확인합니다.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 본 자습서의 강좌 1과 2를 완료했습니다.

필요한 시간

- 이 단원을 완료하는 데 20분이 소요됩니다.

작업 1. 2개의 데이터 소스에서 조인 분석 수행

이 태스크에서는 Boston_Customers와 LA_Customers 데이터 소스에 조인 분석을 수행하여 조인 조건을 확인합니다.

1. 자습서 폴더를 선택하고 **파일 > 새로 만들기 > 프로필**을 클릭합니다.
2. **엔터프라이즈 검색 프로필**을 선택합니다.
3. **다음**을 클릭합니다.
4. 이름 필드에 **Tutorial_Profile**을 입력합니다.
5. **마침**을 클릭합니다.
Tutorial_Profile 프로필이 개체 탐색기에 표시됩니다.
6. **Boston_Customers**와 **LA_Customers** 데이터 소스를 오른쪽의 편집기로 끌어옵니다.
팁: Shift 키를 누른 상태에서 선택하면 여러 데이터 개체를 선택할 수 있습니다.
7. 데이터 개체 이름을 마우스 오른쪽 단추로 클릭하고 **조인 프로필**을 선택합니다.
새 조인 프로필 마법사가 표시됩니다.
8. 이름 필드에 **JoinAnalysis**를 입력합니다.
9. Boston_Customers와 LA_Customers가 데이터 개체로 표시되는지 확인하고 **다음**을 클릭합니다.
10. 두 데이터 소스에서 **CustomerID** 열이 선택되어 있는지 확인합니다.
마법사 창을 아래로 스크롤하여 두 데이터 집합의 열을 봅니다.
다음을 클릭합니다.
11. **추가**를 클릭하여 조인 조건을 추가합니다.
조인 조건 대화 상자가 표시됩니다.
12. **열** 섹션에서 **행 추가**를 클릭합니다.

13. 왼쪽 열의 첫 번째 행을 두 번 클릭하고 **CustomerID**를 선택합니다.
14. 오른쪽 열의 첫 번째 행을 두 번 클릭하고 **CustomerID**를 선택합니다.
15. **확인**을 클릭한 후 **마침**을 클릭합니다.
16. 변경 사항을 저장할 것인지 묻는 메시지가 표시되면 **예**를 클릭합니다.
Developer 도구가 프로필을 실행합니다.

참고: 프로필을 닫지 마십시오. 다음 태스크에서 프로필 결과를 볼 것입니다.

작업 2. 조인 분석 결과 보기

이 태스크에서는 JoinAnalysis 프로필의 조인 결과 보기에서 조인 분석 결과를 봅니다.

1. 편집기에서 **JoinAnalysis** 탭을 클릭합니다.
2. **조인 결과** 섹션에서 첫 번째 행을 클릭합니다.
세부 정보 섹션에 조인 분석 결과를 자세히 보여주는 벤다이어그램 및 색상 키가 표시됩니다.
3. **행 조인** 열에 조인을 포함하는 행의 수가 0으로 표시되는지 확인합니다.
이 값은 CustomerID 필드에 중복 항목이 없다는 것을 나타냅니다. 이제 두 개의 데이터 소스를 성공적으로 병합할 수 있습니다.
4. LA_Customers 데이터 개체에 대한 CustomerID 값을 보려면 벤다이어그램에서 **LA_Customers**라는 이름의 원을 두 번 클릭합니다.
팁: 벤다이어그램에서 원을 두 번 클릭하면 데이터 행을 볼 수 있습니다. 벤다이어그램에서 원이 교차하는 경우 교집합을 두 번 클릭하면 두 데이터 집합에 공통되는 데이터 값을 볼 수 있습니다.
데이터 뷰어에 LA_Customers 데이터 개체의 CustomerID 값이 표시됩니다.

작업 3. 데이터 소스에서 프로필 실행

이 태스크에서는 All_Customers 데이터 소스에 프로필을 실행하여 데이터의 콘텐츠와 구조를 봅니다.

1. **개체 탐색기** 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
2. **All_Customers** 데이터 소스를 선택합니다.
3. **파일 > 새로 만들기 > 프로필**을 클릭합니다.
새로 만들기 대화 상자가 나타납니다.
4. **프로필**을 선택합니다.
5. **다음**을 클릭합니다.
6. **이름** 필드에 **All_Customers**를 입력합니다.
7. **마침**을 클릭합니다.
All_Customers 프로필이 편집기에서 열리고 프로필이 실행됩니다.

작업 4. 열 프로파일링 결과 보기

이 태스크에서는 All_Customers 데이터 개체에 대한 열 프로파일링 결과를 보고 데이터에 포함된 값 및 패턴을 검사합니다.

1. **창 > 보기 표시 > 진행률**을 클릭하여 All_Customers 프로파일의 진행률을 봅니다.
진행률 보기가 열립니다.
2. 진행률 보기에 All_Customers 프로파일 실행이 완료된 것으로 표시되면 편집기에서 **결과** 보기를 클릭합니다.
3. **열 프로파일링** 섹션에서 **CustomerTier** 열을 클릭합니다.
세부 정보 섹션에 CustomerTier 열에 포함된 모든 값이 표시되고 데이터 집합에서 해당 값이 발생하는 빈도에 대한 정보가 표시됩니다.
4. **세부 정보** 섹션에서 **Ruby**를 두 번 클릭합니다.
데이터 뷰어가 실행되고 CustomerTier 열에 **Ruby** 값이 포함되는 레코드가 표시됩니다.
5. **열 프로파일링** 섹션에서 **OrderAmount** 열을 클릭합니다.
6. **세부 정보** 섹션에서 **표시** 목록을 클릭하고 **패턴**을 선택합니다.
세부 정보 섹션에 OrderAmount 열에서 발견된 패턴이 표시됩니다. 패턴 열의 문자열 **9(5)**는 5자리 주문 액수를 포함하는 레코드를 나타냅니다. 문자열 **9(4)**는 4자리 주문 액수를 포함하는 레코드를 나타냅니다.
7. **패턴** 열에서 문자열 **9(4)**를 두 번 클릭합니다.
데이터 뷰어가 실행되고 OrderAmount 열에 4자리 주문 액수가 포함되는 레코드가 표시됩니다.
8. **세부 정보** 섹션에서 **표시** 목록을 클릭하고 **통계**를 선택합니다.
세부 정보 섹션에 평균 값, 표준 편차, 최대 및 최소 길이, 가장 많은 5개 값, 가장 적은 5개 값 등 OrderAmount 열에 대한 통계가 표시됩니다.

데이터 프로파일링 요약

이 단원에서는 프로파일을 실행하여 데이터의 콘텐츠 및 구조에 대한 정보를 얻을 수 있다는 것을 배웠습니다.

데이터 개체 2개에 조인 분석을 수행하여 데이터 개체 간의 겹침 수준을 볼 수 있다는 것을 배웠습니다. 또한 데이터 개체에 열 프로파일을 실행하여 데이터 개체의 각 열에 관련된 값, 패턴 및 통계를 볼 수 있다는 것도 배웠습니다.

이 단원에서는 JoinAnalysis 프로파일을 작성하여 Boston_Customers 데이터 개체의 데이터를 LA_Customers 데이터 개체의 데이터와 병합할 수 있는지 여부를 결정했습니다. 이 프로파일의 결과를 보고 CustomerID 열의 모든 값이 고유하므로 데이터 개체를 성공적으로 병합할 수 있다는 것을 확인했습니다.

All_Customers 프로파일을 작성하고 All_Customers 데이터 개체에 열 프로파일을 실행했습니다. 이 프로파일의 결과에서 All_Customers 데이터 개체의 열에 대한 값, 패턴 및 통계를 검색했습니다. 마지막으로 데이터의 품질을 확인하기 위해 데이터 뷰어를 실행하여 선택한 값 및 패턴을 포함하는 행을 표시했습니다.

제 13 장

4과. 데이터를 구문 분석하는 중

이 장에 포함된 항목:

- [데이터 구문 분석 개요, 61](#)
- [태스크 1. 대상 데이터 개체 생성, 62](#)
- [태스크 2. 데이터 구문 분석용 매핑 생성, 63](#)
- [태스크 3. 파서 변환에서 프로필 실행, 65](#)
- [태스크 4. 매핑 실행, 66](#)
- [태스크 5. 매핑 출력 보기, 66](#)
- [데이터 구문 분석 요약, 66](#)

데이터 구문 분석 개요

입력 필드에서 하나 이상의 데이터 요소를 식별하고 서로 다른 출력 필드에 각 요소를 쓰려면 데이터를 구문 분석합니다.

구문 분석을 통해 각 열의 정보를 더 효과적으로 제어할 수 있습니다. 예를 들어 전체 이름이 **Bob Smith**라는 사람의 이름이 포함된 데이터 필드가 있다고 생각해 보십시오. 이 전체 이름을 이름과 성으로 구분된 데이터 열로 분리하는 데 파서 변환을 사용할 수 있습니다. 데이터를 새로운 열로 구문 분석한 후 각 열에 대해 사용자 지정 데이터 품질 작업을 생성할 수 있습니다.

데이터 열을 구성 요소 문자열로 구문 분석하는 데 **토큰 집합**을 사용할 수 있도록 파서 변환을 구성할 수 있습니다. 토큰 집합은 단어, 우편 번호, 전화 번호, 주민등록번호 등의 데이터 요소를 식별합니다.

또한 참조 테이블 항목 또는 사용자가 입력하는 사용자 지정 정규식과 일치하는 데이터를 구문 분석하는 데 파서 변환을 사용할 수 있습니다.

사례

HypoStores는 로스앤젤레스 사무실의 고객 데이터 파일 형식을 보스턴 사무실의 데이터 파일 형식과 일치시키 고자 합니다. 로스앤젤레스 사무실의 고객 데이터는 고객 이름을 **FullName** 열에 저장하는 반면 보스턴 사무실의 고객 데이터는 고객 이름을 **FirstName** 및 **LastName** 열로 구분하여 저장합니다. HypoStores는 로스앤젤레스의 데이터 형식이 보스턴의 데이터 형식과 일치하도록 만들기 위해 로스앤젤레스의 **FullName** 열 데이터를 이름과 성으로 구문 분석해야 합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 태스크를 완료합니다.

- 구문 분석을 거친 데이터를 포함한 **LA_Customers_tgt** 데이터 개체를 생성하고 구성합니다.

- FullName 열을 FirstName 및 LastName 열로 구분하여 구문 분석하는 매핑을 생성합니다.
- 소스 데이터를 연결하려면 LA_Customers 데이터 개체를 매핑에 추가합니다.
- 대상 데이터 개체를 생성하려면 LA_Customers_tgt 데이터 개체를 매핑에 추가합니다.
- 전체 이름을 이름과 성으로 구문 분석하려면 파서 변환을 매핑에 추가하고 토큰 집합을 사용하도록 이를 구성합니다.
- 대상 데이터 소스를 생성하기 전에 데이터를 검토하려면 파서 변환에서 프로필을 실행합니다.
- 구문 분석된 이름을 생성하는 매핑을 실행합니다.
- 매핑 출력을 보려면 데이터 뷰어를 실행하십시오.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 본 자습서의 강좌 1과 2를 완료했습니다.

소요 시간

이 단원의 태스크를 완료하는 데 20분 걸립니다.

태스크 1. 대상 데이터 개체 생성

이 태스크에서는 구문 분석된 이름을 쓸 수 있는 LA_Customers_tgt 데이터 개체를 생성합니다.

대상 데이터 개체를 생성하려면 다음 단계를 수행하십시오.

1. LA_Customers.csv 파일을 바탕으로 LA_Customers_tgt 데이터 개체를 생성하십시오.
2. 파일 위치 및 파일 이름을 포함하여 데이터 개체에 대한 읽기 및 쓰기 옵션을 구성하십시오.
3. LA_Customers_tgt 데이터 개체에 Firstname 및 Lastname 열을 추가하십시오.

1단계. LA_Customers_tgt 데이터 개체 생성

이 단계에서는 LA_Customers.csv 파일을 바탕으로 LA_Customers_tgt 데이터 개체를 생성합니다.

1. **파일 > 새로 만들기 > 데이터 개체**를 클릭하십시오.
 새로 만들기 창이 열립니다.
2. **플랫 파일 데이터 개체**를 선택하고 **다음**을 클릭하십시오.
3. **기존 플랫 파일에서 생성**이 선택되었는지 확인하십시오.
4. 다음 디렉터리에서 **찾아보기**를 클릭하고 LA_Customers.csv로 이동하십시오. <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
5. **열기**를 클릭합니다.
6. 이름 필드에 LA_Customers_tgt를 입력하십시오.
7. **다음**을 클릭합니다.
8. **다음**을 클릭합니다.
9. **미리보기 옵션** 섹션에서 **첫 번째 행에서 열 이름 가져오기**를 선택하고 **다음**을 클릭하십시오.
10. **마침**을 클릭합니다.
 편집기에 LA_Customers_tgt 데이터 개체가 표시됩니다.

2단계. 읽기 및 쓰기 옵션 구성

이 단계에서는 파일 위치와 파일 이름을 포함해 LA_Customers_tgt 데이터 개체에 대한 읽기 및 쓰기 옵션을 구성합니다.

1. 편집기에서 LA_Customers_tgt 데이터 개체가 열려 있는지 확인하십시오.
2. 편집기에서 **읽기** 보기를 선택하십시오.
3. **창 > 보기 표시 > 속성**을 클릭하십시오.
4. **속성** 보기에서 **런타임** 보기를 선택하십시오.
5. **값** 열에서 소스 파일 이름을 두 번 클릭하고 LA_Customers_tgt.csv를 입력하십시오.
6. **값** 열에서 소스 파일 디렉터리를 강조 표시하려면 두 번 클릭하십시오.
7. 강조 표시된 이름을 마우스 오른쪽 단추로 클릭하고 **복사**를 선택하십시오.
8. 편집기에서 **쓰기** 보기를 선택하십시오.
9. **속성** 보기에서 **런타임** 보기를 선택하십시오.
10. **값** 열에서 **출력 파일 디렉터리** 항목을 두 번 클릭하십시오.
11. **읽기** 보기에서 복사한 디렉터리 위치를 붙여넣기 하려면 마우스 오른쪽 단추로 클릭하고 **붙여넣기**를 선택하십시오.
12. **값** 열에서 **헤더 옵션** 항목을 두 번 클릭하고 출력 필드 이름을 선택하십시오.
13. **값** 열에서 **출력 파일 이름** 항목을 두 번 클릭하고 LA_Customers_tgt.csv를 입력하십시오.
14. 데이터 개체를 저장하려면 **파일 > 저장**을 클릭하십시오.

3단계. 데이터 개체에 열 추가

이 단계에서는 LA_Customers_tgt 데이터 개체에 Firstname 및 Lastname 열을 추가합니다.

1. 개체 탐색기 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
2. LA_Customers_tgt 데이터 개체를 두 번 클릭하십시오.
편집기에서 LA_Customers_tgt 데이터 개체가 열립니다.
3. **개요** 보기가 선택되었는지 확인하십시오.
4. 열을 추가하려면 **FullName** 열을 선택하고 새로 만들기 단추를 클릭하십시오.
FullName1이라고 이름이 지정된 열이 표시됩니다.
5. 해당 열의 이름을 Firstname으로 바꾸십시오. **전체 자릿수** 필드를 클릭하고 "30."을 입력하십시오.
6. 열을 추가하려면 Firstname 열을 선택하고 새로 만들기 단추를 클릭하십시오.
FirstName1이라고 이름이 지정된 열이 표시됩니다.
7. 해당 열의 이름을 Lastname으로 바꾸십시오. **전체 자릿수** 필드를 클릭하고 "30."을 입력하십시오.
8. 데이터 개체를 저장하려면 **파일 > 저장**을 클릭하십시오.

태스크 2. 데이터 구문 분석용 매핑 생성

이 태스크에서는 매핑을 생성하고 데이터 개체와 파서 변환을 사용하도록 이 매핑을 구성합니다.

데이터를 구문 분석하는 매핑을 생성하려면 다음 단계를 수행하십시오.

1. 매핑을 생성하십시오.

2. 소스 및 대상 데이터 개체를 매핑에 추가하십시오.
3. 매핑에 파서 변환을 추가하십시오.
4. 고객의 전체 이름을 포함한 소스 열을 이름과 성을 포함한 개별 대상 열로 구문 분석하도록 파서 변환을 구성합니다.

1단계. 매핑 생성

이 단계에서는 매핑을 생성하고 이름을 지정합니다.

1. 개체 탐색기 보기에서 자습서 프로젝트를 선택하십시오.
2. **파일 > 새로 만들기 > 매핑**을 클릭하십시오.
새 매핑 창이 열립니다.
3. 이름 필드에 **ParserMapping**을 입력하십시오.
4. **마침**을 클릭합니다.
편집기에서 매핑이 열립니다.

2단계. 데이터 개체를 매핑에 추가

이 단계에서는 **LA_Customers** 데이터 개체와 **LA_Customers_tgt** 데이터 개체를 매핑에 추가합니다.

1. 개체 탐색기 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
2. **LA_Customers** 데이터 개체를 선택하고 이를 편집기로 끄십시오.
실제 데이터 개체를 매핑에 추가 창이 열립니다.
3. **읽기**가 선택되었는지 확인하고 **확인**을 클릭하십시오.
편집기에 이 데이터 개체가 표시됩니다.
4. 개체 탐색기 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
5. **LA_Customers_tgt** 데이터 개체를 선택하고 이를 편집기로 끄십시오.
실제 데이터 개체를 매핑에 추가 창이 열립니다.
6. **쓰기**를 선택하고 **확인**을 클릭하십시오.
편집기에 이 데이터 개체가 표시됩니다.
7. **LA_Customers** 데이터 개체에서 **CustomerID**, **CustomerTier** 및 **FullName** 포트를 선택하십시오. **LA_Customers_tgt** 데이터 개체에서 해당 포트를 **CustomerID** 포트로 끄십시오.
팁: 여러 포트를 선택하려면 **CTRL** 키를 누른 상태에서 선택하십시오.
LA_Customers 데이터 개체의 포트는 **LA_Customers_tgt** 데이터 개체의 해당 포트에 연결됩니다.

3단계. 파서 변환을 매핑에 추가

이 단계에서는 파서 변환을 **ParserMapping** 매핑에 추가합니다.

1. **ParserMapping** 매핑을 포함하는 편집기를 선택하십시오.
2. 변환 색상표에서 파서 변환을 선택하십시오.
3. 편집기를 클릭하십시오.
새 파서 변환 창이 열립니다.
4. **토른 파서**가 선택되었는지 확인하고 **마침**을 클릭하십시오.
편집기에 파서 변환이 표시됩니다.

5. LA_Customers 데이터 개체에서 FullName 포트를 선택하고 해당 포트를 파서 변환의 입력 그룹으로 끄십시오.
FullName 포트는 파서 변환에 표시되고 데이터 개체의 FullName 포트에 연결됩니다.

4단계. 파서 변환 구성

이 단계에서는 고객의 전체 이름을 포함한 열을 이름과 성을 포함한 개별 열로 구문 분석하도록 파서 변환을 구성합니다.

1. ParserMapping 매핑을 포함하는 편집기를 선택하십시오.
2. 파서 변환을 클릭하십시오.
3. 창 > 보기 표시 > 속성을 클릭하십시오.
4. 속성 보기에서 전략 보기를 선택하십시오.
5. 새로 만들기를 클릭합니다. 새로운 전략 마법사가 표시됩니다.
6. 입력 열에서 선택 화살표를 클릭하고 FullName 포트를 선택하십시오.
7. 공백 문자 구분자 [\s]를 선택하십시오.
8. 다음을 클릭합니다.
9. 토큰 집합을 사용하여 구문 분석 작업을 선택하고 다음을 클릭하십시오.
10. 고정 토큰 집합(단일 출력만)을 선택하고 정의되지 않음 토큰 집합을 선택하십시오.
11. 출력 필드를 클릭하고 새로 만들기를 선택하십시오.
12. 작업 출력 대화 상자에서 출력 이름을 Undefined_Output으로 변경하십시오.
13. 마침을 클릭합니다.
14. 파서 변환에서 Undefined_Output 포트를 클릭하고 LA_customers_tgt 데이터 개체의 FirstName 포트에 이를 끄십시오.
포트 사이에 연결이 나타납니다.
15. 파서 변환에서 OverflowField 포트를 클릭하고 LA_customers_tgt 데이터 개체의 LastName 포트에 이를 끄십시오.
포트 사이에 연결이 나타납니다.
16. 매핑을 저장하려면 파일 > 저장을 클릭하십시오.

태스크 3. 파서 변환에서 프로필 실행

이 태스크에서는 전체 이름을 올바르게 구문 분석하도록 파서 변환을 구성했는지 확인하기 위해 파서 변환에서 프로필을 실행합니다.

1. ParserMapping 매핑을 포함하는 편집기를 선택하십시오.
2. 파서 변환을 마우스 오른쪽 단추로 클릭하고 지금 프로파일링을 선택하십시오.
프로필이 편집기에서 실행되고 열립니다.
3. 프로파일링 작업의 결과를 표시하려면 편집기에서 결과 보기를 클릭하십시오.
4. 세부 정보 섹션에서 열에 관한 정보를 표시하려면 Undefined_output 열을 선택하십시오.
Undefined_output 열에 포함된 값이 각 값에 대한 빈도 및 백분율 통계와 함께 세부 정보 섹션에 나타납니다.
5. 데이터를 보고 Undefined_output 열에 이름만 표시되는지 확인하십시오.

태스크 4. 매핑 실행

이 태스크에서는 매핑을 실행하여 매핑 출력을 생성합니다.

1. **ParserMapping** 매핑을 포함하는 편집기를 선택하십시오.
2. **실행 > 매핑 실행**을 클릭하십시오.
매핑을 실행하여 출력을 **LA_Customers_tgt.csv** 파일에 씁니다.

태스크 5. 매핑 출력 보기

이 태스크에서는 매핑 출력을 보기 위해 데이터 뷰어를 실행합니다.

1. 개체 탐색기 보기에서 **LA_Customers_tgt** 데이터 개체를 자습서 프로젝트에서 찾고 해당 데이터 개체를 두 번 클릭하십시오.
편집기에서 데이터 개체가 열립니다.
2. **창 > 보기 표시 > 데이터 뷰어**를 클릭하십시오.
데이터 뷰어 보기가 열립니다.
3. 데이터 뷰어 보기에서 **실행**을 클릭하십시오.
데이터 뷰어는 데이터를 실행하고 이를 표시합니다.
4. **FirstName** 및 **LastName** 열에 올바르게 구문 분석된 데이터가 표시되는지 확인하십시오.

데이터 구문 분석 요약

이 강좌에서는 데이터 구문 분석이 입력 필드의 데이터 요소를 식별하고 각 요소를 새 열에 쓴다는 것을 학습했습니다.

데이터 구문 분석을 위해 파서 변환을 사용한다는 것을 학습했습니다. 해당 변환에서 출력을 분석하기 위해 매핑에서 변환에 대한 프로필을 생성할 수 있다는 것도 학습했습니다. 마지막으로 데이터 뷰어를 사용하여 매핑 출력을 볼 수 있다는 것을 학습했습니다.

구문 분석을 거친 출력을 포함하는 **LA_Customers_tgt** 데이터 개체를 생성하고 구성했습니다. 데이터를 구문 분석하는 매핑을 생성했습니다. 이 매핑에서는 로스앤젤레스 고객 파일의 **FullName** 열에서 이름과 성을 구문 분석하기 위해 토큰 집합으로 파서 변환을 구성했습니다. 구문 분석을 거친 데이터를 **LA_Customers_tgt** 데이터 개체의 **Firstname** 및 **Lastname** 열에 쓰기 위한 매핑을 구성했습니다. 매핑을 실행하기 전에 변환의 출력을 보기 위해 프로필을 실행했습니다. 마지막으로 매핑을 실행하고 데이터 뷰어를 사용하여 **LA_Customers_tgt** 데이터 개체의 새로운 데이터 열을 확인했습니다.

제 14 장

5과. 데이터 표준화

이 장에 포함된 항목:

- [데이터 표준화 개요, 67](#)
- [태스크 1. 대상 데이터 개체 생성, 68](#)
- [태스크 2. 데이터를 표준화하는 매핑 생성, 69](#)
- [태스크 3. 매핑 실행, 71](#)
- [태스크 4. 매핑 출력 보기, 71](#)
- [데이터 표준화 요약, 72](#)

데이터 표준화 개요

데이터 표준화는 데이터의 오류와 불일치를 제거함으로써 데이터 품질을 개선합니다.

데이터 품질을 개선하려면 다음과 같은 값 유형을 포함한 데이터를 표준화하십시오.

- 잘못된 값
- 정보는 올바르나 형식이 잘못된 값
- 새 정보를 파생시키고자 하는 시작점이 되는 값

데이터에서 이러한 값을 검색하려면 표준화 변환을 사용하십시오. 다음 검색 작업 유형 중 하나를 선택할 수 있습니다.

- **텍스트.** 입력하는 사용자 지정 문자열을 검색합니다. 이러한 문자열을 제거하고 사용자 지정 텍스트로 교체하십시오.
- **참조 테이블.** 선택하는 참조 테이블에 포함된 문자열을 검색합니다. 이러한 문자열을 제거하거나 참조 테이블 항목 또는 사용자 지정 텍스트로 교체하십시오.

예를 들어 **ST.**라는 교체 문자열을 사용하여 **Street** 및 **St.**라는 사용자 지정 문자열을 포함한 주소 데이터를 표준화하도록 표준화 변환을 구성할 수 있습니다. 표준화 변환은 검색 용어를 **ST.**로 교체하며 그 결과를 새로운 데이터 열에 씁니다.

사례

HypoStores는 모든 주소가 일관된 형식의 용어를 사용하도록 고객 주소 데이터를 표준화해야 합니다.

All_Customers 데이터 개체의 주소 데이터에는 **Street**, **Boulevard**, **Avenue**, **Drive** 및 **Park** 등 일반 용어처럼 일관된 형식으로 작성되지 않은 항목이 포함되어 있습니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 태스크를 완료합니다.

- 표준화된 데이터를 포함하도록 All_Customers_Stdz_tgt 데이터 개체를 생성하고 구성하십시오.
- 주소 용어 Street, Boulevard, Avenue, Drive 및 Park 등을 일관된 형식으로 표준화하는 매핑을 생성하십시오.
- 소스 데이터에 연결하려면 매핑에 All_Customers 데이터 개체를 추가하십시오.
- 대상 데이터 개체를 생성하려면 매핑에 All_Customers_Stdz_tgt 데이터 개체를 추가하십시오.
- 매핑에 표준화 변환을 추가하고 주소 용어를 표준화하도록 이를 구성하십시오.
- 표준화된 주소 데이터를 생성하는 매핑을 실행합니다.
- 매핑 출력을 보려면 데이터 뷰어를 실행하십시오.

선행 조건

이 강좌를 시작하기 전에 다음 선행 조건을 확인하십시오.

- 본 자습서의 강좌 1과 2를 완료했습니다.

소요 시간

이 강좌를 완료하는 데 15분 정도를 할애하십시오.

태스크 1. 대상 데이터 개체 생성

이 태스크에서는 표준화된 데이터를 쓸 수 있는 All_Customers_Stdz_tgt 데이터 개체를 생성합니다.

대상 데이터 개체를 생성하려면 다음 단계를 수행하십시오.

1. All_Customers.csv 파일을 바탕으로 All_Customers_Stdz_tgt 데이터 개체를 생성하십시오.
2. 파일 위치 및 파일 이름을 포함하여 데이터 개체에 대한 읽기 및 쓰기 옵션을 구성하십시오.

1단계. All_Customers_Stdz_tgt 데이터 개체 생성

이 단계에서는 All_Customers.csv 파일을 바탕으로 All_Customers_Stdz_tgt 데이터 개체를 생성합니다.

1. **파일 > 새로 만들기 > 데이터 개체**를 클릭하십시오.
 새로 만들기 창이 열립니다.
2. **플랫 파일 데이터 개체**를 선택하고 **다음**을 클릭하십시오.
3. **기존 플랫 파일에서 생성**이 선택되었는지 확인하십시오.
4. 다음 디렉터리에서 **찾아보기**를 클릭하고 All_Customers.csv로 이동하십시오. <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
5. **열기**를 클릭합니다.
6. 이름 필드에 All_Customers_Stdz_tgt를 입력하십시오.
7. **다음**을 클릭합니다.
8. **다음**을 클릭합니다.
9. **미리보기 옵션** 섹션에서 **첫 번째 행에서 열 이름 가져오기**를 선택하고 **다음**을 클릭하십시오.
10. **마침**을 클릭합니다.
 편집기에 All_Customers_Stdz_tgt 데이터 개체가 표시됩니다.

2단계. 읽기 및 쓰기 옵션 구성

이 단계에서는 파일 위치와 파일 이름을 포함해 All_Customers_Stdz_tgt 데이터 개체에 대한 읽기 및 쓰기 옵션을 구성합니다.

1. 편집기에서 All_Customers_Stdz_tgt 데이터 개체가 열려 있는지 확인하십시오.
2. 편집기에서 **읽기** 보기를 선택하십시오.
3. **창 > 보기 표시 > 속성**을 클릭하십시오.
4. **속성** 보기에서 **런타임** 보기를 선택하십시오.
5. **값** 열에서 소스 파일 이름을 두 번 클릭하고 All_Customers_Stdz_tgt.csv를 입력하십시오.
6. **값** 열에서 **소스 파일 디렉터리** 항목을 두 번 클릭하십시오.
7. 강조 표시된 이름을 마우스 오른쪽 단추로 클릭하고 **복사**를 선택하십시오.
8. 편집기에서 **쓰기** 보기를 선택하십시오.
9. **속성** 보기에서 **런타임** 보기를 선택하십시오.
10. **값** 열에서 **출력 파일 디렉터리** 항목을 두 번 클릭하십시오.
11. **읽기** 보기에서 복사한 디렉터리 위치를 붙여넣기 하려면 마우스 오른쪽 단추로 클릭하고 **붙여넣기**를 선택하십시오.
12. **값** 열에서 **헤더 옵션** 항목을 두 번 클릭하고 출력 필드 이름을 선택하십시오.
13. **값** 열에서 **출력 파일 이름** 항목을 두 번 클릭하고 All_Customers_Stdz_tgt.csv를 입력하십시오.
14. 데이터 개체를 저장하려면 **파일 > 저장**을 클릭하십시오.

태스크 2. 데이터를 표준화하는 매핑 생성

이 태스크에서는 매핑을 생성하고 데이터 개체와 표준화 변환을 사용하도록 이 매핑을 구성합니다.

데이터를 표준화하는 매핑을 생성하려면 다음 단계를 수행하십시오.

1. 매핑을 생성하십시오.
2. 소스 및 대상 데이터 개체를 매핑에 추가하십시오.
3. 매핑에 표준화 변환을 추가하십시오.
4. 일반 주소 용어를 일관된 형식으로 표준화하도록 표준화 변환을 구성하십시오.

1단계. 매핑 생성

이 단계에서는 매핑을 생성하고 이름을 지정합니다.

1. **개체 탐색기** 보기에서 자습서 프로젝트를 선택하십시오.
2. **파일 > 새로 만들기 > 매핑**을 클릭하십시오.
새 매핑 창이 열립니다.
3. **이름** 필드에 StandardizerMapping을 입력하십시오.
4. **마침**을 클릭합니다.
편집기에서 매핑이 열립니다.

2단계. 데이터 개체를 매핑에 추가

이 단계에서는 매핑에 All_Customers 데이터 개체와 All_Customers_Stdz_tgt 데이터 개체를 추가합니다.

1. **개체 탐색기** 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.

2. All_Customers 데이터 개체를 선택하고 이를 편집기로 끄십시오.

실제 데이터 개체를 매핑에 추가 창이 열립니다.

3. **읽기**가 선택되었는지 확인하고 **확인**을 클릭하십시오.

편집기에 이 데이터 개체가 표시됩니다.

4. **개체 탐색기** 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.

5. All_Customers_Stdz_tgt 데이터 개체를 선택하고 이를 편집기로 끄십시오.

실제 데이터 개체를 매핑에 추가 창이 열립니다.

6. **쓰기**를 선택하고 **확인**을 클릭하십시오.

편집기에 이 데이터 개체가 표시됩니다.

7. All_Customers 데이터 개체에서 모든 포트를 선택하십시오. All_Customers_Stdz_tgt 데이터 개체에서 해당 포트를 CustomerID 포트로 끄십시오.

팁: 여러 포트를 선택하려면 **Shift** 키를 누른 상태에서 선택하십시오. 모두 선택하려면 포트 목록을 아래로 스크롤해야 할 수도 있습니다.

All_Customers 데이터 개체의 포트는 All_Customers_Stdz_tgt 데이터 개체의 해당 포트에 연결합니다.

3단계. 매핑에 표준화 변환 추가

이 단계에서는 주소 데이터의 문자열을 표준화하기 위해 표준화 변환을 추가합니다.

1. StandardizerMapping 매핑을 포함하는 편집기를 선택하십시오.

2. 변환 색상표에서 표준화 변환을 선택하십시오.

3. 편집기를 클릭하십시오.

매핑에 NewStandardizer라고 이름이 지정된 표준화 변환이 표시됩니다.

4. 표준화 변환의 이름을 바꾸려면 해당 변환의 제목 표시줄을 두 번 클릭하고 AddressStandardizer를 입력하십시오.

5. All_Customers 데이터 개체에서 주소1 포트를 선택하고 해당 포트를 표준화 변환의 **입력** 그룹으로 끄십시오.

입력 그룹에 주소1이라고 이름이 지정된 포트가 표시됩니다. 해당 포트는 All_Customers 데이터 개체의 주소1 포트에 연결됩니다.

참고: 표준화 전략을 구성할 때는 출력 포트를 변환에 추가합니다.

4단계. 표준화 변환 구성

이 단계에서는 소스 데이터의 주소 용어를 표준화하기 위해 표준화 변환을 구성합니다.

참고: 이 태스크에서는 다섯 가지 표준화 작업을 정의합니다. 각 작업은 입력 열의 문자열을 새 문자열로 교체합니다.

1. StandardizerMapping 매핑을 포함하는 편집기를 선택하십시오.

2. 표준화 변환을 클릭하십시오.

3. **창 > 보기 표시 > 속성**을 클릭하십시오.

4. 속성 보기에서 **전략**을 선택하십시오.

5. **새로 만들기**를 클릭합니다. 새로운 전략 마법사가 표시됩니다.
6. 입력 열에서 선택 화살표를 클릭하고 주소1 입력 포트를 선택하십시오.
출력 필드는 주소1을 출력 포트로 표시합니다.
7. 공백 문자, 쉼표 구분자 [,] 및 \s]를 선택하십시오. 선택적으로 후행 공백을 제거하는 옵션을 선택하십시오.
8. **다음**을 클릭합니다.
9. **사용자 지정 문자열 바꾸기** 작업을 선택하고 **다음**을 클릭하십시오.
10. **속성** 아래에서 **새로 만들기**를 클릭하십시오.
11. 다음 테이블에서 문자열의 첫 번째 쌍을 포함할 수 있도록 **사용자 지정 문자열**과 **교체할 내용** 필드를 편집하십시오.

사용자 지정 문자열	교체할 내용
STREET	ST.
BOULEVARD	BLVD.
AVENUE	AVE.
DRIVE	DR.
PARK	PK.

12. 테이블의 모든 문자열에 대한 표준화 작업을 정의하려면 9부터 12까지의 단계를 반복하십시오.
13. 주소1 출력 포트를 All_Customers_Stdz_tgt 데이터 개체의 주소1 포트로 끄십시오.
14. 매핑을 저장하려면 **파일 > 저장**을 클릭하십시오.

태스크 3. 매핑 실행

이 태스크에서는 출력 데이터 개체에 표준화된 주소를 쓰는 매핑을 실행합니다.

1. StandardizerMapping 매핑을 포함하는 편집기를 선택하십시오.
2. **실행 > 매핑 실행**을 클릭하십시오.
매핑을 실행하여 출력을 All_Customers_Stdz_tgt.csv 파일에 씁니다.

태스크 4. 매핑 출력 보기

이 태스크에서는 매핑 출력을 보고 주소 데이터가 올바르게 표준화되었는지 확인하기 위해 데이터 뷰어를 실행합니다.

1. **개체 탐색기** 보기에서 All_Customers_Stdz_tgt 데이터 개체를 자습서 프로젝트에서 찾고 해당 데이터 개체를 두 번 클릭하십시오.
편집기에서 데이터 개체가 열립니다.
2. **창 > 보기 표시 > 데이터 뷰어**를 클릭하십시오.

데이터 뷰어 보기가 열립니다.

3. 데이터 뷰어 보기에서 **실행**을 클릭하십시오.

데이터 뷰어는 매핑 출력을 표시합니다.

4. 주소1 열이 올바르게 표준화된 데이터를 표시하는지 확인하십시오. 예를 들어 **STREET** 문자열의 모든 인스턴스는 **ST.** 문자열로 교체되어야 합니다.

데이터 표준화 요약

이 강좌에서는 데이터의 오류와 불일치를 제거하기 위해 데이터를 표준화할 수 있음을 학습했습니다.

입력 열에서 문자열을 표준화하는 데 표준화 변환을 사용할 수 있음을 학습했습니다. 데이터 뷰어를 사용하여 매핑 출력을 볼 수 있다는 것도 학습했습니다.

표준화된 출력을 포함하는 **All_Customers_Stdz_tgt** 데이터 개체를 생성하고 구성했습니다. 데이터를 표준화하는 매핑을 생성했습니다. 이 매핑에서는 **All_Customers** 데이터 개체에서 주소1 열을 표준화하도록 표준화 변환을 구성했습니다. **All_Customers_Stdz_tgt** 데이터 개체에 표준화된 출력을 쓰는 매핑을 구성했습니다. 마지막으로 매핑을 실행하고 데이터 뷰어를 사용하여 **All_Customers_Stdz_tgt** 데이터 개체의 표준화된 데이터를 확인했습니다.

제 15 장

6과. 주소 데이터의 유효성 검사

이 장에 포함된 항목:

- [주소 데이터 유효성 검사 개요, 73](#)
- [태스크 1. 대상 데이터 개체 생성, 74](#)
- [태스크 2. 주소의 유효성을 검사하는 매핑 생성, 76](#)
- [태스크 3. 주소 유효성 검사기 변환 구성, 77](#)
- [태스크 4. 매핑 실행, 80](#)
- [태스크 5. 매핑 출력 보기, 80](#)
- [주소 데이터 유효성 검사 요약, 82](#)

주소 데이터 유효성 검사 개요

주소 유효성 검사는 우편 주소의 품질을 평가하고 개선하는 프로세스입니다. 이는 입력 주소를 유효한 주소의 참조 데이터 집합과 비교함으로써 주소 품질을 평가합니다. 이는 올바르게 않은 주소 값을 식별하고 올바른 값이 포함된 필드를 생성하는 참조 데이터 집합을 사용함으로써 주소 품질을 개선합니다.

배달 가능한 주소의 경우 해당 주소는 유효한 것입니다. 주소가 올바른 형식으로 작성되어 있고 해당 주소에 실제 거리, 시 및 우편 번호 정보가 포함되어 있더라도 이 데이터를 통해 배달 가능한 주소가 도출되지 않는다면 그 주소는 유효한 것이 아닙니다. 개발자 도구는 입력 주소의 배달 가능성을 확인하는 데 주소 참조 데이터 집합을 사용합니다. Informatica는 주소 참조 데이터 집합을 제공합니다.

주소 참조 데이터 집합은 국가 내의 모든 배달 가능한 주소를 설명하는 데이터를 포함합니다. 주소 유효성 검사 프로세스는 입력 주소 데이터와 가장 유사한 주소에 대한 참조 데이터 집합을 검색합니다. 프로세스가 참조 데이터 집합에서 매우 유사한 항목을 찾는 경우 올바르게 않거나 불완전한 데이터 값에 대해 새로운 값을 씁니다. 프로세스는 입력 주소와 참조 주소 사이에서 발견된 일치 유형을 설명하는 영숫자 코드 집합을 생성합니다. 또한 이는 주소를 재구성할 수 있으며 미국 주소의 네 자리 우편 번호 접미사 등 입력 주소에 포함되지 않은 정보를 추가할 수도 있습니다.

개발자 도구에서 주소 유효성 검사 프로세스를 구축하는 데 주소 유효성 검사기 변환을 사용하십시오. 이 다중 그룹 변환에는 입력 주소의 가능한 모든 필드에 해당하는 미리 정의된 입력 포트 및 출력 포트 집합이 포함됩니다. 주소 유효성 검사기 변환을 구성할 때 기본 참조 데이터 집합을 선택하고 변환 포트를 사용하여 입력 및 출력 주소 구조를 생성합니다. 이 강좌에서는 미국 주소 데이터의 유효성을 검사하도록 변환을 구성합니다.

사례

HypoStores는 광고용 우편물 캠페인과 기타 소비자 우편물이 고객에게 전달되도록 보장하기 위한 올바르고 완전한 주소 데이터를 필요로 합니다. 올바르고 완전한 주소 데이터 또한 조직의 우편 발송 작업 비용을 감소시킵니다. 더불어 HypoStores는 다양한 길이의 주소를 포함할 수 있도록 유연하고 인쇄가 가능한 형식으로 주소를 포함할 수 있는 고객 데이터를 필요로 합니다.

HypoStores ICC 팀은 이러한 비즈니스 요구 사항을 충족하기 위해 개발자 도구에서 주소 유효성 검사 매핑을 생성합니다.

학습 목표

이 학습 과정을 진행하는 동안 다음 태스크를 완료합니다.

- 유효성 검사를 거친 주소 필드와 일치 코드를 포함할 대상 데이터 개체를 생성합니다.
- 소스 데이터 개체, 대상 데이터 개체 및 주소 유효성 검사기 변환을 포함한 매핑을 생성합니다.
- 고객의 주소 데이터의 유효성을 검사하도록 주소 유효성 검사기 변환을 구성합니다.
- 주소 데이터의 유효성을 검사하는 매핑을 실행하고 일치 코드 출력을 검토하여 해당 주소 데이터의 유효성을 확인합니다.

선행 조건

이 학습 과정을 시작하기 전에 다음 선행 조건이 충족되었는지 확인하십시오.

- 본 자습서의 강좌 1과 2를 완료했습니다.
- 미국 주소 참조 데이터는 도메인에 설치되며 Administrator 도구에 등록됩니다. 미국 주소 데이터가 사용자 시스템에 설치되어 있는지 확인하려면 Informatica 관리자에게 문의하십시오. 참조 데이터는 Data Quality 콘텐츠 설치 프로그램을 통해 설치됩니다.

소요 시간

이 강좌를 완료하는 데 25분 정도를 할애하십시오.

태스크 1. 대상 데이터 개체 생성

이 태스크에서는 대상 데이터 개체를 생성하고 쓰기 옵션을 구성하고 포트를 추가합니다.

대상 데이터 개체를 생성하고 구성하려면 다음 단계를 수행하십시오.

1. All_Customers.csv 파일을 바탕으로 All_Customers_av_tgt 데이터 개체를 생성하십시오.
2. 파일 위치 및 파일 이름을 포함하여 데이터 개체에 대한 읽기 및 쓰기 옵션을 구성하십시오.
3. 주소 유효성 검사기 변환으로 생성된 일치 코드 값을 받으려면 데이터 개체에 포트를 추가하십시오.

1단계. All_Customers_av_tgt 데이터 개체 생성

이 단계에서는 All_Customers.csv 파일을 바탕으로 All_Customers_av_tgt 데이터 개체를 생성합니다.

1. **파일 > 새로 만들기 > 데이터 개체**를 클릭하십시오.
 새로 만들기 창이 열립니다.
2. **플랫 파일 데이터 개체**를 선택하고 **다음**을 클릭하십시오.
3. **기존 플랫 파일에서 생성**이 선택되었는지 확인하십시오. 이 선택 사항 옆의 **찾아보기**를 클릭하고 All_Customers.csv 파일을 찾은 후 **열기**를 클릭하십시오.
4. 이름 필드에 All_Customers_av_tgt를 입력하십시오.
5. **다음**을 클릭합니다.
6. **다음**을 클릭합니다.
7. **미리보기 옵션** 섹션에서 **첫 번째 행에서 열 이름 가져오기**를 선택하고 **다음**을 클릭하십시오.

8. **마침**을 클릭합니다.
편집기에 All_Customers_av_tgt 데이터 개체가 표시됩니다.

2단계. 읽기 및 쓰기 옵션 구성

이 단계에서는 대상 파일 이름 및 위치를 포함해 All_Customers_av_tgt 데이터 개체에 대한 읽기 및 쓰기 옵션을 구성합니다.

1. 편집기에서 All_Customers_av_tgt 데이터 개체가 열려 있는지 확인하십시오.
2. 편집기에서 **읽기** 보기를 선택하십시오.
3. **창 > 보기 표시 > 속성**을 선택하십시오.
4. **속성** 보기에서 **런타임** 보기를 선택하십시오.
5. **값** 열에서 소스 파일 이름을 두 번 클릭하고 All_Customers_av_tgt.csv를 입력하십시오.
6. **값** 열에서 소스 파일 디렉터리 경로를 강조 표시하려면 두 번 클릭하십시오.
7. 강조 표시된 경로와 이름을 마우스 오른쪽 단추로 클릭하고 **복사**를 선택하십시오.
8. 편집기에서 **쓰기** 보기를 선택하십시오.
9. **속성** 보기에서 **런타임** 보기를 선택하십시오.
10. **값** 열에서 **출력 파일 디렉터리** 항목을 두 번 클릭하십시오.
11. **읽기** 보기에서 복사된 경로를 추가하려면 이 항목을 마우스 오른쪽 단추로 클릭하고 **붙여넣기**를 선택하십시오.
12. **값** 열에서 **헤더 옵션** 항목을 두 번 클릭하고 출력 필드 이름을 선택하십시오.
13. **값** 열에서 **출력 파일 이름** 항목을 두 번 클릭하고 All_Customers_av_tgt.csv를 입력하십시오.
14. 데이터 개체를 저장하려면 **파일 > 저장**을 선택하십시오.

3단계. 데이터 개체에 포트 추가

이 단계에서는 주소 유효성 검사기 변환이 대상 파일에 일치 코드 값을 쓸 수 있도록 All_Customers_av_tgt 데이터 개체에 2개의 포트를 추가합니다. 포트의 이름을 MailabilityScore 및 MatchCode로 지정하십시오.

MailabilityScore 값은 입력 주소의 배달 가능성을 설명합니다. MatchCode 값은 변환이 입력 주소와 참조 데이터 주소 사이에서 만드는 일치 유형을 설명합니다.

1. **개체 탐색기** 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
2. All_Customers_av_tgt 데이터 개체를 두 번 클릭하십시오.
편집기에서 All_Customers_av_tgt 데이터 개체가 열립니다.
3. **개요**가 선택되었는지 확인하십시오.
4. 포트 목록에서 최종 포트를 선택하십시오. 이 포트의 이름은 MiscDate로 지정되었습니다.
5. **새로 만들기**를 클릭합니다.
MiscDate1이라고 이름이 지정된 포트가 나타납니다.
6. MiscDate1 포트의 이름을 MailabilityScore로 바꾸십시오.
7. MailabilityScore 포트를 선택하십시오.
8. **새로 만들기**를 클릭합니다.
MailabilityScore1로 이름이 지정된 포트가 나타납니다.
9. MailabilityScore1 포트의 이름을 MatchCode로 바꾸십시오.
10. 데이터 개체를 저장하려면 **파일 > 저장**을 클릭하십시오.

태스크 2. 주소의 유효성을 검사하는 매핑 생성

이 태스크에서는 매핑을 생성하고 데이터 개체와 주소 유효성 검사기 변환을 추가합니다.

매핑을 생성하고 필요한 개체를 추가하려면 다음 단계를 수행하십시오.

1. 매핑 개체를 생성하십시오.
2. 소스 및 대상 데이터 개체를 매핑에 추가하십시오.
3. 주소 유효성 검사기 변환을 매핑에 추가하십시오.

1단계. 매핑 생성

이 단계에서는 매핑을 생성하고 이름을 지정합니다.

1. 개체 탐색기 보기에서 자습서 프로젝트를 선택하십시오.
2. **파일 > 새로 만들기 > 매핑**을 선택하십시오.
새 매핑 창이 열립니다.
3. 이름 필드에 `ValidateAddresses`를 입력하십시오.
4. 마침을 클릭합니다.
편집기에서 매핑이 열립니다.

2단계. 데이터 개체를 매핑에 추가

이 단계에서는 소스 및 대상 데이터 개체를 매핑에 추가합니다.

`All_Customers`는 매핑을 위한 소스 데이터 개체입니다. 주소 유효성 검사기 변환이 이 개체에서 데이터를 읽습니다. `All_Customers_av_tgt`는 매핑을 위한 데이터 대상 개체입니다. 이 개체는 주소 유효성 검사기 변환에서 데이터를 읽습니다.

1. 개체 탐색기 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
2. `All_Customers` 데이터 개체를 선택하고 이를 편집기로 끄십시오.
실제 데이터 개체를 매핑에 추가 창이 열립니다.
3. 읽기가 선택되었는지 확인하고 **확인**을 클릭하십시오.
편집기에 이 데이터 개체가 표시됩니다.
4. 개체 탐색기 보기에서 자습서 프로젝트의 데이터 개체로 이동하십시오.
5. `All_Customers_av_tgt` 데이터 개체를 선택하고 이를 편집기로 끄십시오.
실제 데이터 개체를 매핑에 추가 창이 열립니다.
6. 쓰기를 선택하고 **확인**을 클릭하십시오.
편집기에 이 데이터 개체가 표시됩니다.
7. 저장을 클릭합니다.

3단계. 매핑에 주소 유효성 검사기 변환을 추가

이 단계에서는 주소 유효성 검사기 변환을 소스 및 데이터 개체를 포함하는 매핑에 추가합니다.

이 단계가 완료되면 변환을 구성하고 해당 포트를 데이터 개체에 연결할 수 있습니다.

1. `ValidateAddresses` 매핑을 포함한 편집기를 선택하십시오.
2. 변환 색상표에서 주소 유효성 검사기 변환을 선택하십시오.

3. 편집기를 클릭하십시오.
편집기에 주소 유효성 검사기 변환이 표시됩니다.

태스크 3. 주소 유효성 검사기 변환 구성

이 태스크에서는 `All_Customers` 데이터 소스에서 주소를 읽고 유효성을 검사하도록 주소 유효성 검사기 변환을 구성합니다.

참고: 주소 유효성 검사기 변환에는 일련의 미리 정의된 입력 및 출력 포트가 포함됩니다. 필요한 포트를 선택하고 이를 매핑의 개체에 연결하십시오.

변환을 구성하려면 다음 단계를 수행하십시오.

1. 주소 유효성 검사를 위한 기본 국가를 선택합니다.
2. 변환 입력 포트를 구성합니다.
3. 변환 출력 포트를 구성합니다.
4. 사용되지 않은 소스 포트를 데이터 대상에 연결하십시오.

1단계. 주소 유효성 검사를 위한 기본 국가 설정

이 단계에서는 주소 유효성 검사를 위한 기본 국가를 선택합니다. 주소 유효성 검사기 변환이 사용하는 주소 참조 데이터 파일은 국가로 구성됩니다. 기본 국가를 선택하는 경우 변환이 국가 정보가 포함되지 않은 모든 입력 주소에 적용하는 국가 데이터 집합을 식별합니다.

1. 편집기에서 주소 유효성 검사기 변환을 선택하십시오.
2. **속성** 아래에서 **일반 설정**을 클릭하십시오.
3. **기본 국가** 메뉴에서 미국을 선택하십시오.

2단계. 주소 유효성 검사기 변환 입력 포트 구성

이 단계에서 변환 입력 포트를 선택하고 이러한 포트를 `All_Customers_av` 데이터 개체에 연결합니다.

주소 유효성 검사기 변환에는 미리 정의된 입력 포트의 일부 그룹이 포함됩니다. 입력 주소의 필드에 해당하는 입력 포트를 선택하고 이러한 포트를 변환에 추가하십시오.

단일 작업으로 여러 포트를 선택하려면 아래의 단계에서 포트를 선택할 때 **Ctrl** 키를 누른 상태에서 선택하십시오.

1. 편집기에서 주소 유효성 검사기 변환을 선택하십시오.
2. **속성** 아래에서, **템플릿**을 클릭합니다.
3. **기본 모델** 포트 그룹을 확장합니다.
4. **하이브리드** 입력 포트 그룹을 확장하고 다음 포트를 선택하십시오.

포트 이름	설명
배달 주소 행 1	거리 이름 및 건물 번호와 같은 거리 주소 데이터.
완전한 로컬리티 1	도시 또는 타운 이름.

포트 이름	설명
우편 번호 1	우편 번호.
시/도 1	시/도 또는 주 이름.
국가 이름	국가 이름 또는 약어.

참고: 단일 작업에서 다중 포트를 선택하려면 Ctrl 키를 누른 상태에서 선택합니다.

5. 포트 이름 목록 위의 도구 모음에서 **변환에 포트 추가**를 클릭합니다.
이 도구 모음은 **템플릿**을 선택할 때 표시됩니다.
선택된 포트는 매핑 편집기의 변환에 나타납니다.
6. 다음과 같이 소스 포트를 주소 유효성 검사기 변환 포트에 연결하십시오.

소스 포트	주소 유효성 검사기 변환 포트
주소 1	배달 주소 행 1
도시	완전한 로컬리티 1
우편 번호	우편 번호 1
상태	시/도 1
국가	국가 이름

3단계. 주소 유효성 검사기 변환 출력 포트 구성

이 단계에서 변환 출력 포트를 선택하고 이러한 포트를 All_Customers_av_tgt 데이터 개체에 연결합니다.

주소 유효성 검사기 변환에는 미리 정의된 출력 포트의 일부 그룹이 포함됩니다. 필요한 주소 구조를 정의하는 포트를 선택하고 이러한 포트를 변환에 추가합니다.

각 주소에 대해 수행된 유효성 검사 유형에 대한 정보를 담고 있는 포트를 선택할 수도 있습니다.

1. 매핑 편집기에서 주소 유효성 검사기 변환을 선택합니다.
2. **속성** 아래에서, **템플릿**을 클릭합니다.
3. **기본 모델** 포트 그룹을 확장합니다.
4. **주소 요소** 출력 포트 그룹을 확장하고 다음 포트를 선택합니다.

포트 이름	설명
완전한 거리 1	거리 이름 및 건물 번호와 같은 거리 주소 데이터.

5. **마지막 행 요소** 출력 포트 그룹을 확장하고 다음 포트를 선택합니다.

포트 이름	설명
완전한 로컬리티 1	도시 또는 타운 이름.

포트 이름	설명
우편 번호 1	우편 번호.
시/도 약어 1	시/도 또는 주 식별자.

참고: 단일 작업에서 다중 포트를 선택하려면 Ctrl 키를 누른 상태에서 선택합니다.

6. 국가 출력 포트 그룹을 확장하고 다음 포트를 선택합니다.

포트 이름	설명
국가 이름 1	국가 이름.

7. 상태 정보 출력 포트 그룹을 확장하고 다음 포트를 선택합니다.

포트 이름	설명
편지 점수	성공적인 우편 배달 확률을 나타내는 점수.
일치 코드	입력 주소와 참조 데이터 사이의 유사성 정도를 나타내는 코드.

8. 포트 이름 목록 위의 도구 모음에서 **변환에 포트 추가**를 클릭합니다.

이 도구 모음은 **템플릿**을 선택할 때 표시됩니다.

9. 다음과 같이 주소 유효성 검사기 변환 포트를 All_Customers_av_tgt 포트에 연결합니다.

주소 유효성 검사기 변환 포트	대상 포트
완전한 거리 1	주소 1
완전한 로컬리티 1	도시
우편 번호 1	우편 번호
시/도 약어 1	상태
국가 이름 1	국가
편지 점수	MailabilityScore
일치 코드	MatchCode

4단계. 데이터 대상에 사용되지 않은 데이터 소스 포트 연결

이 단계에서는 All_Customers 데이터 소스에서 사용되지 않은 포트를 데이터 대상에 연결합니다.

- ▶ 데이터 소스에서 사용되지 않은 포트를 데이터 대상에서 이름이 동일한 포트에 연결하십시오.

태스크 4. 매핑 실행

이 태스크에서는 매핑을 실행하여 매핑 출력을 생성합니다.

1. **ValidateAddresses** 매핑을 포함한 편집기를 선택하십시오.
2. **실행 > 매핑 실행**을 선택하십시오.
매핑을 실행하여 출력을 **All_Customers_av_tgt.csv** 파일에 씁니다.

태스크 5. 매핑 출력 보기

이 태스크에서는 매핑 출력을 보기 위해 데이터 뷰어를 실행합니다. 대상 데이터 개체에서 편지 점수와 일치 코드 열에 쓰여진 값을 검사함으로써 유효성 검사를 거친 주소의 품질을 검토하십시오.

일치 코드 값은 주소에 대해 매핑이 수행했던 유효성 검사 유형을 나타내는 영숫자 코드입니다.

편지 점수 값은 주소의 배달 가능성을 요약하는 한 자리 값입니다.

1. 개체 탐색기 보기에서 **All_Customers_av_tgt** 데이터 개체를 자습서 프로젝트에서 찾고 해당 데이터 개체를 두 번 클릭하십시오.
편집기에서 데이터 개체가 열립니다.
2. **창 > 보기 표시 > 데이터 뷰어**를 선택하십시오.
데이터 뷰어가 열립니다.
3. 데이터 뷰어에서 **실행**을 클릭하십시오.
데이터 뷰어는 매핑 출력을 표시합니다.
4. 편지 점수와 일치 코드 열이 표시되도록 매핑 결과 전체를 스크롤하십시오.
5. 편지 점수 열에서 값을 검토하십시오.
점수는 0에서 5까지의 범위로 표시됩니다. 점수가 더 높은 주소는 성공적으로 배달될 가능성이 더 큼을 의미합니다.
6. 일치 코드 열에서 값을 검토하십시오.
일치 코드는 영숫자 코드입니다. 영숫자 문자는 변환이 수행했던 유효성 검사 유형을 나타내며 자릿수는 최종 주소의 품질을 나타냅니다.
다음 테이블은 일치 코드 값을 설명합니다.

코드	설명
A1	주소 코드 조회로 입력 코드에 대한 부분적 주소 또는 전체 주소를 찾았습니다.
A0	주소 코드 조회로 입력 코드에 대한 주소를 찾지 못했습니다.
C4	수정됨. 모든 우편 관련 요소가 확인되었습니다.
C3	수정됨. 일부 요소를 확인할 수 없습니다.
C2	수정되었지만 참조 데이터가 없으므로 배달 상태가 확실하지 않습니다.

코드	설명
C1	수정되었지만 사용자 표준화로 인한 오류가 유발되어 배달 상태가 확실하지 않습니다.
I4	데이터는 완전하게 수정될 수 없지만 참조 데이터에는 단일 주소와 일치하는 단일 항목이 있습니다.
I3	데이터는 완전하게 수정될 수 없으며 참조 데이터에는 여러 주소와 일치하는 여러 항목이 있습니다.
I2	데이터를 수정할 수 없습니다. 일괄 모드는 부분적인 제안 주소를 반환합니다.
I1	데이터를 수정할 수 없습니다. 일괄 모드는 주소를 제안할 수 없습니다.
N7	유효성 검사 오류가 발생했습니다. 단일 행 유효성 검사가 잠금 해제되지 않았으므로 유효성 검사가 수행되지 않았습니다.
N6	유효성 검사 오류가 발생했습니다. 대상 국가에 대한 단일 행 유효성 검사가 지원되지 않으므로 유효성 검사가 수행되지 않았습니다.
N5	유효성 검사 오류가 발생했습니다. 참조 데이터베이스가 오래되어 유효성 검사가 수행되지 않았습니다.
N4	유효성 검사 오류가 발생했습니다. 참조 데이터가 손상되거나 잘못된 형식이므로 유효성 검사가 수행되지 않았습니다.
N3	유효성 검사 오류가 발생했습니다. 국가 데이터의 잠금이 해제될 수 없으므로 유효성 검사가 수행되지 않았습니다.
N2	유효성 검사 오류가 발생했습니다. 필요한 참조 데이터베이스를 사용할 수 없으므로 유효성 검사가 수행되지 않았습니다.
N1	유효성 검사 오류가 발생했습니다. 국가가 인식되지 않거나 지원되지 않아 유효성 검사가 수행되지 않았습니다.
Q3	제안 목록 모드. 주소 유효성 검사는 입력 주소에 해당하는 주소 참조 데이터에서 하나 이상의 완전한 주소를 검색할 수 있습니다.
Q2	제안 목록 모드. 주소 유효성 검사는 완전한 주소를 생성하기 위해 입력 주소 요소와 주소 참조 데이터의 요소를 결합할 수 있습니다.
Q1	제안 목록 모드. 주소 유효성 검사는 완전한 주소를 제안할 수 없습니다. 완전한 주소 제안을 생성하려면 데이터를 입력 주소에 추가하십시오.
Q0	제안 목록 모드. 제안을 생성하기에는 입력 데이터가 부족합니다.
RB	국가는 약어로 인식됩니다. ISO 두 문자 및 ISO 세 문자 국가 코드를 인식합니다. 독일의 "GER"과 같은 일반적 약어로도 국가를 인식할 수 있습니다.
RA	변환의 국가 강제 적용 설정에서 국가가 인식되었습니다.
R9	변환의 기본 국가 설정에서 국가가 인식되었습니다.
R8	국가 이름에서 국가가 인식되었습니다.

코드	설명
R7	국가명에서 국가가 인식되었으나 변환이 국가 데이터에서 오류를 식별했습니다.
R6	지역 데이터에서 국가가 인식되었습니다.
R5	시/도 데이터에서 국가가 인식되었습니다.
R4	주요 타운 데이터에서 국가가 인식되었습니다.
R3	주소 형식에서 국가가 인식되었습니다.
R2	스크립트에서 국가가 인식되었습니다.
R1	사용할 수 있는 일치 항목이 여러 개이므로 국가가 인식되지 않았습니다.
R0	국가가 인식되지 않았습니다.
S4	구문 분석 모드. 주소가 완전하게 구문 분석되었습니다.
S3	구문 분석 모드. 주소 구문 분석을 통해 다수의 결과가 도출되었습니다.
S1	구문 분석 모드. 입력 형식 불일치로 인해 구문 분석 오류가 발생했습니다.
V4	확인됨. 입력 데이터가 정확합니다. 주소 유효성 검사는 모든 우편 관련 요소를 확인했고 입력은 완전하게 일치되었습니다.
V3	확인됨. 입력 데이터는 올바르지만 일부 또는 모든 요소가 표준화되었거나 입력에 오래된 이름 또는 외국어 지명이 포함되어 있습니다.
V2	확인됨. 입력 데이터는 올바르지만 참조 데이터가 완전하지 않아 일부 요소를 확인할 수 없습니다.
V1	확인됨. 입력 데이터는 올바르지만 사용자 표준화가 배달 가능성에 부정적인 영향을 미쳤습니다. 예를 들어 우편 번호 길이가 너무 짧습니다.

주소 데이터 유효성 검사 요약

이 강좌에서는 주소 유효성 검사가 입력 주소 데이터와 참조 데이터를 비교하여 가능한 한 가장 정확한 버전의 주소를 반환한다는 것을 학습했습니다.

주소 유효성 검사 프로세스 또한 각 주소의 품질에 대한 상태 정보를 반환한다는 것도 학습했습니다.

Administrator 도구 사용자는 **Data Quality** 콘텐츠 설치 프로그램을 실행하여 주소 참조 데이터를 설치한다는 것을 학습했습니다.

주소 유효성 검사기 변환은 다중 그룹 변환이고, 포트 그룹에서 변환에 대한 입력 및 출력 포트를 선택한다는 것도 학습했습니다. 선택하는 입력 포트는 유효성 검사를 거치는 주소 콘텐츠를 결정합니다. 출력 포트는 최종 주소 레코드의 콘텐츠를 결정합니다.

부록 A

FAQ

이 부록에 포함된 항목:

- [Informatica Analyst FAQ, 83](#)
- [Informatica Developer FAQ, 83](#)

Informatica Analyst FAQ

Informatica Analyst에 관해 생길 수 있는 의문의 답을 구하기 위해 FAQ를 검토하십시오.

Administrator 도구, 개발자 도구 및 Analyst 도구에 액세스하는 데 단일 사용자 계정을 사용할 수 있습니까?

예. 세 가지 도구에 모두 액세스할 수 있는 사용자 권한을 제공할 수 있습니다. 각 클라이언트 응용 프로그램에 각각 별도의 사용자 계정을 생성할 필요는 없습니다.

참조 데이터는 어디에 저장됩니까?

참조 데이터 개체를 생성하고 공유하는 데 개발자 도구와 Analyst 도구를 사용할 수 있습니다. 모델 리포지토리는 참조 데이터 개체 메타데이터를 저장합니다. 참조 데이터 데이터베이스는 참조 테이블 데이터 값을 저장합니다. 콘텐츠 관리 서비스에서 참조 데이터 데이터베이스를 구성하십시오.

Informatica Developer FAQ

Informatica Developer에 관해 생길 수 있는 의문의 답을 구하기 위해 FAQ를 검토하십시오.

맵렛과 규칙 사이의 차이는 무엇입니까?

맵렛의 유효성을 특정 규칙으로 검사할 수 있습니다. 규칙은 예를 들어 프로필을 실행할 때 소스 데이터에 적용되는 조건을 정의하는 비즈니스 논리입니다. 맵렛이 다음 요구 사항을 충족할 때 맵렛의 유효성을 특정 규칙으로 검사할 수 있습니다.

- 이는 입력 및 출력 변환을 포함합니다.
- 활성 변환을 포함하지 않습니다.
- 이는 입력 그룹 간의 카디널리티를 지정하지 않습니다.

Data Engineering 제품 라이선스가 있습니다. Developer tool을 사용하여 PowerCenter로 개체를 내보낼 수 있습니까?

아니요. Engineering 제품은 PowerCenter와 통합되지 않습니다.

PowerCenter의 소스 및 대상과 Developer tool의 실제 데이터 개체 사이의 차이는 무엇입니까?

PowerCenter에서 매핑 소스로 포함시킬 소스 정의를 생성합니다. 매핑 대상으로 포함시킬 대상 정의를 생성합니다. Developer tool에서 매핑 소스 또는 대상으로 사용할 수 있는 실제 데이터 개체를 생성합니다.

Developer tool의 매핑과 PowerCenter의 매핑 사이의 차이는 무엇입니까?

PowerCenter 매핑은 소스와 대상 간에 데이터를 이동하는 방법을 지정합니다. Developer tool 매핑은 매핑 입력과 출력 간에 데이터를 이동하는 방법을 지정합니다.

PowerCenter 매핑은 하나 이상의 소스 정의, 소스 한정자 및 대상 정의를 포함해야 합니다. PowerCenter 매핑은 또한 바로 가기, 변환 및 맵셋을 포함할 수 있습니다.

Developer tool 매핑은 매핑 입력과 출력을 포함해야 합니다. Developer tool 매핑 또한 변환 및 맵셋을 포함할 수 있습니다.

Developer tool에는 다음과 같은 매핑 유형이 있습니다.

- 소스 및 대상 간에 데이터를 이동시키는 매핑. 이러한 매핑 유형은 바로 가기를 사용할 수 없고 소스 한정자를 사용하지 않는다는 면에서만 PowerCenter 매핑과 다릅니다.
- 논리적 데이터 개체 매핑. 논리적 데이터 개체 모델에서의 매핑. 논리적 데이터 개체 매핑은 논리적 데이터 개체를 매핑 입력으로, 데이터 개체를 매핑 출력으로 포함할 수 있습니다. 또는 하나 이상의 실제 데이터 개체를 매핑 입력으로, 논리적 데이터 개체를 매핑 출력으로 포함할 수 있습니다.
- 가상 테이블 매핑. SQL 데이터 서비스에서의 매핑. 이는 데이터 개체를 매핑 입력으로, 가상 테이블을 매핑 출력으로 포함합니다.
- 가상 저장 프로시저 매핑. SQL 데이터 서비스에서 비즈니스 논리 집합을 정의합니다. 이는 입력 매개 변수 변환 또는 실제 데이터 개체를 매핑 입력으로, 출력 매개 변수 변환 또는 실제 데이터 개체를 매핑 출력으로 포함합니다.

PowerCenter의 맵셋과 Developer tool의 맵셋 사이의 차이는 무엇입니까?

PowerCenter 및 Developer tool의 맵셋은 변환 집합을 포함하며 재사용이 가능한 개체입니다. 여러 매핑에서 변환 논리를 재사용할 수 있습니다.

PowerCenter 맵셋은 소스 정의 또는 입력 변환을 맵셋 입력으로 포함할 수 있습니다. 이는 출력 변환을 맵셋 출력으로 포함해야 합니다.

Developer tool 맵셋은 데이터 개체 또는 입력 변환을 맵셋 입력으로 포함할 수 있습니다. 이는 데이터 개체 또는 출력 변환을 맵셋 출력으로 포함할 수 있습니다. 또한 Developer tool의 매핑에는 다음 피처가 포함됩니다.

- 맵셋의 유효성을 특정 규칙으로 검사할 수 있습니다. 프로필에서 규칙을 사용합니다.
- 맵셋은 다른 맵셋을 포함할 수 있습니다.

인덱스

A

Analyst 도구 설정
개요 [18](#)

D

Developer tool 설정
개요 [44](#)

N

데이터 개체 작성
개요 [21](#)
데이터 프로파일링
개요 [57](#)

사용자 지정 프로필 작성
개요 [27](#)
성과 기록표 작성
개요 [33](#)
식 규칙 작성
개요 [30](#)
실제 데이터 개체 가져오기
개요 [48](#)
열에서 참조 테이블 작성
개요 [38](#)
참조 테이블
개요 [41](#)

기

기본 프로필 작성
개요 [24](#)