

# Informatica Enterprise Data Preparation

## 이점

- 셀프서비스 방식으로 분석 프로젝트에 적합한 데이터 검색
- 필요한 데이터를 신속하게 준비 및 공유
- 재사용성 증진을 위한 간편한 데이터 준비 운영화

## 끊임없는 원시 데이터 탐색과 고급 분석

데이터가 디지털 변혁의 근간이라는 데 이견은 없습니다. 조직마다 이전에는 상상할 수 없었던, 또는 손에 넣기란 불가능했던 통찰력을 도출하기 위해 Apache Hadoop과 같은 새로운 데이터 처리 플랫폼을 활용합니다. Apache Hadoop과 데이터 레이크 접근 방식의 등장으로 이제 조직은 모든 데이터를 통합하여 사용자가 언제든지 데이터에 액세스해 모든 유형의 분석을 수행할 수 있는 이점을 누리게 되었습니다.

Hadoop 시스템으로 수집되고 있는 막대한 양의 데이터가 IT 조직에 큰 부담으로 다가오고 있습니다. 현업 부서 분석가는 Hadoop에서 고품질 데이터를 기다리지만, IT 담당자로서는 원시 데이터를 용도에 맞는 데이터 자산으로 준비하는 데 있어 시간 소모가 큰 수작업 프로세스 때문에 부담을 느끼고 있습니다.

데이터 준비를 위한 확장 가능하고 반복 가능한 지능형 메커니즘 없이는 데이터 레이크가 약속한 모든 기회가 위험으로 인한 정체를 야기합니다. 이른바 데이터 늪이라고 불리는 위기를 극복할 열쇠는 바로 CLAIRE™ 엔진으로 알려진 Informatica®의 메타데이터 기반 인공지능 기술로서, 데이터 자산과 관련된 여러 관계를 자동으로 탐색, 프로파일링, 추론할 수 있습니다.

Informatica Enterprise Data Preparation이 체계적인 원시 빅 데이터 탐색을 돕기 때문에 현업 부서 분석가는 데이터 세트를 신뢰할 만한 정보로 셀프서비스 방식에 따라 변환할 수 있게 됩니다. 데이터 과학자와 현업 부서 분석가는 시맨틱 및 패시 검색을 사용해 신속하게 원하는 데이터를 찾는 동시에 자동으로 데이터 연계 및 데이터 관계를 파악할 수 있습니다. 또한 데이터 분석가 팀은 프로젝트 작업 공간에서 다른 사람과 쉽게 협업하고 결과를 공유할 수 있습니다. 데이터 세트를 프로젝트 작업 공간에 추가하면 기계 학습 알고리즘에서 분석가가 사용할 만한 대체 데이터 세트를 추천하도록 배경에서 작업을 수행합니다.

사용자는 Enterprise Data Preparation에서 Excel처럼 이용이 편리한 데이터 준비 도구를 사용해 데이터 세트를 프로젝트 작업 공간 내에서 열 수 있습니다. 메타데이터 기반의 데이터 준비 접근 방식은 빅 데이터를 지속 가능한 비즈니스 가치를 제공하는 믿을 만한 정보 자산으로 변환하는 지능적 방안이라 할 수 있습니다.

## 주요 특징

### 지능형 검색 및 지능형 가시화

지능형 시맨틱 검색 및 추론 기반 결과를 사용하여 데이터 레이크와 그 외의 기업 시스템에서 데이터를 검색해 보십시오. 시스템 속성을 이용하여 동적 패킷 기반 데이터 자산을 필터링하고 Apache Zeppelin 기반 차트를 통해 가시화할 수 있습니다.

### 포괄적 데이터 탐색

맞춤 특성, 데이터 품질에 대한 프로파일링 통계, 비즈니스 콘텐츠를 위한 데이터 도메인 및 사용 정보를 포함하여 데이터 자산에 대한 개요를 확인해 보십시오. 태깅을 통해 데이터 세트에 대한 정보를 클라우드소싱하고 메타데이터를 보강하여 비즈니스 컨텍스트를 추가할 수 있습니다. 사용자 자격 증명을 기반으로 샘플 데이터를 미리 보고 데이터를 빠르게 이해할 수 있습니다. 다른 테이블 또는 뷰, 사용자, 보고서 및 데이터 도메인과의 연결을 기반으로 데이터 자산이 기업의 다른 자산과 어떻게 관련되어 있는지 파악할 수 있습니다.

직관적인 Excel과 유사한 인터페이스를 사용하여 데이터를 필터링, 집계, 병합 및 결합하는 기본적인 변환을 통해, 대화형 방식으로 분석을 위해 데이터를 준비하십시오. 문자열, 수식, 날짜 및 논리 작업을 사용하여 열 수준 데이터 정제 및 데이터 변환을 수행하십시오. 안내형 인텔리전스가 데이터 세트 혼합 시 결합 키 추천과 같이 데이터 세트 준비에 도움을 줍니다. 시트 수준 및 열 수준의 기술통계 개요를 참조하십시오(값 분포 및 숫자, 날짜 분포 포함). 데이터 흐름을 자동으로 생성하는 데 사용 가능한 레시피에 모든 절차가 기록되어 있어 분석 통찰력의 운영화를 반복해서 계획할 수 있습니다.

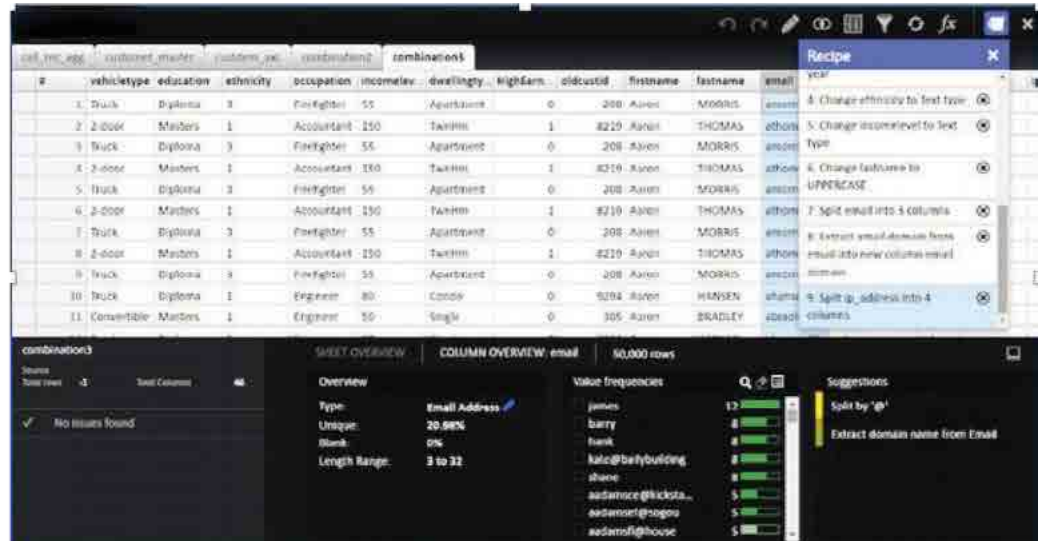


그림 1: 지능형 시맨틱 검색 및 동적 패킷을 사용하여 데이터 세트를 신속하게 찾습니다.

### 데이터 준비 운영화

현업 부서 분석가는 사전에 마련된 수많은 비즈니스 규칙을 사용하여 제품의 데이터 품질을 체계적으로 개선하는 한편 반복 실행을 위해 데이터 준비 절차를 IT 담당자에게 맡길 수 있습니다.

### 기업 협업

데이터 퍼블리케이션을 셀프서비스 방식으로 관리하는 한편 데이터 자산을 프로젝트 작업 공간에 추가하여 작업을 조직해 보십시오. 팀원을 공동 소유자, 편집자, 보기 전용 사용자 등 개별적 권한을 지닌 여러 역할로 프로젝트에 추가하여 다른 분석가와 협업할 수 있습니다.

### 데이터 자산 추천

다른 사용자의 공유 지식 및 행동에 적용된 기계 학습 알고리즘을 기반으로, 자동화된 추천을 통해 생산성을 향상시키고 신뢰할 수 있는 자산의 재사용을 늘리십시오. 프로젝트에 추가된 데이터 자산을 기반으로 하는 프로젝트에 대해 대체 자산 및 추가 자산이 추천됩니다.

### 마법사 기반 데이터 업로드

마법사 기반 인터페이스를 사용하여 개인의 구분된 파일을 데이터 레이크에 업로드하십시오. 가장 적절한 형식으로 업로드할 수 있도록 Hive 테이블이 자동으로 생성됩니다. 업로드된 데이터에 대해 자산을 생성, 추가 또는 덮어쓸 수 있습니다.

## 핵심 장점

### 모든 데이터의 검색과 액세스

현업 부서 분석가는 일반적으로 빅 데이터와 관련된 점점 증가하는 '볼륨, 다양성 및 속도'를 효율적으로 관리할 수 있는 방법을 찾고 있습니다. 지능형 시맨틱 검색 및 동적 패킷을 활용하여 결과를 필터링하고 믿을 만한 데이터 자산을 손쉽게 검색해 보십시오. 자동화된 기계 학습 기반 검색 프로세스는 관련된 데이터 자산을 분석가가 관심을 가질 만한 새로운 데이터 자산에 대한 지능형 추천으로 변환합니다. 그 결과 신뢰성이 크게 증가하고 유사한 프로젝트에 대해 생성되는 중복 데이터 세트가 줄어듭니다.

### 거버넌스와의 협업

빅 데이터 분석 프로젝트의 효율성을 높이기 위해 현업 부서 분석가는 프로젝트 작업 공간을 사용하여 데이터 세트에 대한 협업을 진행합니다. 데이터 세트를 프로젝트 작업 공간에 추가하면 프로파일 통계, 데이터 세트의 엔드 투 엔드 계보(Lineage) 및 모든 관련된 데이터 자산, 데이터 도메인, 사용자 등을 확인할 수 있습니다. 이렇게 하면 데이터 품질 평가, 신뢰할 수 있는 데이터 공유 및 프로젝트에 유용한 기타 데이터 세트를 적극적으로 검색하는 데 도움이 됩니다. 또한 현업 부서 분석가는 데이터 퍼블리케이션을 셀프서비스 방식으로 관리할 수도 있습니다. 역할 기반 보안은 프로젝트에 추가된 분석가에게만 데이터 액세스 권한을 보장합니다.

## Informatica 정보

디지털 변혁은 우리의 기대치를 바꿔 놓았습니다. 이제 더 적은 비용으로 더 나은 서비스를 더 빠르게 제공할 수 있어야 합니다. 이러한 상황에 부응하기 위해서는 기업이 변화해야 하며, 데이터가 이에 대한 해답을 쥐고 있습니다.

엔터프라이즈 클라우드 데이터 관리 분야의 세계적인 선도 기업인 Informatica는 모든 부문, 카테고리, 틈새시장에서 지능적인 방식으로 고객을 지원할 준비가 되어 있습니다. Informatica는 더욱 민첩하게 운영하고, 새로운 성장 기회를 발견하거나 새로운 혁신을 이룰 수 있는 통찰력을 기업에 제공합니다. 또한 모든 종류의 데이터에 100% 집중하여 성공에 필요한 다양한 서비스를 제공하고 있습니다.

Informatica가 제공하는 모든 서비스에 대해 알아보고 데이터의 힘을 활용하여 미래의 지능형 혁신을 실현하시기 바랍니다.

## 필요한 데이터를 신속하게 준비 및 공유

비즈니스 주기가 계속해서 단축됨에 따라, 속도는 현업 부서 분석가가 비즈니스 가치를 더하기 위한 경쟁에서 의존할 수 있는 얼마 안 되는 경쟁 우위 중 하나가 되었습니다. 경쟁력 있는 분석을 제공하는 데 있어 중요한 데이터를 신속하게 준비하고 공유하십시오. Informatica의 셀프서비스 데이터 준비 기능은 현업 부서 분석가를 위해 친숙하고 사용하기 쉬운 Excel과 유사한 인터페이스를 제공하며, 이를 통해 현업 부서 분석가는 필요로 하는 통찰력에 데이터를 신속하게 결합, 필터링 및 혼합할 수 있습니다. 클라우드소싱된 데이터 자산 태깅 및 공유 기능을 통해 현업 부서 분석가는 데이터 분류 프로세스를 관리하여 운영 효율성을 개선할 수 있습니다.

## 데이터 준비를 재사용 가능한 워크플로우로 운영화

사용하고 있는 자동화 및 셀프서비스 툴과 관계없이 분석가는 새로운 데이터 세트에서 데이터 준비 작업을 자주 반복해야 합니다. 이는 지속적인 확장 및 재사용성을 통해 누릴 수 있는 혜택을 방해합니다. Informatica Enterprise Data Preparation은 사용자가 반복적 자동 데이터 파이프라인으로 계획할 수 있는 데이터 준비 절차를 기록합니다. 이러한 과정을 통해 데이터 준비를 수동 프로세스에서 재사용 가능하고 지속 가능한 시스템으로 변환합니다.

보다 자세한 내용은 [Informatica Enterprise Data Preparation 제품 페이지](#)를 참조하십시오.



한국 인포매티카, 한국인포매티카 06611 서울시 서초구 서초동 강남대로 465 교보타워 B동 13층, 대표 전화: +82 2 6293 5001

IN06\_0319\_03237

© Copyright Informatica LLC 2019. Informatica, Informatica 로고 및 CLAIRE는 미국 및 기타 국가에서 Informatica LLC의 상표 또는 등록 상표입니다. Informatica 상표의 최신 목록은 웹페이지 (<https://www.informatica.com/trademarks.html>)에서 확인할 수 있습니다. 기타 회사 및 제품 이름은 해당 소유주의 상품명 또는 등록 상표일 수 있습니다. 이 문서의 정보는 예고 없이 변경될 수 있으며 일체의 명시적 또는 묵시적인 보증 없이 '있는 그대로' 제공됩니다.