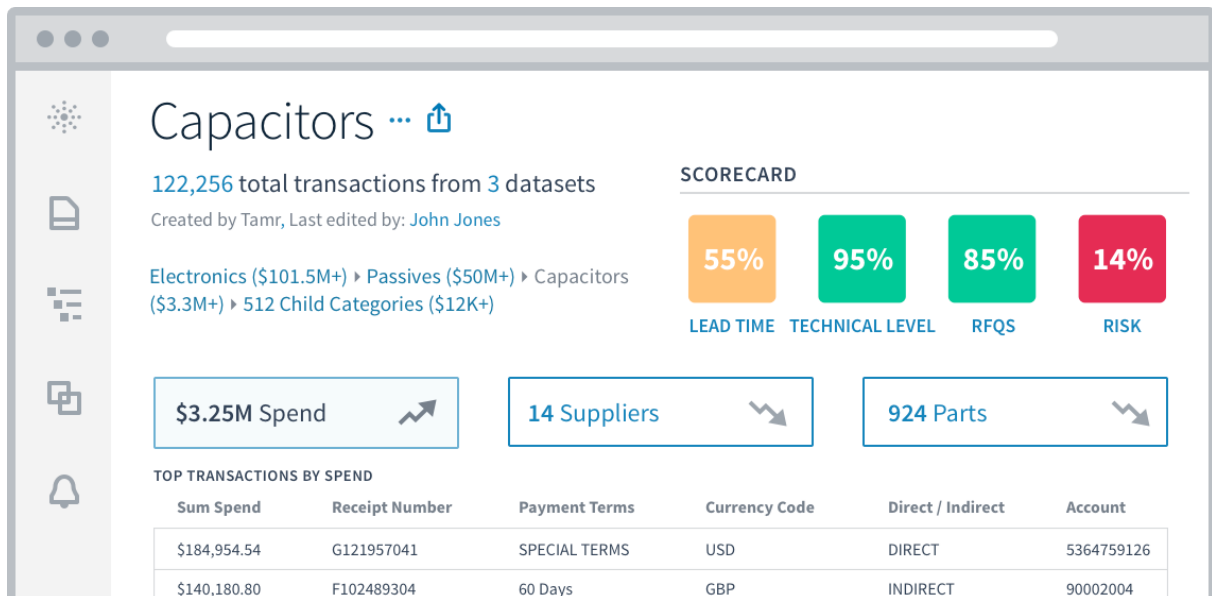


TAMR PRODUCT INTRODUCTION

The Contents

INTRODUCTION	03
ABOUT Tamr	04
핵심역량	05
CONNECT	06
CLEAN	08
CLASSIFY	09
제품구조	10
MARKET POSITIONING	12
USE CASES	14

TAMR

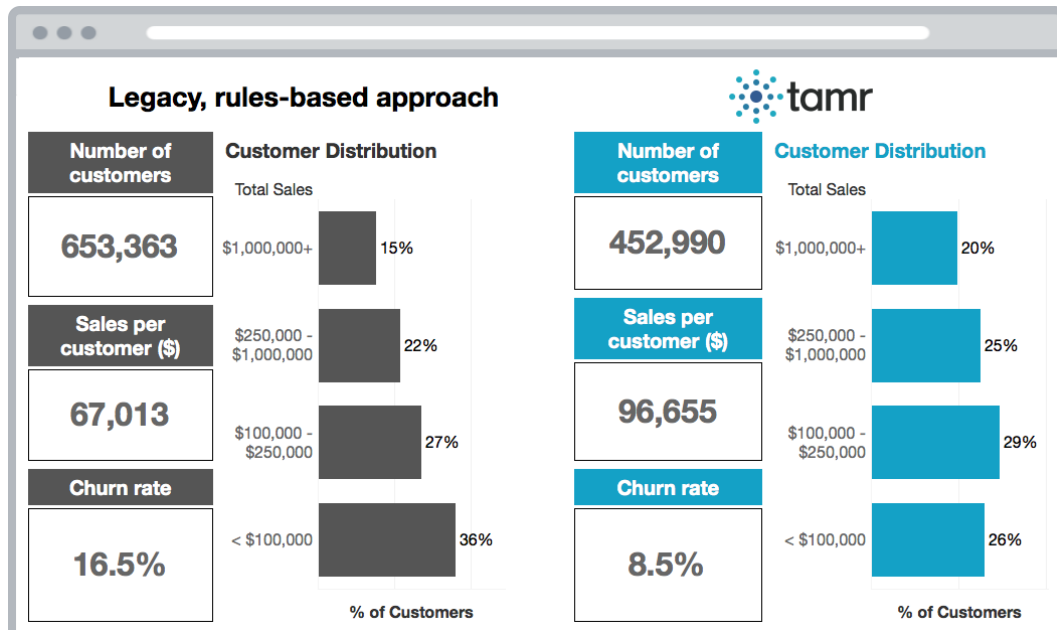


TAMR는 2013년 Andy Palmer와 Mike Stonebraker 두 공동 창업자에 의해 설립 되었습니다. 이 두 창업주는 데이터베이스 업계의 거장으로, 현재는 휴렛패커드(HP)에 인수된 Vertica System를 공동 설립했고, 그 외에도 여러 다른 회사와 협력했습니다.

그들은 지난 20년 간의 데이터 관리 방법이 오늘날 기업들의 요구를 충족시키지 못하고 있다는 공통된 생각을 가지고 있었습니다. 기존의 방법으로는 폭발적으로 증가하고 다양해지는 기업의 데이터를 감당 할 수 없다고 판단하여 2012년 MIT 의 컴퓨터 과학 & AI 연구소에서 볼륨(Volume), 속도(Velocity)와 다양한 (Variety) 종류의 기업 데이터를 관리하기 위한 상향식 솔루션에 대한 연구를 시작했습니다.

마침내 2013년, “Data Curation at Scale: The Data Tamer System” 논문을 통해, 머신 러닝과 인적 전문가 가이드를 결합하여 수 천개의 데이터 소스를 통합하는 획기적인 접근법을 설명하였습니다. 그 논문을 바탕으로 “TAMR”가 탄생했습니다

엔터프라이즈 데이터 통합 솔루션



▶ 기존의 Rules-based 접근방식과 Tamr의 접근방식의 차이

TAMR 는 방대하고 다양한 대규모의 데이터 관리 문제를 해결하기 위해 기존의 기술 (legacy technologies)과는 다른 접근방식을 가지고 설립되었습니다. 대부분의 솔루션은 규칙기반의 Top-Down 방식의 데이터관리로 진행해왔다면, TAMR는 조직 내의 이질적인 데이터셋을 통합하기 위한 머신러닝 기반의 Bottom-up방식에 초점을 맞추고 있습니다.

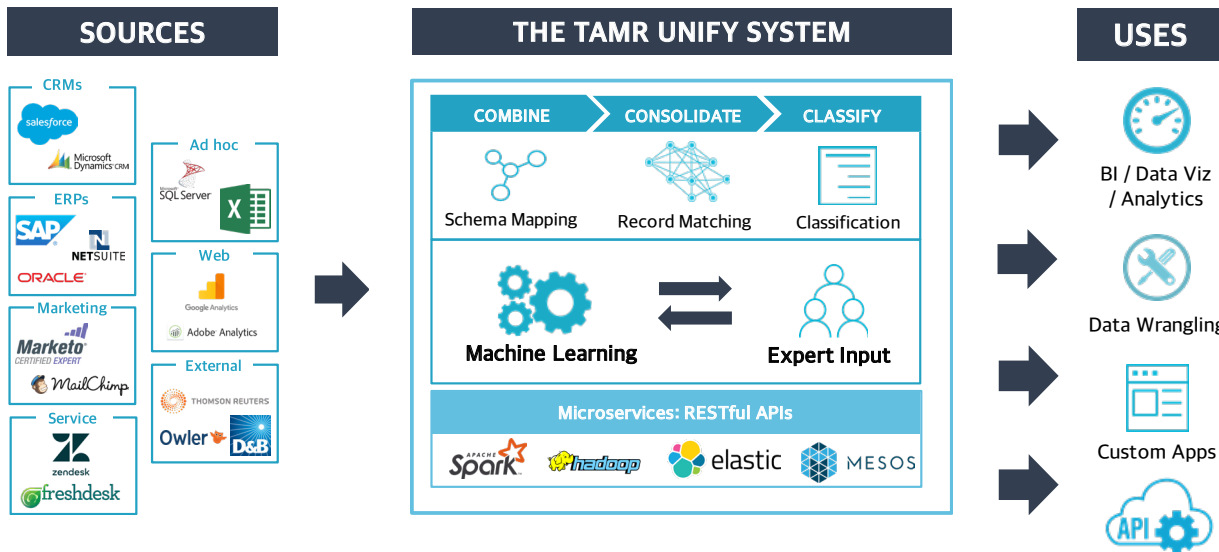
TAMR는 엔터프라이즈급 데이터 통합 솔루션입니다. 머신러닝의 기술과 전문가의 피드백을 결합하여 조직 전반에 걸친 데이터 소스를 신속하고, 정확하게, 확장가능한 형태로 통합할 수 있습니다.

플랫폼의 핵심 기능은 조직 전체에 데이터 소스를 "연결(Connecting)"하여 관련 데이터셋을 타깃 스키마에 맞추고, 레코드 중복 제거 및 마스터링을 통해 통합 데이터셋을 "정리(Cleansing)"하여 통합 데이터셋을 "분류(Classifying)"한 결과를 고객에 제공함으로써 더 강력한 다운 스트림 분석을 제공합니다.

TAMR로 만든 통합 데이터셋은 분석 툴에서 데이터 웨어하우스 (Data Warehouses)까지 다양한 제품들에서 활용될 수 있습니다.

또한 TAMR는 기존 솔루션(MDM,ETL)뿐만 아니라 더 나은 기술(Data Catalogs, Self-Service Preparation Tool, and Analytic Tools)과 상호 보완적인 데이터 관리 기술로써 사용될 수 있습니다.

엔터프라이즈급 데이터 통합기술은 산업 전반에서 입증된 요구사항이며, TAMR는 다양한 적용 사례 속에서, 복잡한 환경에서의 통합 구현을 실현시킴으로써 수백만 달러 이상의 가치를 제공해 왔습니다.



핵심역량

TAMR는 대기업의 분산된 데이터 사일로(Silos)의 통합을 자동화하기 위해 개발된 머신 러닝과 휴먼 가이드(human expert guidance)를 결합한 엔터프라이즈 데이터 통합 플랫폼입니다. TAMR는 이전에는 불가능했던 획기적인 분석 돌파구를 제공합니다. TAMR는 방대한 데이터를 신속하고 정확하고 효율적인 비용으로 통합할 수 있는 유일한 시스템입니다. 다운스트림 분석을 위해 기업 차원에서 신속하고 정확하게 데이터를 준비할 수 있도록 TAMR는 MIT의 특허 받은 휴먼 가이드 머신 러닝 기반의 접근 방식을 사용하는 세가지 핵심 기능을 설계했습니다.

Connect

“Connect”단계에서 Tamr는 모든 관련 소스의 데이터셋을 프로젝트 목표와 관련하여 가장 효과적이고 적절한 통합 스키마로 정렬합니다.

휴먼 가이드 머신 러닝 기반의 데이터셋을 결합하는 방식은 개발자들에게 의존하는 기존 방식과 비교할 때, 속도와 규모를 대폭 개선합니다.

Clean

“Clean”단계에서는 머신러닝 엔진을 통해 통합 데이터셋 내에서 불필요한 엔티티의 제거 및 표준화 작업을 자동화 합니다.

이 단계의 핵심 기능은 사용자가 통합 데이터셋 내에서 중복된 데이터를 식별하고 내부에 포함된 레코드를 파악할 수 있도록 허용함으로써 정확한 결과를 얻는 것입니다. 중복 제거 레코드식별에 대한 학습이 가능한 경우, TAMR는 모델에 적용합니다.

Classify

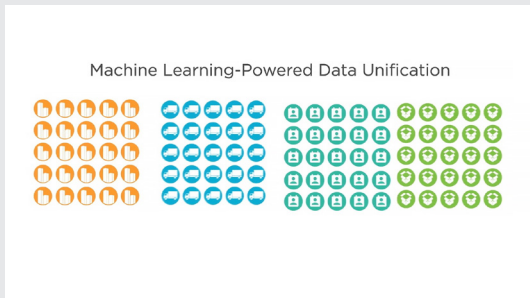
TAMR에 의해 특정 엔티티의 정제된 통합 데이터셋이 작성되면, 사용자는 좀 더 심층적인 분석 기능을 위해 해당 레코드를 기업에 특화된 또는 일반적으로 사용되는 분류 체계(Taxonomy)로 “Classifying” 할 수 있습니다.

TAMR의 분류단계는 Connect 및 Clean 단계와 동일하게 작동하며, 제품의 고유한 조합을 휴먼 가이드 머신러닝 엔진을 사용하여 제공된 분류 체계의 가장 심층적인 수준으로 기록을 빠르고 정확하게 분류할 수 있습니다.

CONNECT

“Connect” 단계는 프로젝트 목표와 엔티티(개인, 장소 또는 사물)를 정의하고, 사용자가 다운 스트림 분석의 목적을 위한 통일된 뷰를 추구하는 것으로 시작합니다. “Connect” 단계에서 TAMR는 모든 관련 소스 데이터셋 속성(attribute)¹을 프로젝트 목표에 가장 효과적이고 적절한 타깃 스키마로 정렬합니다. 휴먼 가이드 머신 러닝은 이러한 데이터셋을 결합하는데 사용되며, 쓰기 스크립트에 의존하는 기존 방식에 비해 속도와 규모가 상당히 향상됩니다. TAMR는 다음과 같은 방식으로 이 작업을 수행합니다.

1) 속성(attribute)은 관계형 데이터베이스 용어에서 컬럼과 같은 의미로 사용됩니다.



TAMR는 APIs, JDBC 커넥터 또는 커넥터 세트를 통해 소스 시스템 (Databases, HDFS, CSVs and Flat Files)으로부터 데이터를 수집합니다. TAMR 플랫폼은 데이터를 수집하기 전에 상대적으로 구조화되고, 태그 지정된 데이터를 요구할 수 있으며, JSON, Multivalues, tabular 및 XML과 같은 데이터 형식을 사용할 수 있습니다.

이러한 데이터소스는 사용자가 고객, 공급 업체 등에 포함된 엔티티 유형을 식별할 수 있도록 프로파일링합니다. 사용자는 분석에 적합한 데이터를 선택 할 수 있습니다.

TAMR 시스템에서 타깃 스키마를 식별하여 최종 통합 데이터셋에 최적의 특성이 표현 되는지 확인합니다. 사용자는 데이터 소스의 특성을 선택하여 타깃 스키마를 만들거나, 각 속성에 대한 값 샘플을 사용하여 자체 스키마를 구축 할 수 있습니다.

이 단계에서 TAMR는 소스 기반 데이터 집합 특성을 타깃 스키마에 맞추기 위해 “human-guided machine learning”을 사용합니다.

Unify Your Attributes

</

▶ TAMR는 모든 관련 소스 데이터셋 속성을 프로젝트 목표에 가장 효과적이고 적절한 타깃 스키마로 정렬합니다.

++ 첫번째로 TAMR는 타깃 스키마 내의 값을 비교하기 위한 기준으로 자료를 수집합니다. 여기에는 필드 이름, 설명/ 주석, 데이터 형식 및 유효성 검사와 같은 메타데이터가 포함됩니다.

++ TAMR는 소스 속성안에 동일한 정보를 수집하고 비지도 학습(Un-supervised machine learning)을 사용하여 소스와 타깃 스키마 속성 (TAMR가 생성한 초기모델을 기반으로 만든)간의 잠재적인 일치 정도를 확인합니다.

++ TAMR의 매칭의 정확도를 평가하고 시스템으로부터 학습할 수 있습니다. 다른 특성들을 대상으로 타깃 스키마와 소스 데이터셋 간의 일치성 정도에 대해 yes, no의 간단한 질문으로 TAMR는 'high impact questions'를 생성합니다.

++ 예를 들어, TAMR가 'mail address'와 통합 속성 'street address'가 같은 소스 속성인지에 대해 낮은 신뢰도를 가지고 있을때, 고객집단에 대한 피드백을 얻기 위해 사용자에게 질문을 할 것입니다. 이 프로세스는 정확성을 향상시켜 신뢰성 높은 결과를 보장하고, TAMR의 알고리즘이 인사이트를 학습함으로써 유사한 속성에 대한 배치 작업이 사람의 개입 없이도 자동으로 일치할 수 있도록 합니다.

++ Tamr의 운영자는 조직 내의 전문가가 이메일과 같은 시스템 또는 Out-of-Band 메커니즘에 대한 직접 로그인을 통해 Tamr의 매칭을 검증하거나 무효화할 수 있도록 합니다. 시스템을 사용하는 전문가 수의 허용치는 제한이 없습니다.

++ 그러면 전문가의 피드백이 즉각적으로 통합되거나 워크플로우를 통해 적절한 조치를 결정합니다.

++ 예를 들어, 두가지 특성이 일치하는지 여부를 평가할 때 사용자는 여러 전문가의 피드백을 반영 할 수 있습니다. 이러한 환경에서, 배정된 전문가 그룹들이 매칭가능성이 있는 속성들에 대해 가장 일반적으로 선택한 yes, no를 인식하도록 규정할 수 있습니다.

++ 전문가의 피드백을 통합한 후, Tamr는 향후 자동화된 사용을 위해 모델을 세분화합니다.

타깃 스키마를 통합하는 엔티티의 데이터셋는 TAMR를 통해 구체화되고 Downstream 기능에서 내보내거나 사용할 수 있습니다.

CLEAN

Tamr의 “Clean”단계는 휴먼 가이드 머신 러닝을 사용하여 통합된 데이터셋 내의 엔티티를 효율적이고 정확하게 파악하고 중복을 제거하도록 설계되었습니다. 일반적으로 엔터프라이즈급 데이터 시스템은 지저분하거나 중복된 데이터의 문제를 가지고 있고, 이 문제를 기존의 데이터 관리 기법으로는 해결하기 어려웠습니다. 이 단계의 핵심 기능은 사용자가 통합 데이터셋 내에서 중복된 데이터를 식별하고 내부에 포함된 레코드를 파악할 수 있도록 허용함으로써 정확한 결과를 얻는 것입니다.



TAMR는 완벽하게 정리되지 않은 통합 데이터셋으로부터 작업을 시작하므로, 다량의 중복레코드를 제거할 수 있습니다. 중복 제거 레코드식별에 대한 학습이 가능한 경우, TAMR는 해당 학습을 모델에 적용합니다. “Connect”단계와 유사한 프로세스에서 TAMR는 동일한 엔티티와 연관 가능성이 높은 레코드를 통합 데이터셋으로 군집화(Clustering)하기 위해 휴먼 가이드 머신러닝을 적용합니다.



++ 학습 데이터가 사전에 제공되지 않을 경우, TAMR는 비지도 학습(Un-supervised machine learning : 데이터 자체를 분석하나 군집화하면서 학습)을 통해 한 쌍의 레코드에 대한 모든 속성과 속성 값을 분석하여 유사성을 탐지합니다.

++ TAMR는 모델링에 따라 어떤 레코드가 중복 될 수 있는지 제안하고 데이터셋 내의 다른 잠재적 레코드 쌍에 대해 간단하고 중요한 질문을 생성합니다.



++ 예를 들어, TAMR는 “J.Smith”라는 레코드를 “John S” 레코드와 동일하게 인식합니다. 그런 다음, 조직의 전문가에게 이들이 다른 레코드인지, 동일한 레코드 인지 피드백을 요청합니다. “Connection” 단계와 마찬가지로 이는 프로세스의 정확성을 향상시켜 신뢰할 수 있는 결과를 보장하고, TAMR의 알고리즘이 통찰력을 학습할 수 있게 하여 차후에 유사한 작업이 자동으로 진행될 수 있게 해 줍니다.

Responses	address	customerID	email	firstName
⚡	375 Dwight Ter, Hampton, VA 23668	1017970		Kathleen
⚡	375 Dwight Terrace, Hampton, VA 23668	289929	ccoxph@who.int	Kitty
⚡	305 Glendale Trl, Los Angeles, CA 90030	1003298	jgraham6t@bigca	Jac
⚡	305 Glendale Trail, Los Angeles, CA 90030	275257		Johnny
⚡	93 Barnett Plaza, Fresno, CA 93786	1021662		Ben
⚡		293621	bevansgx@cdc.go	Benjamin
⚡		1068275		Karen
⚡		340234	kjones66@woothr	
⚡	10 Ruskin Park, Lansing, MI 48919	3419		Tina

▶ TAMR는 휴먼 가이드 머신 러닝을 사용하여 통합된 데이터셋 내의 엔티티를 효율적이고 정확하게 파악하고 중복을 제거합니다.

++ TAMR의 관리운영자는 조직 내에서 관련 전문가를 통하여 TAMR의 매칭을 확인하거나 무효화 하고 전문가의 피드백을 반영하여 사전에 정의한 워크 플로우를 통해 적절한 조치를 결정합니다.

++ 전문가의 피드백을 통합한 TAMR는, 차후에 더 많은 자동화된 사용을 위해 모델을 개선합니다.

++ 속성의 가장 일반적인 값

++ “신뢰할 수 있는 소스” 데이터셋의 값 선택 (i.e. 사용자가 가장 신뢰할 만한 것으로 알고 있는 데이터셋)

++ 전문가가 특정 속성에 대해 선택한 적절한 값

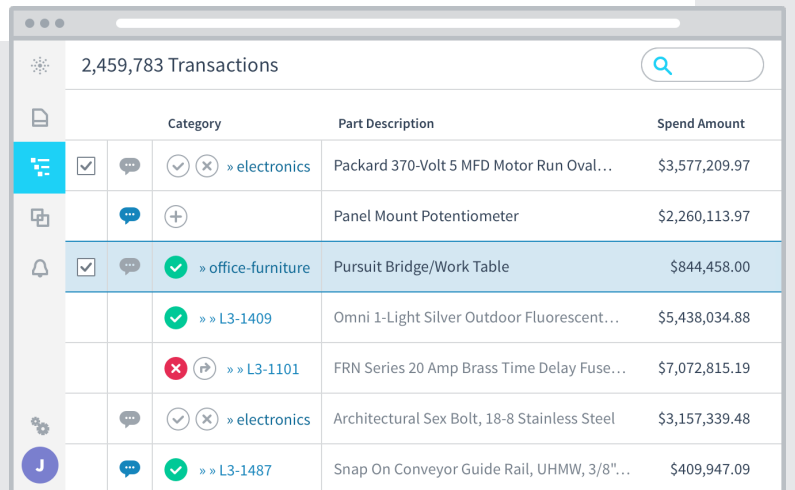
일부 다운스트림 사용 사례의 경우, 레코드 군집화를 사용하기에 충분합니다. 또는 조직이 레코드 그룹을 반영하여 단일 레코드로 병합하는 것을 선호할 수 있습니다. TAMR의 병합 로직은 강력하며 다음과 같은 클러스터 레코드 내의 특성에 대한 값을 선택하는 옵션을 포함합니다

논리적 엔티티가 정리되고 통합된 데이터셋이 구체화 되어, TAMR를 통해 Downstream으로 내보내거나 사용할 수 있습니다.

CLASSIFY

TAMR에 의해 특정한 엔티티가 통합된 데이터셋이 생성된 후에, 사용자는 더 심층적인 분석 기능을 위해 특정한 사용자 또는 일반적으로 사용되는 분류 체계로 레코드를 분류할 수 있습니다. 이는 공급 사슬(Supply Chain)이나 조달(Procurement) 분석과 같은 사례에서 사용 될 수 있습니다. 분류작업은 사업 및 분석 목적을 위해 논리적 그룹으로 엔티티를 구성할 수 있습니다. TAMR의 분류 단계는 “Connect” 및 “Clean” 단계와 동일한 방식으로 작동하며, 이를 통해 사용자에게 가장 심층적인 수준의 분류체계를 정확하게 제공합니다.

TAMR는 깨끗하게 통합되어있지만 아직 분류되어 있지 않은 논리적 엔티티의 데이터셋으로 시작합니다. 분류 프로세스의 첫 번째 단계는 포함된 각 분기와 관련된 엔티티의 샘플 레코드를 사용하여 대상 분류를 TAMR에 로드하는 것입니다. 여기서 TAMR은 분류의 각 분기와 관련된 단어/토큰을 식별 할 수 있습니다.



2,459,783 Transactions			
	Category	Part Description	Spend Amount
☒	» electronics	Packard 370-Volt 5 MFD Motor Run Oval...	\$3,577,209.97
☒		Panel Mount Potentiometer	\$2,260,113.97
☒	» office-furniture	Pursuit Bridge/Work Table	\$844,458.00
☒	» L3-1409	Omni 1-Light Silver Outdoor Fluorescent...	\$5,438,034.88
☒	» L3-1101	FRN Series 20 Amp Brass Time Delay Fuse...	\$7,072,815.19
☒	» electronics	Architectural Sex Bolt, 18-8 Stainless Steel	\$3,157,339.48
☒	» L3-1487	Snap On Conveyor Guide Rail, UHMW, 3/8"...	\$409,947.09

▶ TAMR는 심층적인 분석 기능을 위해 다양한 분류 체계로 레코드를 분류합니다

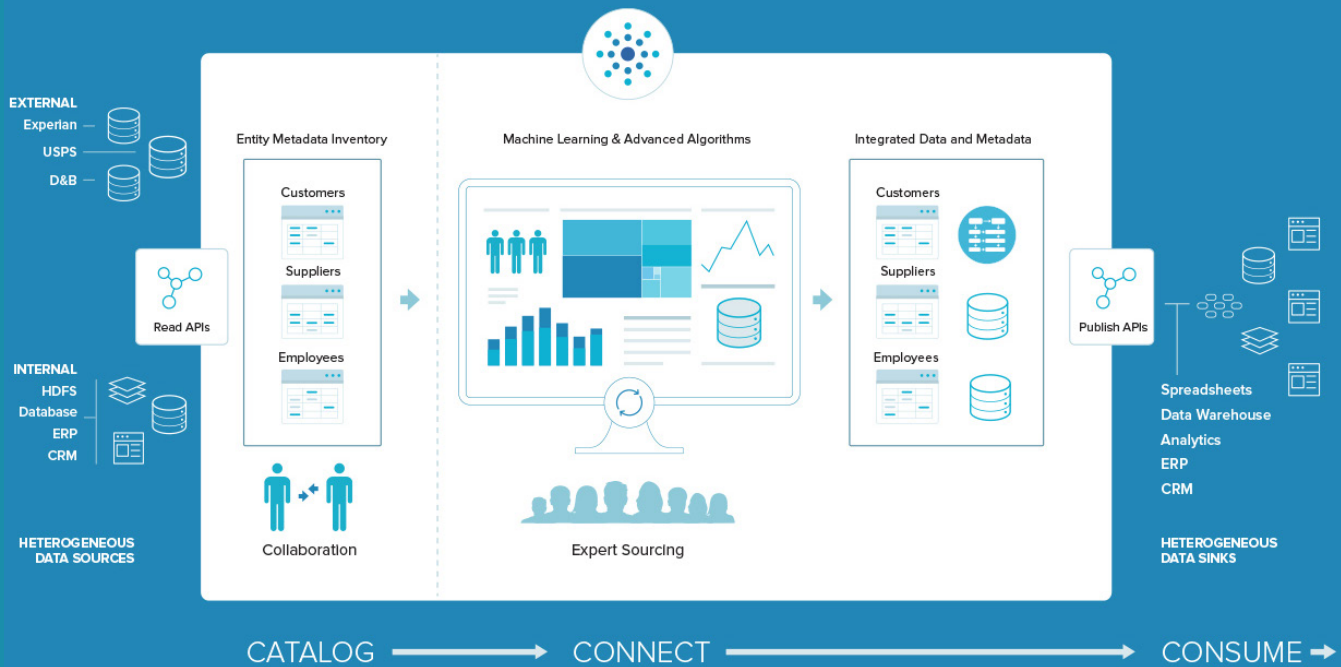
++ TAMR는 통합 데이터셋의 각 레코드 내에서 값을 분석하고 머신 러닝 알고리즘을 사용하여 각 데이터 셋 레코드에 포함된 단어와 각각의 카테고리 분류된 관련된 단어와의 일치율을 확인합니다. 이를 통해 TAMR는 초기 모델을 기반으로 각 레코드에 대한 분류를 정확하게 제안 할 수 있습니다.

++ TAMR는 통일된 데이터 셋 레코드의 상당부분을 대표하는 특정 레코드가 적절하게 분류되었는지 확인하기 위해 간단하고 중요한 질문을 제시합니다. 예를 들어 TAMR는 “1인치 터빈 볼트”와 관련된 기록이 실제로 조직 분류 내의 “볼트” 범주의 일부인지 판단하여, 판단의 신뢰도가 낮으면 조직 내의 전문가에게 의견을 요청합니다 -이 과정은 정확도를 높이고 향후 자동화를 개선 할 수 있습니다.

++ 전자 메일과 같은 직접 로그인 또는 Out-of-Band(ID검증과 같은 소프트웨어에 액세스하는 주요 방법)를 통한 전문가 피드백은 고객사의 피드백 워크 플로우에 따라 데이터 집합과 TAMR의 모델에 통합됩니다.

TAMR의 워크 플로우를 통해 레코드가 분류되면 TAMR은 각 레코드가 분류된 방법을 나타내는 새로운 필드를 각 통합된 데이터셋 레코드에 추가합니다. 예를 들어, TAMR가 레코드를 조직의 분류 체계의 네 번째 수준으로 분류하는 경우 각 레코드에 4개의 필드를 추가하여 분류 방법을 나타냅니다. 분류 후에는 연결(connected), 정리(cleaned) 및 분류된(classified) 데이터셋을 사용할 수 있습니다. TAMR는 분석 도구, 운영 데이터베이스 또는 단순한 Excel/CSV파일을 API를 통해 유연하게 사용할 수 있습니다.

제품구조



TAMR 플랫폼은 최근 향상된 유연성, 확장성, 관리의 편의성을 활용 하도록 설계되었습니다. 응용 프로그램의 계층은 다양한 기능을 제공 하는 마이크로 서비스의 결합으로 구성되어 있으며, 응용 프로그램 배 포 방식과 개별 응용 프로그램 구성 요소의 확장성에 유연성을 제공합니다.

데이터 처리 계층은 대용량 데이터 처리와 빠른 검색 및 필터링을 제공 하기 위해 확장성이 뛰어난 구성 요소들을 결합합니다. 전반적인 시스템 스케일은 트라이얼을 위한 단일 서버로 축소되고, 당면한 대규모의 생산 과제를 해결하기 위해 멀티 노드 데이터 레이크 인프라(multi-node data lake infrastructure)로 확장됩니다.

응용 프로그램을 구성하는 개별 마이크로 서비스는 개별로 실행되는 단일 응용 프로그램과 같은 모양과 느낌을 최종 사용자에게 제공합니다. 또한 사용자 인증을 위한 LDAP 및 사용자 기본 설정 등을 저장하

기 위한 관계형 데이터베이스와 같은 조직적 백업 서비스와 상호 작용합니다. 데이터 자체는 분산되고 안정적인 스토리지를 위한 HDFS, 분산 메모리 내 계산을 위한 Spark와 같은 기술을 활용하는 다중 노드, 스케일 아웃 인프라 내에 있습니다.

마지막으로, Elastic search 클러스터는 쿼리 대기 시간과 페이지로 드 시간을 짧게 유지하면서 풍부한 대화식 프론트엔드(front-end)를 제공하므로 사용자가 데이터에서 통찰력을 얻을 수 있도록 집중 할 수 있습니다.

MARKET POSITIONING

TAMR는 기업의 데이터 환경에서 기록 시스템으로도, 참조 시스템으로도 운영될 수 있는 다양한 가능성을 지니고 있습니다. TAMR 플랫폼은 대부분의 빅데이터 환경 투자를 보완하기 위해, 기존의 환경에서 해결하지 못했던 “garbage 입력, garbage 출력” 문제의 해결을 위해 고안되었습니다.



Data Catalogs

TAMR는 데이터셋의 프로파일링에 대해 몇 가지 기능을 제공합니다. 그러나 특정한 엔티티와 관련된 데이터셋을 발견할 수 있고, 사용자 간의 안전한 공동 작업을 지원할 수 있는 데이터 카탈로그는 TAMR의 밸류 포지션(Value Position) 다운스트림을 강화합니다. 데이터 카탈로그 안에서 발견되고 분석되는 데이터는 TAMR에서 입력자료로 사용될 수 있습니다.



Master Data Management

TAMR는 사실상의 마스터 데이터 관리(MDM) 솔루션으로 배포 될 수 있지만, 기존 데이터 관리 솔루션을 보완하는 경우가 가장 많습니다. 특히 TAMR는 레코드 통합을 위한 참조 시스템으로 작동함으로써 MDM 시스템을 지원하여 작업을 보다 빠르고 확장성 있게 만듭니다.



ETL

마스터 데이터 관리(MDM)와 마찬가지로 TAMR는 ETL 솔루션에 대한 참조 시스템 역할을 합니다. 특히 TAMR는 특정 레코드가 실제로 동일한 엔터티를 참조하는 ETL 솔루션의 변환을 제안하며, 변환 실행 속도와 규모를 지원합니다.



Self-Service Data Preparation

TAMR는 셀프-서비스 데이터 준비 도구를 보완합니다. 이러한 도구들은 대부분 이미 연결되고 정리된 데이터셋을 대상으로 하며, 강력한 머신 러닝을 필요로 하지 않습니다. 그러나 셀프서비스 데이터 준비 도구는 TAMR 플랫폼의 다운스트림 기능인 개별 유저의 데이터 큐레이션(예: 원치 않는 레코드 제거)을 허용합니다.



Analytic Tools

TAMR는 분석 및 시각화 툴을 보완 하는데 이용될 수 있으며, 이는 TAMR의 소비 방식으로 이용될 수 있습니다. 통일되고 깨끗한 데이터셋은 분석 및 시각화 플랫폼에 아주 중요하고, TAMR는 분석을 수행하는 기관들의 딜레마인 “garbage입력, garbage출력”문제를 해결할 수 있기 때문입니다.

USE CASES

엔터프라이즈 데이터 통합에 대한 TAMR는 모든 분야의 회사에 적합하며 다양한 사례에 적용 할 수 있습니다. TAMR는 공급 업체 및 부품에서 제품 및 고객에 이르는 기업의 데이터를 중앙에서 관리 할 수 있게 해줍니다.

TOYOTA 고객 360° 통합 뷰 기능을 통해 세분화된 데이터 수집

고객 데이터에 대한 일관된 시각이 부족하여 혁신을 추진하거나 고객의 기대를 충족할 수 없었습니다. 이 문제를 해결하기 위해 Tamr를 통해 기업의 데이터를 통합했습니다.

Matt Stevens
Director of Information Systems



기술적인 과제

- + 30개 국가의 로컬 관리 고객 데이터를 효과적으로 통합
- + 국가 수준에서 데이터 수집 및 관리를 위한 유연성 유지
- + 새로운 소스를 신속하게 통합
- + 로컬 Migration 프로젝트 지원

Tamr를 통한 해결

- + 현재 125개의 소스가 통합되어 있으며, 프로젝트 종료 시점 기준 500개
- + 로컬 시스템에 대한 운영 중단은 없지만, 새로운 중앙 집중식 데이터도 사용가능
- + 평균 1~2주 소요
- + TME France CRM migration 6주만에 완료

비즈니스 과제

- + 국가 간을 이동하는 고객에게 더 나은 서비스 제공
- + 모든 고객 접점에 걸쳐 일관된 환경 제공
- + 고객의 요구 사항을 이해하고 예측하여 기대치를 초과 달성하는 능력 향상

Tamr를 통한 해결

- + CSR에 이제 Tamr-Mastered고객 데이터를 검색하는 단일 UI 제공.
- + 판매 / 서비스 시점에서 더 나은 고객 지식 제공
- + 새로운 분석 및 운영 활용 사례를 지원하는 고객에 대한 최초의 통합 뷰



민첩한 멀티 도메인 접근을 통해 5억달러 이상의 가치 창출 **GE**

Tamr의 기술이 GE의 규모에 맞게 수행한 변화를 통해 얻은 결과를 우리는 직접 보았습니다. 데이터 셋을 결합하는 비용과 복잡성이 대폭 감소하면 분석결과를 통해 이전에는 접근할 수 없었던 기회를 발견하게 됩니다.

Lisa Coca

Managing Director, GE Ventures

기술적인 과제

- + 공급 업체 : 75개 이상의 ERP시스템과 2M 공급 업체 레코드로 통합 뷰 구축
- + 부품 : 8개 BU에 걸친 구매 시스템에서 25M의 비일상적 부품
- + 인수 합병 : 취득한 기업의 데이터를 기존 마스터 뷰와 통합

Tamr를 통한 해결

- + 파일럿에서 글로벌 구축 데이터 파이프 라인까지 6개월 이내, 2M레코드를 700k로 통합
- + 25M에서 간소화된 6.4M 고유 부품, 5개의 분석 가능한 분류로 데이터 통합
- + GE와의 3개 인수합병에서 얻은 데이터를 2주 이내에 통합

비즈니스 과제

- + 공급 업체 : GE를 모든 협상에서 주어진 공급 업체와 최상의 조건으로 협상
- + 부품 : 가장 비용 효과적인 공급 업체를 위한 소싱 전략 최적화
- + 기업 인수 합병 : 사후 취득 후 시너지 효과를 실현하는 속도를 높임

Tamr를 통한 해결

- + Tamr-mastered 통합 공급 업체 관점에서 1년차에 80만달러 절감
- + 연간 3억달러의 비용 절감 효과 파악(직접 비용의 0.5%감소)
- + 공급 업체, 구매 업체 및 고객 기반 기회를 신속하게 식별

THOMSON REUTERS 정보 기업의 데이터 큐레이션 운영 최적화

Tamr의 새로운 통합 플랫폼을 통해 자체적인 엔티티 통합 작업속도를 가속화하고 수동 작업을 40%이상 줄이는 상당한 성과를 거두었습니다.

Tim Baker

Global Head of Content Initiatives



**THOMSON
REUTERS®**

기술적인 과제

- + 데이터 처리 자동화 수준 향상
- + 데이터 관리자의 전문 지식 확보 및 활용
- + 데이터 품질 향상

Tamr를 통한 해결

- + 주요 프로젝트에서 최대 90%의 자동화
- + SME's knowledge 이 Tamr의 모델로 통합
- + 지속 가능한 정밀도 및 회수율 95%초과

비즈니스 과제

- + 출시 기간 단축
- + 수작업 감소
- + 클라우드 채택률 향상 지원
- + 이전에는 '너무 어려운' 것으로 간주되었던 데이터 통합 프로젝트 수행

Tamr를 통한 해결

- + 신제품 소개에서 수 개월 단축
- + 40% 수동 통합작업 축소
- + 하이브리드 사내 / 클라우드 구축
- + 이전에는드로잉 보드에 붙어 있던 새로운 TR 제품이 출시

P & D

S o l u t i o n



ADDRESS

14057 경기도 안양시 동안구
벌말로 126, 3010호
(관양동, 평촌오비즈타워)



CONTACT

T : 031 450 3011
F : 031 450 3012



EMAIL

tamr@pndsolution.com
www.pndsolution.com