



최신 분석 플랫폼의 구성 요소

작성자: Dan Kogan, Jen Underwood

‘최신 분석’이란 무엇이며 왜 필요합니까?

우리는 가속화된 혁신, 글로벌 경쟁의 증가, 그리고 유례 없는 변화와 재발견의 기회가 있는 격동의 시대에 살고 있습니다. 디지털 기술의 급속한 성장, 유비쿼터스 인터넷 연결성 및 스마트 기기의 확산 등이 데이터의 홍수를 일으켜 4차 산업 혁명이 도래하고 있으며, 이 모든 것을 빠르고 정확하게 이해할 수 있는 사람이 경쟁 우위를 지니게 됩니다.

최신 기술의 잠재력은 과소평가되어서는 안 됩니다. 시장 진입 장벽이 낮아지고 완전히 새로운 비즈니스 모델이 생겨나면서 실제와 가상 세계 간의 분명했던 경계가 모호해지고 있습니다. 모든 산업, 십여 년 전에는 그 자신이 혁신을 가져왔던 산업조차도 혁신의 대상이 되고 있습니다. 중소기업 용자에 대한 시민 기금, 더욱 광대해진 온라인 소매업 시장, 기존 유선 TV 채널에서의 미디어 서비스 구독 등의 영향을 생각해 보십시오. 우리 주위 어디에서나 이러한 변화가 가속화되고 있습니다. 자동화된 프로세스, 메신저 봇, 인공지능은 불확실하나 흥미진진한 미래로 가는 길에 초석이 되는 새로운 기술 중 몇몇 예에 불과합니다.



변화하는 BI의 역할을 가장 포괄적으로 잘 나타내는 용어에 대해 자세히 알아보려면, 오늘날의 분석에 관련된 주요 용어를 소개하는 **‘분석의 정의’** 백서를 참조하십시오.

그런데 이러한 모든 동향에 얽혀 있는 하나의 공통된 실마리가 있습니다. 바로 대량의 데이터입니다.

데이터가 원자재라면 분석은 그것을 유용한 정보로 정제하는 과정, 궁극적으로는 비즈니스에 경쟁력을 가져오는 과정입니다. 데이터는 오늘날 그 어느 때보다 필수적입니다. 비즈니스가 변화하는 산업에 발맞추어 진화함에 따라, 신속하고 스마트한 결정을 내리는 데 지속적으로 증가하고 있는 최신의 정확한 데이터에 의존해야 합니다. 그러나 전통적으로 분석 기술은 데이터 형식의 변화와 증가에 번잡하고 느리게 대응해 왔습니다. 매주 새로운 유형의 데이터베이스, 새로운 기기들에서 생성되는 데이터는 모두 완전히 새로운 기술에서 만들어진 것입니다. 분석은 이 속도를 따라가지 못했습니다.



엔터프라이즈 분석에 대한 최신 접근 방식을 구현하기 위해 IT와 비즈니스가 어떻게 협력해야 하는지에 대해 자세히 알아보려면, **분석 문화를 구축하는 방법** 백서를 참조하십시오.

디지털 시대인 오늘날, 현재 비즈니스 당면 과제를 정확성과 속도로 상대할 준비가 되어 있는 데이터 기반 문화를 조성하려면, 조직은 새로운 기술에 투자해야 할 뿐만 아니라 변화를 이끄는 사람들을 포함하여 정보를 제공하는 새로운 방식에도 투자해야 합니다. 이러한 평행적인 문화 이동이 바로 데이터를 수집하고 발굴해내며 또한 정제하고 필요에 맞는 적절한 정보를 제공하기 위해 협력하는 두 파트너인 IT와 비즈니스 간의 관계에서 기본적인 변화입니다. IT와 비즈니스가 협력할 때에 조직이 셀프 서비스 분석 문화라는 결코 이룰 수 없던 꿈을 실현할 수 있게 됩니다.

이 백서에서는 IT와 비즈니스가 회사 전반에 데이터, 가치 및 의사 결정을 제공하기 위해 함께 사용할 수 있는 최신 분석 플랫폼을 이루는 구성 요소를 살펴볼 것입니다. 여기에는 오늘날의 최신 도구는 물론 수십 년간 비즈니스 인텔리전스의 기반이 되어 온 기존 도구도 포함됩니다. 각각의 구성 요소가 데이터를 인사이트로 바꾸는 더 큰 프로세스에 어떻게 맞는지, 데이터를 캡처하고 보고하기 위해 필요한 것이 도구인지, 아니면 대화형 인사이트를 공유하는 새로운 기술인지 보게 될 것입니다. 또한 Tableau가 어떻게 최신 분석 플랫폼의 기반인 동시에 새로운 데이터 기반 분석 문화의 기폭제가 될 수 있는지 논의해 보겠습니다.



IT 조직이 셀프 서비스 분석의 변혁을 주도할 때, 규모에 맞는 거버넌스와 보안을 보장할 수 있습니다. 데이터 기반의 뛰어난 대응력을 갖는 비즈니스로 역량을 강화함으로써, IT는 비즈니스의 신뢰받는 파트너가 됩니다.

- COLIN REES, CIO, DOMINO'S PIZZA 그룹

목차

최신 분석이란 무엇이며 왜 필요합니까?	2
1. 오늘날 당면한 3가지 데이터 과제	4
데이터는 어디에나 있습니다	4
모든 사람이 데이터를 필요로 합니다	5
데이터는 항상 변하고 있습니다	5
2. 최신 분석 플랫폼의 구성 요소	6
IT의 지원	7
작성 및 사용	8
3. 종합 정리	11
4. 부록	11
스트림 수집	11
통합 허브 조정	12
구조화되지 않은 데이터, NoSQL 및 데이터 레이크에 대한 추가 정보	12
서비스형 데이터(DaaS)	13
논리적 데이터 웨어하우스	13
기계 학습	14
자연어	15
권장 데이터 검색	15
검색	15
알림	15
스토리텔링	15
Tableau 소개	17

1. 오늘날 당면한 3가지 데이터 과제

데이터는 어디에나 있으며, 매 순간 많은 양이 생성되고 있습니다. 예를 들면 칫솔은 이제 하나의 스마트 기기가 되어, 언제 칫솔을 사용하는지, 사용 시간은 얼마인지, 그리고 내부 부품의 상태를 기록할 수 있습니다. 또한 이 모든 정보를 담당 치과의사한테 보낼 수도 있습니다. 간단한 일이 이제 수천 개의 데이터 요소가 되었습니다. 이 하나의 예를 수백만의 기기에 대응시켜 보면, 사용자는 불과 몇 년 전에는 존재하지 않았던, 데이터의 산을 만들어내는 하나의 소규모 산업을 가진 것입니다. 이벤트 로깅, API, 소셜 미디어, 웹 사이트 추적, 다른 인터넷 기술 호스트 등에서 매 순간 데이터가 넘쳐나고 있습니다.

이러한 현대 환경에서 다음과 같은 3가지 비즈니스 과제를 보게 됩니다.

1. 데이터는 어디에나 있습니다
2. 모든 사람이 데이터를 필요로 합니다
3. 데이터는 항상 변하고 있습니다

데이터는 어디에나 있습니다

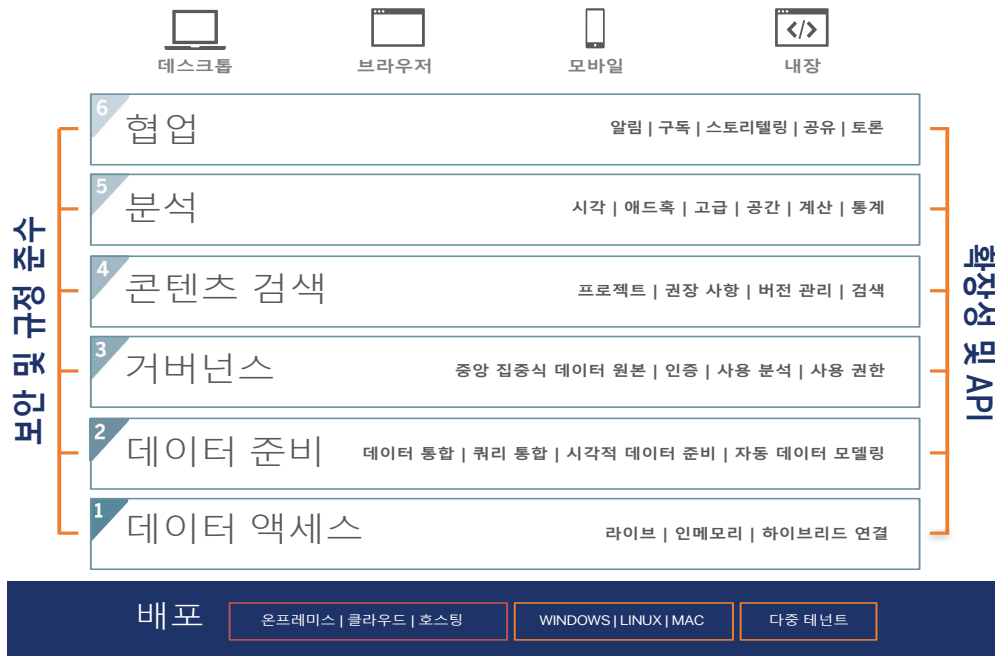


그림 1 데이터가 어디에 있든지 연결 가능한 Tableau

과거에는 모든 조직에서 데이터를 온프레미스로 보유하고 있었습니다. 조직에서는 생성된 모든 데이터를 잘 정의된 데이터 웨어하우스에 저장하고, 이를 원활하게 제어하기 위해 공을 들였습니다. 캡처하지 않은 데이터가 있다면 그것은 중요하지 않았고 아마도 신경 쓸 필요가 없는 것이었습니다.



클라우드 마이그레이션의 현재 동향에 대해 자세히 알아보려면 **클라우드 데이터 개요**를 살펴보세요.

그러한 사고방식은 웹 사이트, 모바일 기기 및 클라우드 애플리케이션이 전부 데이터를 조직 밖에서 생성하고 있는 오늘날에는 비즈니스의 중단을 초래할 수 있습니다. Google Analytics, Splunk, ServiceNow, Salesforce 등이 현재 동향의 몇몇 예가 될 것입니다. 외부 제공업체는 더욱 많은 유용한 데이터를 클라우드에서 생성할 것이며, 이러한 동향은 가속화될 것입니다. 이 클라우드로의 편향성에 한층 더해, 온프레미스 인프라조차 클라우드로 이전하는 조직도 있습니다.

모든 사람이 데이터를 필요로 합니다

지난 이십 년간 시장이 디지털 비즈니스와 클라우드로 이동하는 것을 보았습니다. 이것이 현대 분석 혁명의 핵심 요소 중 하나입니다. 다른 하나는 셀프 서비스 비즈니스 인텔리전스를 통한 데이터 기반 문화로의 이동입니다. 분석 문화는 질문이 있는 비즈니스 사용자가 그 질문에 대한 답을 스스로 발견해 낼 수 있는 오늘날 가장 혁신적인 회사들에 스며들고 있습니다. 이는 조직이 데이터를 제련하여 디지털 황금인 정보로 빠르게 변환하는 전문가로 바뀌게 하고 있습니다. 분석 문화를 완전히 조성하려면, 조직은 단언컨대 조직의 가장 가치 있는 두 자산인 직원과 데이터를 같이 놓고, 모든 사람에게 올바른 데이터로의 액세스를 부여하며, 탐색하고 협업하는 것을 장려해야 합니다.

분석에 대한 최신 접근 방식은 IT와 비즈니스가 서로 협력하는 것입니다. IT는 비즈니스 사용자가 신뢰할 수 있는 데이터 및 콘텐츠를 찾을 수 있는 중앙 집중된 환경을 제공하고, 모든 사람이 그것을 안전하게 사용하고, 질문, 실험, 의사 결정을 생각하는 즉시 할 수 있도록 해줍니다. 이것은 비즈니스 전문가가 유연한 민첩성을 제공하고 지속적으로 프로세스 개선을 촉진하는 메타데이터, 비즈니스 규칙 및 보고 모델을 만드는 상향식 방법론입니다.



처음 Tableau를 시작했을 때는 대시보드와 보고서를 만드는 것만 생각했습니다. Tableau가 조직의 DNA를 근본적으로 변화시킬 것이라고 결코 생각하지 못했습니다. 중요한 것은 솔루션이나 기술이 아니라 데이터에 대한 문화가 변화된 방식입니다.

- ASHISH BRAGANZA, 글로벌 비즈니스 인텔리전스 담당 이사

데이터는 항상 변하고 있습니다

우리가 알고 있는 '변하지 않는' 한 가지 사실은 '모든 것이 변한다'는 것입니다. 최신 분석 플랫폼은 유연성에 우선순위를 둡니다. 유연성이란 데이터를 플랫폼 간에 이동하고, 필요에 따라 인프라를 조정하고, 새로운 데이터 유형의 이점을 활용하고, 새로운 사용 사례를 지원하는 기능을 말합니다. 뿐만 아니라, 매일 데이터 분석의 새로운 기술이 세상에 등장하는 것처럼 보입니다. 기계 학습, 음성 인식 비서 프로그램, 자연어 쿼리 등의 기술을 예로 들 수 있습니다. 어떤 기술은 적어도 지금은 실용적이기 보다는 허구에 가깝게 보일 수 있지만, 새로운 방법 및 기술은 발전하여 그 가치를 고객과 회사에 증명해 보이게 될 것입니다.

급속히 진화하는 데이터 세계에서 인프라 확장 요구와 새로운 기술 모두에 있어 유연성이 가장 중요합니다. 유연성은 경쟁 우위를 확보하고 유지하는 데 꼭 필요합니다. 미래 지향적인 분석 아키텍처를 고려할 때, 미래의 민첩성을 크게 저하시킬 수 있는 독점 아키텍처 구현 공급 업체에 대한 종속성은 피해야 합니다.

2. 최신 분석 플랫폼의 구성 요소

오늘날 비즈니스가 직면하고 있는 세 가지 과제는 보이는 것만큼 극복하기 어려운 것은 아닙니다. 오늘날의 진화하는 비즈니스 세계에서 데이터가 공통된 문제라면, 최신 분석 플랫폼이 잠재력을 열어주는 열쇠입니다. 그러나 최신 분석 플랫폼은 인프라의 한 부분이지 단일 요소가 아니며, 여러 개의 독립적인 구성 요소로 구성되어 있습니다. 이들 중 일부는 데이터 웨어하우스 같이 최근에 활용되는 기존 비즈니스 인텔리전스 요소가 있고, 시각적 분석 같이 처음부터 비즈니스가 데이터 분석에 접근하는 방식을 혁신한 완전히 새로운 개념이 있습니다. 이들 모두는 오늘날의 비즈니스 과제에 대응하도록 도울 수 있는 분석 플랫폼의 두 가지 구성 요소입니다.

최신 분석 플랫폼은 다음과 같이 양분해 볼 수 있습니다.

- **IT의 지원:** 데이터의 수집, 정리 및 준비가 포함됩니다.
- **작성 및 사용:** 데이터 분석과 인사이트를 올바른 이해 관계자에게 전달하는 것을 포함합니다.

기존에는 두 측면 모두 IT 영역에서 단일 프로세스로 결합되었습니다. 이제 IT가 지원하는 첫 번째 측면인 데이터 원본의 작성과 처리를 살펴봅시다. 두 번째 측면인 분석과 제공은 여전히 IT에서 지원하지만 비즈니스 사용자가 직접 실행하고 있습니다.

이러한 두 측면의 조합이 비즈니스와 IT간의 진정한 파트너십이며, 신속하게 데이터 기반 의사 결정을 하려는 모든 조직이 실행할 수 있는 현대적 방법입니다. 이원화된 BI라고도 부르는 이 방식은 최신 분석의 셀프 서비스 메커니즘을 채용하면서 기존 BI와 운영 보고의 장점만을 취한 것입니다.

이러한 IT와 비즈니스 간의 관계에서 IT는 데이터 아키텍처를 설계하고 적절한 데이터 보안 및 액세스 제어를 제공합니다. 그러면 비즈니스 전문가가 필요한 분석 자산을 필요할 때에 작성합니다. 그 결과 IT는 모든 사람이 그들 자신의 필수적인 질문에 효율적으로 대답할 수 있게 하며, 비즈니스 사용자는 그 질문이 나온 순간에 답을 찾게 되어, 현대 비즈니스 상황의 당면 과제에 맞설 준비가 된 민첩한 조직을 만들어 냅니다.

최신 분석 플랫폼의 각 부문을 이루는 다양한 구성 요소 중에서 몇몇 상위 동향, 그리고 기억해야 할 몇몇 중요한 개념을 개략적으로 살펴보겠습니다. 특정 구성 요소와 그것을 제공하는 시장 주도 공급 업체에 대한 세부 정보와 그것의 적용 여부에 대해 살펴하려면 **부록**에서 각 구성 요소의 해당 섹션을 참조하십시오.



그림 2 기본 구성 요소

IT의 지원

기존 IT 주도 비즈니스 인텔리전스와는 달리, 오늘날의 가장 효율적인 IT 조직은 사용자와 전문가가 작성하고 사용할 수 있도록 데이터를 조정, 정리 및 통합함으로써 분석 데이터 원본을 사용 가능하게 하는 데 중점을 둡니다. 역할은 분명하나, 그렇다고 과대 평가해서는 안됩니다. 다른 사람이 쓸 수 있도록 데이터를 수집하고, 원본을 관리하고, 데이터를 처리하는 작업은 항상 비즈니스 인텔리전스에 필수적이었고 여전히 최신 분석 플랫폼의 핵심으로 자리합니다. 정제할 원자재가 없다면 거기서 무슨 인사이트를 발견할 수 있겠습니까?

최신 분석 플랫폼이 갖는 차이점은 바로 비즈니스와 IT 간의 파트너십입니다. 비즈니스 사용자가 스스로 데이터를 분석할 도구가 있을 때, 데이터 자체를 신뢰할 수 있다고 인식한다면 즉석에서 질문에 대한 답을 찾을 것입니다. 이것은 정확하고 신속한 보고서와 대시보드로 이어집니다. 또한 IT는 대시보드와 변경 요청에서 벗어나, 마침내 데이터 자체에 우선순위를 둘 수 있게 됩니다. 즉, 데이터 거버넌스 및 보안을 지키고, 데이터의 정확성을 보장하고, 데이터 수집, 처리 및 저장의 가장 효율적인 진행 단계를 확립할 것입니다.

데이터 우선순위 지정에 더 적절한 시기는 없습니다. 비즈니스는 규모에 관계 없이 이미 데이터를 수집하고 있으며, 분명히 그 중 일부를 분석하고 있을 것입니다. 나머지는 **다크 데이터**입니다. 데이터를 수집할 수 있는 곳이 수십 억 곳에 이르고, 가능한 한 많이 수집할 수 있도록 돕는 더 많은 도구가 시장에 나오고 있습니다. 많은 데이터 양, 데이터 위치, 다양한 데이터 원본 유형 등에 따른 다양한 특성을 다루기 위해 사용되는 기술에는 여러 가지가 있습니다. 모든 조직은 제각각 다르므로, 어떤 구성 요소가 현재와 미래에 적용하기에 가장 적합할지 충분히 시간을 갖고 우선순위를 결정해야 합니다.

다음 사항을 고려해 보십시오.

데이터 스트리밍, 서비스형 데이터(DaaS) 등 특정 기술과 특정 공급 업체 선택에 대한 세부 정보는 아래 **부록**을 참조하십시오.

데이터베이스와 데이터 웨어하우스

데이터베이스와 데이터 웨어하우스는 수십 년간 비즈니스 인텔리전스의 기반이 되어왔습니다. 이들 중 다수는 시장 지위를 상실한 반면, 일부는 계속하여 최신 분석 아키텍처로서 자리매김하고 있습니다.

가장 오래된 데이터베이스의 하나는 OLAP(온라인 분석 처리)로 알려져 있습니다. 이들은 느린 데이터베이스 기술에 대한 대응으로 시작되어 예측 가능한 쿼리에 대해 집계 및 캐시로 응답 시간을 줄이는 데에 사용되었습니다. 그러나 회사의 질문이 더 복잡해지고 예측하기 어려워짐에 따라 OLAP는 유지되지 못하고 종종 완전히 새로운 집계기 처음부터 다시 작성되어야 했습니다. 또한 보다 개선된 데이터베이스 기술에도 불구하고 연관성을 잃어가고 있습니다.

오늘날의 데이터베이스는 인메모리, 대규모 병렬 처리(MPP) 등의 고급 컴퓨팅 기술을 활용하고 있습니다. 이를 통해 데이터베이스는 데이터 저장, 하드웨어 메모리 사용과 때때로 내장 계산 및 데이터 과학 기능까지 최적화하는 동시에 선형 확장성을 통해 매우 빠른 성능을 제공합니다.

뿐만 아니라, 클라우드의 도래는 온프레미스 버전은 맞지 않는 데이터베이스 기술에 새로운 활력을 가져왔습니다. 여기에는 하드웨어 구매나 인프라 관리 팀을 필요로 하지 않고 변화하는 비즈니스 요구에 따라 탄력적으로 규모를 확장할 수 있는 기능이 포함됩니다.

최신 분석 아키텍처에는 언제나 데이터베이스와 데이터 웨어하우스의 자리가 있을 것이며, 엔터프라이즈 전반의 셀프 서비스 보고를 위한 거버넌스, 정확성 및 적절한 차원형 데이터의 제공에 있어 지속적으로 중요한 역할을 수행할 것입니다. Hadoop, 데이터 레이크 등의 다른 기술을 채용한 회사도 보통 관계형 데이터베이스를 혼합된 데이터 원본의 일부로 유지합니다.

NoSQL, 구조화되지 않은 데이터 및 데이터 레이크

데이터베이스와 데이터 웨어하우스는 예측 가능한 원본과 형식에서 오는 데이터의 분석을 지원할 때 강력한 성능을 발휘합니다. 그런데, 모든 데이터가 예측 가능한 것은 아닙니다. 최신 분석 아키텍처에서 NoSQL 데이터베이스는 잘 정의된 스키마나 형식이 없는 데이터 원본을 포함하여 모든 원본의 데이터를 빨리 로드할 수 있다는 장점 때문에 모든 조직에서 전략의 일부가 되고 있습니다. Non-SQL 또는 Not-Only-SQL이라고도 불리는 NoSQL 데이터베이스는 기존 관계형 데이터베이스와 비교하여 열, 문서, 키 값 및 그래프 스토리지 유형을 포함하는 데이터 스토리지의 대체 유형을 제공합니다.

구조화되지 않은 데이터와 관련하여 빅 데이터 및 데이터 레이크의 개념이 있습니다. 데이터는 어디에서든지 무작위로도 생성되므로, 그것을 수집하고 사용 가능한 형식으로 만드는 것은 불가능해 보일 수도 있습니다. 기술은 데이터를 먼저 특정 형식에 억지로 맞추는 대신, 분석 도구가 원시 데이터에 연결할 수 있게 하는 방향으로 발달해 왔습니다.

이러한 기술 중의 하나는 데이터 레이크라 부르는 광대한 양의 데이터를 구조화되어 있든지 아니든지 원래의 형식으로 보유할 수 있는 스토리지 저장소입니다. 그런 다음, 데이터를 특정 형식으로 바꾸는 사전 처리 없이 데이터를 신속하게 변환하는 API, SQL 등의 최적화된 처리 메커니즘을 사용하여 데이터를 분석합니다.

이러한 모든 도구는 생성되는 데이터의 양과 위치를 예측하기 어려운 사물 인터넷, 데이터 과학, 데이터 스트리밍, 기타 구조화되지 않은 분석 사용 사례와 관련된 프로젝트에 주로 쓰입니다.

NoSQL, Hadoop 및 데이터 레이크와 관련된 데이터 기술의 목록은 [부록](#)을 참조하십시오.

플랫 파일

Excel과 CSV 파일의 전성기는 오래 전에 끝났습니다. 소규모 조직이든 대규모 조직이든 간에 겉보기에 부적절해 보이는 이러한 플랫 파일은 계속해서 아마도 영원히 존재할 것입니다. 실제로 그 어느 때보다 많은 곳에 있습니다. 전에는 누군가의 컴퓨터에 물리적으로 존재했었습니다. 이제는 Google Drive나 Dropbox 같은 클라우드 저장 시스템에 존재합니다. 외부 조직에서 서비스형 데이터(DaaS)의 일부로 플랫 파일을 만들어 냅니다. 이러한 파일은 작성이 쉽기 때문에 다양한 데이터 필드, 추가적인 고객 조사 또는 기존 데이터 집합에 추가되는 일부 정보의 범례로서 유용합니다.

부가적으로, 올바른 시기에 플랫 파일에 대해 적절한 보안을 적용하십시오. 특히 일회성 시나리오에서 필요에 따라 사용하기를 권장합니다. 특정 파일의 사용이 증가하면, 적합한 사람들이 안전하게 액세스할 수 있도록 적절한 보안 프로토콜을 적용합니다.

작성 및 사용

최신 비즈니스 인텔리전스의 특징은 비즈니스 사용자가 비즈니스 인텔리전스 플랫폼을 사용하도록 한 것입니다. 의사 결정권자가 IT 부서에 보고서를 요청하고, 며칠을 기다려서 자신의 질문에 대한 답은 찾을 수 없는, 이미 최신이 아닌 보고서를 받던 시대는 지나갔습니다. 오늘날, 질문이 있는 의사 결정권자는 스스로 도구를 사용하여 답을 찾을 수 있습니다. IT가 전체 조직이 사용 가능한 데이터를 신뢰할 수 있도록 해 놓았으므로, 비즈니스 사용자는 어려움 없이 필요할 때마다 스마트한 데이터 기반 의사 결정을 내릴 수 있습니다.

여기서의 중요한 요소는 첫 날부터 바로 사용할 수 있는 실제 분석 도구입니다. 조직에 가장 적합한 구성 요소가 무엇인지 알아보기 전에, 전체 아키텍처 구축을 준비하면서 귀사의 데이터 원본이 유효하고 신속한 탐색이 가능한지 데이터 위치에 상관없이 모든 데이터에 연결해보아야 합니다.

데이터 분석에는 여러 구성 요소가 있지만, 핵심은 조직의 누구나 프로그래밍 경험에 상관없이 데이터 원본에 연결하여 인사이트를 끌어낼 수 있게 하는 수단인 시각적 분석입니다. 교사, 의사, 영업 사원의 손에 이러한 도구를 쥐어 준다면, 조직은 변경 요청과 분투하는 대신 기름칠한 기계처럼 잘 작동할 것입니다.

이제 비즈니스 인텔리전스는 인사이트를 다른 사람과 어떻게 공유할 것인지에 새로이 초점을 맞추고 있습니다. 이제 사람들은 대시보드와 보고서에만 그치지 않고 완전한 대화형 애플리케이션, 데이터가 연결된 긴 양식의 문서, 텍스트, 그림 또는 모바일 최적화된 뷰 환경까지 작성할 수 있습니다.

또한 비즈니스와 소규모 부서가 성장함에 따라 정보를 신속하게 공유하고, 데이터 원본을 발견하고, 대시보드를 최신 상태로 유지하며, 가장 중요한 메트릭을 따르는 스마트한 생산성 도구에 의지하게 됩니다.

이 섹션에서 데이터 전달을 강화하기 위해 업계 최고의 최신 분석 도구가 제공하는 몇 가지 구성 요소를 살펴봅니다. 스토리텔링, 알림 등에 대한 추가 정보는 [부록](#)을 참조하십시오.

시각적 분석

인간의 시각 시스템은 세상에서 가장 강력한 도구입니다. 그것이 오늘날 비로소 데이터 분석에 통합된 방식이 되었습니다. 패턴 인식에 기초하여 인간의 뇌는 이미 매일같이 상향 또는 하향 기울기의 추세, 불규칙한 활동의 급증, 또는 이상값을 보이는 특정 레코드 등의 데이터 패턴을 근접하게

밝혀낼 수 있는 시각적 분석을 사용하고 있습니다.

기존 스프레드시트에서는 데이터를 열과 행으로 분석하고 공유할 하위 집합을 선택한 후에 차트를 작성해야 합니다. 어색한 마법사나 텍스트 기반 명령어를 통해서 이 차트는 간혹 질문에 대한 답을 주기도 했고, 때로는 완전히 새로운 의문을 제기하기도 했는데, 늘 거기서 더 나아갈 수 없었습니다. 이와는 대조적으로, 시각적 분석은 시각적 요소를 프로세스의 일부로 만드는 간단한 드래그 앤 드롭 동작으로 우아하고 직관적인 분석 환경을 제공하며, 최종 결과로서 단순한 차트 이상을 제공합니다. 인사이트를 얻는 과정은 답을 얻는 것만큼이나 가치가 있습니다.

시각적 분석은 단지 보기 좋은 시각화 도구가 아닙니다. 시각적 분석은 데이터를 결합하게 해주는 언어이며, 이상 현상을 감지하고, 계산, 그룹화, 가정 조건 등을 통해 데이터를 보강하며, 이 모든 것을 프로그래밍의 도움 없이 가능하게 합니다.

기존 BI 및 보고

기존의 BI, 대시보드 및 보고가 아직 사용되고 있지만 오늘날 그 작성 프로세스는 달라 보입니다. 요약 대시보드, 재무 감사 등 많은 정적 보고서는 이전에 요청한 분석 쿼리를 통해 작성할 기술 개발 능력과 종종 기초 데이터 모델의 변경까지 필요로 했습니다. 이 모든 것을 개발하는 데는 수일, 수주 또는 수개월이 걸릴 수 있습니다.

최신 분석 플랫폼에서 이러한 대시보드 및 보고서의 다수는 질문의 성격으로 인해 IT 및 데이터 관리자가 구체화 및 검증하고, 궁극적으로 기존 정적 보고서를 대체하는 애드혹 질문으로 시작됩니다. 이 업데이트된 프로세스는 질문이 진화하고, 변화하며, 완전히 새로운 질문으로 이어지는 경우에도 적절한 답을 찾기 위해 데이터를 심도 있게 검색하는 과정에서 비즈니스 사용자의 전문성을 활용합니다. 최신 분석의 유연성을 통해 기존 보고 요구 사항이 지속되는 경우에도 기존 도구를 단순히 대체할 수 있습니다.

개인 데이터 준비 도구

ETL과 달리, 데이터 준비 도구는 IT 직원이 아닌 사용자가 데이터에 대해 강력하고 정밀한 조작을 할 수 있도록 설계된 경량 애플리케이션입니다. 데이터 준비 도구는 비즈니스 사용자가 일상적으로 데이터 집합을 결합하고, 조인을 자동화하고, 필드명을 바꿀 수 있게 하고, 데이터를 분석 준비가 되게 만든다는 목표 아래 데이터를 개선할 수 있게 하는 시각적 분석 도구가 갖는 사용 용이성, 속도 및 민첩성의 원칙에 따라 만들어집니다.

일단 데이터를 사용하기 시작해야 데이터를 어떻게 수정해야 할지 알게 되므로, 이러한 도구는 차후에 우선순위 목록에 추가되어야 할 것입니다. 그런데 조직의 운영에 있어 질문에 답하는 데 드는 시간 중 많은 부분이 데이터를 적절한 양식으로 바꾸는 데 할애됩니다. 개인 데이터 준비 도구는 IT가 제공하는 사전 정의된 보고의 의미론적 계층을 필히 포함하거나 배제하지 않으면서 시간 투자를 줄일 수 있는 강력한 방법입니다.



하버드 비즈니스 리뷰(Harvard Business Review)의 **'당신의 데이터 전략은 무엇입니까?(What's Your Data Strategy?)'** 라는 글에 따르면 "분석가가 쓰는 시간 중 80%는 단순히 데이터를 찾고 준비하는 데 소비됩니다."

Tableau

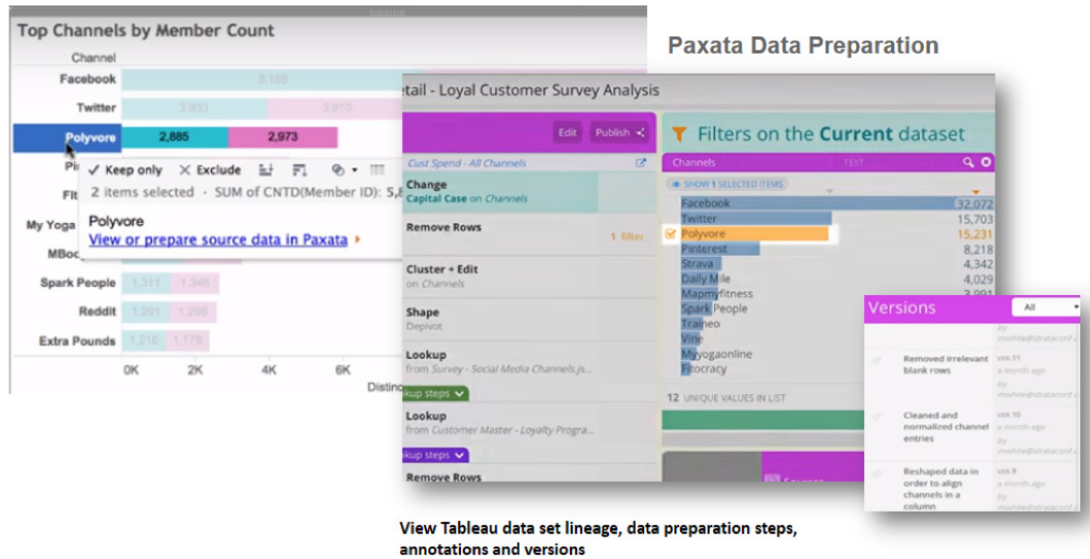


그림 3 Paxata의 개인 데이터 준비와 시각적 분석의 통합

고급 분석

고급 분석은 통계, 예측 알고리즘 및 기계 학습으로 대형 데이터 집합의 가치를 최대화하는 것에 힘을 쏟는 회사들에게 최신 분석 아키텍처의 핵심 구성 요소로 부상했습니다.

과거에는, R, Python, SPSS, SAS 등의 프로그램을 사용하는 숙련된 데이터 과학자만 고급 분석을 사용할 수 있었습니다. 오늘날, 시각적 분석 기능의 향상으로 고급 분석의 범위는 모든 셀프 서비스 분석 사용자에게 기본 제공되는 기능으로 확대되었습니다. 박스 플롯, 트리 맵, 기본 예측 모델링 기술과 같은 기능은 보편화되어 마우스 클릭만으로 쉽게 사용할 수 있습니다.

전용 통계 분석 도구에는 여전히 많은 사용 사례가 있습니다. 고객에게 "다음에 볼 내용"을 제공하기 위해 끊임 없이 진화하는 알고리즘을 만들려는 조직이나, 고객의 신용 카드에서 사기가 발견된 시기를 판단하는 기능을 작성하려는 조직은 특정 전용 도구를 개발하여 자신의 핵심 시각적 분석 도구 모음에 추가하고 싶을 수 있습니다. 이러한 도구를 배포하려면 필수 교육 및 도구별 프로그래밍 경험이 필요한데, 이 과정은 수개월이 걸릴 수 있습니다.

공유 및 협업

최신 BI 플랫폼에서는 인사이트의 공유, 협업 및 사회화가 핵심 기능입니다. 컨텍스트를 평가하는 것에서부터 차선책을 정하는 것까지 인사이트의 영향은 협업을 통해 극대화됩니다. 최신 분석 플랫폼은 토론 포럼, 주석, 해설, 즐겨찾기, 좋아요 수, 그리고 선도적인 생산성 및 포털 애플리케이션에서 채용한 기타 사회 개념을 제공합니다. 도구 자체나 분석 플랫폼에서 얻은 인사이트를 직접 전달할 수 있기 때문에 분석 흐름이 훨씬 단순해지고, 가치 있는 발견 내용에 대한 추가적인 탐색과 토론을 도모할 수 있습니다. 또한 내장 기능을 통해 외부 애플리케이션 및 포털로 협업이 확대될 수도 있습니다.

내장된 분석 기능

가장 강력하지만 자주 간과되는 비즈니스 개념 중 하나는 흐름입니다. 표준 운영 프로세스에서 데이터에 대한 답을 찾을 때 비즈니스 사용자를 분리하지 말고 인사이트를 이미 정립되어 있는 흐름에 그대로 삽입하여 기존 프로세스에 통합하십시오.

최신 분석을 사용하면, 데이터 및 대시보드를 회사 포털 또는 애플리케이션에 직접 내장하거나 생산성 도구에 쉽게 통합할 수 있습니다. 최고의 분석 플랫폼은 API, 소프트웨어 개발자 키트 및 유연한 전달 메커니즘을 통해 이러한 모든 시나리오를 지원하므로, 한 도구에서 다음 도구로 쉽게 전환하거나 모든 도구를 단일 포털 내에 결합할 수도 있습니다.

흐름은 또한 실제 위치로도 확장됩니다. 오늘날의 직원들은 방화벽 뒤의 사내 리소스에 액세스 없이 이동 중인 경우가 많습니다. 최신 분석 기술은 어디서나 모든 종류의 기기에서 데이터에 액세스할 수 있도록 지원합니다. 즉, 영업 사원은 노트북을 꺼내지 않고도 휴대기기에서 정보를 바탕으로 데이터 기반 의사 결정을 내릴 수 있습니다. 또한 건설 관리자가 VPN을 통해 기업 네트워크에 연결하지 않고도 현장에서 중요한 정보에 액세스할 수 있음을 의미합니다. 모바일 및 클라우드는 비즈니스 운영 방식을 영원히 바꿔 놓았으며, 진정한 최신 분석 플랫폼은 또한 비즈니스가 이러한 이점을 활용할 수 있도록 도와야 합니다.

3. 종합 정리

이러한 구성 요소는 비즈니스가 어떤 과제에도 정면으로 맞서 나갈 수 있도록 역량을 강화해 주는 최신 분석 플랫폼의 토대입니다. 비즈니스와 IT 간의 진정한 파트너십을 지원하는 이 플랫폼을 통해 회사의 모든 사람들은 의사 결정을 함에 있어 올바른 도구로 신뢰할 수 있는 데이터에 기반하여 의사 결정이 내려진다는 것을 확신할 수 있게 됩니다.

최신 분석 플랫폼을 전체적으로 한 번에 구현하는 것은 그 자체가 하나의 과제처럼 보일 수 있습니다. 다행히, 시작하기 전에 전체 환경을 완전히 구축할 필요는 없습니다. 오늘날 비즈니스가 가장 성공적이 되는 경우는 오히려 그렇지 않을 때입니다. 시작은 작게 하되, 궁극적으로 비즈니스가 다음에 투자해야 할 곳을 알고 점진적으로 변환해가는 것입니다. 회사에서 기술 솔루션을 부서에 광범위하게 적용하기 전에 파일럿 테스트로 시작해 볼 수 있습니다. 시작부터 전체 전략에 지원하는 모든 개별 구성 요소를 통합할 필요는 없습니다. 예를 들어, 전체 웨어하우스를 완료하기 전에 데이터 경로의 허점을 발견하는 데 시각적 분석 도구를 사용할 수 있습니다. 이렇게 하면, 분석에서 즉각적인 가치를 창출하고 데이터의 결함 및 오류를 찾을 수 있어서, 결과적으로 더 정확하고 기능적인 데이터 웨어하우스를 구축할 수 있습니다.

핵심은 이러한 유형의 점진적인 변화를 가능하게 하는 도구로 작업하는 것이며, 최신 분석 플랫폼은 한 번에 하나씩 배포할 수 있는 구성 요소의 모음입니다. 이를 통해 조직에서는 접근성, 민첩성을 향상하고 광범위한 데이터 원본에서 인사이트를 끌어낼 수 있습니다. 이것이 바로 정상급의 분석 리더가 하는 일입니다. 최고의 빅 데이터 분석, 사물 인터넷, 데이터 과학 솔루션과 함께 시각적 분석을 전략적으로 활용하는 것입니다.

예를 들어, Netflix는 포괄적인 빅 데이터 플랫폼과 데이터 레이크를 구축하여 비즈니스에서 생성되는 엄청난 양의 데이터를 지원하고 있습니다. Tableau는 S3, EMR, Spark와 같은 상이한 도구를 비즈니스를 지원하는 통합 분석 플랫폼에 결합할 수 있게 해주는 필수 구성 요소입니다.

디지털 비즈니스 전환 과정 어디에 있든지 상관없이 현재 갖고 있는 데이터를 사용하여 즉시 시작하는 것이 중요합니다. 비즈니스는 향후의 거대 시장 변화에 맞춰 더욱 신속하게 움직여야 할 것입니다. 최신 분석 구성 요소를 선택하여 비즈니스가 데이터 기반 의사 결정을 통해 혁신을 가져오는 주체가 될 수 있도록 역량을 강화하십시오.

4. 부록

많은 다양한 분석 기술 및 솔루션 옵션이 존재하며, 각각 특성화된 목적과 이점이 있습니다. 이 부록에서는 개요에서 다루지 않은 각각의 옵션을 자세히 살펴보겠습니다.

스트림 수집

스트림 데이터는 소셜 네트워크, 스마트 계량기, 홈 자동화, 비디오 게임, IoT 센서 등 연결된 모든 기기 및 앱에서 계속해서 생성됩니다. 때로는 이러한 데이터가 반구조화된 데이터 경로를 통해 수집되기도 합니다. 실시간 분석 및 예측 알고리즘을 스트림에 적용할 수 있지만, 일반적으로 스트림 데이터는 람다 아키텍처(Lambda Architecture)를 통해 원시 형식으로 전송/저장되거나 분석 용도로 Hadoop과 같은 데이터 레이크에 저장됩니다.

람다 아키텍처는 배치 및 스트림 처리 방법에서 각각의 장점을 살려 방대한 양의 데이터를 처리하도록 설계된 데이터 처리 아키텍처입니다. 이 아키텍처는 지연 시간, 처리량 및 오류 허용 문제의 균형을 맞추도록 설계되었습니다.

현재 스트리밍 데이터에는 Amazon Kinesis, Storm, Flume, Kafka, Informatica Vibe Data Stream과 같은 다양한 옵션이 있습니다.

통합 허브 조정

허브 앤 스포크(Hub-and-spoke) 통합 패턴은 이해하기 쉽고 광범위하게 사용되는 데이터 아키텍처 디자인입니다. 허브는 데이터 원본이 어디에 있든 이를 분리하고, 관리할 점 대 점 인터페이스의 수를 줄임으로써 더 유연한 통합을 가능하게 하는 것을 목표로 합니다. 통합 허브는 게시 및 구독 기능을 통해 데이터 재사용을 촉진하고, 최적화, 데이터 표준 및 거버넌스를 위한 중앙 제어 기능을 제공합니다. 중앙 집중식 관리는 데이터 원본 위치에 상관없이 연결할 수 있는 모든 조정된 데이터 이동 경로에 대한 향상된 가시성을 제공합니다.

차세대 데이터 통합 허브는 기존 기능을 셀프 서비스 분석 사용자로 확장합니다. 누구든지 최소한의 IT 부서의 도움으로 최신 통합 허브 데이터 피드를 게시하거나 구독할 수 있습니다. 데이터 소비자는 인증된 데이터를 활용하고 내역 및 통합 프로세스에 대한 가시성을 얻을 수 있습니다. 최신 데이터 통합 허브의 다른 이점으로는 원활한 데이터 품질 기능, 신속한 데이터 원본 준비 및 크고 작은 데이터 집합의 적시 제공을 들 수 있습니다.

Informatica와 Cisco는 데이터 통합 허브 기술의 시장 선두 주자입니다. Tableau는 Informatica와의 긴밀한 통합을 통해 수백 가지의 서로 다른 데이터 원본을 Tableau 데이터 추출로 결합하고, 조직의 모든 사람이 사용할 수 있도록 [Tableau Data Server](#)에 최신으로 저장 및 보관할 수 있습니다.

구조화되지 않은 데이터, NoSQL 및 데이터 레이크에 대한 추가 정보

데이터 레이크는 누구나 원시 데이터를 다양한 방식으로 신속하게 분석할 수 있도록, 더욱 빠르고 유연한 데이터 수집 및 저장을 통해 최신 빅 데이터 분석 요건을 충족합니다. 데이터 레이크가 데이터 웨어하우스를 대체하는 것은 아닙니다.

최신 수집 및 로드 설계 패턴에서 모든 크기와 모양의 원시 데이터는 통상 데이터 레이크에 전송되어 저장됩니다. 데이터 레이크는 방대한 양의 데이터를 구조화, 반구조화 및 구조화되지 않은 원래 형식으로 보유하고 있는 스토리지 저장소입니다. 데이터 레이크는 또한 '읽기 스키마' 기능으로 원시 데이터를 변환하는 API 또는 SQL을 통해 최적화된 처리 메커니즘을 제공합니다.

Hadoop은 복원력과 저렴한 비용 덕분에 첫 번째 Hadoop 분산 파일 시스템(HDFS) 이래로 데이터 레이크에 사용되고 있지만 유일한 데이터 레이크 구현 옵션은 아닙니다. Amazon Web Services S3 (Simple Storage Service), 유연한 스키마의 NoSQL 데이터베이스와 같은 개체 저장소도 데이터 레이크로 사용될 수 있습니다. Tableau는 이제 Amazon S3에 연결하는 [Amazon Athena 데이터 서비스를 지원](#)하며, NoSQL 데이터베이스에 직접 연결할 수 있는 다양한 도구도 제공합니다.

최신 분석 아키텍처에서 NoSQL 데이터베이스는 데이터가 어디에 있든지 신속한 데이터 로드와 스키마 없는 데이터베이스 개념의 장점으로 인해 표준이 되고 있습니다. NoSQL, Non-SQL 또는 Not-Only-SQL 데이터베이스는 데이터 저장의 대체 유형을 제공합니다. 일반적인 NoSQL 저장 유형에는 열, 문서, 키 값 및 그래프가 포함됩니다.

Tableau에서 자주 사용되는 NoSQL 데이터베이스에는 MongoDB, Datastax, MarkLogic 등이 있습니다.

Hadoop은 빅 데이터 플랫폼으로 자주 사용되지만 데이터베이스가 아닙니다. Hadoop은 범용 하드웨어 클러스터에서 데이터를 저장하고 애플리케이션을 실행하기 위한 오픈 소스 소프트웨어 프레임워크입니다. Hadoop은 모든 종류의 데이터를 위한 대용량 스토리지, 방대한 처리 능력과 엄청난 양의 동시 업무 또는 작업을 처리할 수 있는 기능을 제공합니다.

최신 분석 아키텍처에서 Hadoop은 오래된 기록 데이터를 데이터 웨어하우스에서 온라인 콜드 스토어로 이전하는 데 저렴한 비용으로 스토리지 및 데이터 아카이브 기능을 제공합니다. 또한 IoT, 데이터 과학 및 구조화되지 않은 분석 사용 사례에 사용됩니다.

Hadoop 프레임워크 내의 데이터 로드, 정리 및 쿼리를 지원하는 관련 기술에는 다음이 포함됩니다.

- **Apache Spark** - 고성능 인메모리 분석을 지원하며, 관련 프로젝트가 계속 증가하는 오픈 소스 클러스터 컴퓨팅 프레임워크
- **Apache Impala** - Apache Hadoop용 오픈 소스 분석 MPP 데이터베이스. Tableau를 사용한 성공적인 Hadoop 연관 프로젝트에서 가장 일반적으로 사용되는 데이터 연결입니다.
- **Apache Presto** - 모든 크기의 데이터 집합에서 대화식 쿼리 실행을 지원하는 오픈 소스 분산 SQL 쿼리 엔진. [Tableau는 버전 10에 Presto 지원을 추가했습니다.](#)
- **MapReduce** - 입력을 받은 문제를 소규모로 분할하여 작업자 노드에 배포하는 병렬 처리 소프트웨어 프레임워크

- **Hive** - 데이터 웨어하우징 및 SQL 계열 쿼리 언어. Hive 2.0에는 또한 Hive 쿼리 성능을 획기적으로 향상시키는 LLAP(Live Long and Process)가 포함되어 있습니다.
- **Hadoop 분산 파일 시스템(HDFS)** - 사전 정리 없이 여러 컴퓨터에 걸쳐 데이터를 저장하는 확장 가능 시스템
- **YARN** - (Yet Another Resource Negotiator) Hadoop에서 실행되는 프로세스에 대한 리소스 관리를 제공
- **Ambari** - Hadoop 서비스 및 구성 요소를 관리하기 위한 웹 인터페이스
- **Cassandra** - 분산 데이터베이스 시스템
- **Flume** - 데이터를 HDFS로 스트리밍하는 소프트웨어
- **HBase** - Hadoop 위에서 실행되는 비관계형 분산 데이터베이스
- **HCatalog** - 테이블 및 스토리지 관리 계층
- **Oozie** - Hadoop 작업 스케줄러
- **Pig** - HDFS에 저장된 데이터를 조작하기 위한 플랫폼
- **Solr** - 확장 가능한 검색 도구
- **Sqoop** - Hadoop과 관계형 데이터베이스 간에 데이터 이동
- **Zookeeper** - 분산 처리를 조정하는 애플리케이션

지난 2년간 Apache Spark는 Hadoop 환경의 하나의 구성 요소에서 눈에 띄게 많은 기업이 선택하는 독립형 빅 데이터 분석 플랫폼으로 변모했습니다. Spark는 Hadoop에 비해 데이터 처리 속도가 획기적으로 향상되었습니다. Spark 자체에 핵심 Apache Spark runtime, Spark SQL, Spark Streaming, MLlib, ML, GraphX 등 많은 관련 프로젝트가 포함되어 있습니다. 현재 250개 이상의 조직에서 1,000명이 넘는 사용자가 참여하는 가장 큰 빅 데이터 오픈 소스 프로젝트가 되었습니다.

Tableau는 빅 데이터 관련 분석 연결 및 시각적 데이터 분석에서 시장 선두 주자입니다. 업계 최고의 빅 데이터 분석 프로그램에서 Cloudera, Spark SQL, Amazon EMR, Hortonworks, Microsoft HDInsight/Data Lake, MapR 등과 함께 Tableau를 사용하고 있습니다. 이러한 지원되는 기술이나 드라이버를 통해 다른 많은 빅 데이터 기술을 Tableau에 연결할 수 있습니다.

서비스형 데이터(DaaS)

디지털 세계에서 데이터가 금이라면 데이터는 모든 사람이 소비할 수 있는 상품이라는 뜻이 됩니다. 고객, 금융, 시장, 날씨, 지리 및 인구 통계학 데이터는 이미 데이터 시장 및 거래 플랫폼에서 구매 서비스로 제공되고 있습니다.

서비스형 데이터(DaaS)는 클라우드를 통한 데이터 전달을 위한 유연한 서비스 지향 아키텍처(SOA) 패턴을 반영합니다. 이 접근법은 SOA 아키텍처가 단순하기 때문에 최고의 민첩성을 제공합니다. 오늘날, ISV, CRM, ERP에서 통합 또는 외부 보고 시나리오를 위한 표준 서비스형 데이터(DaaS) REST API를 제공하고 있습니다.

Tableau의 **웹 데이터 커넥터 SDK**를 사용하면 기존 커넥터 외부에 있는 데이터로의 연결을 구성할 수 있습니다. 셀프 서비스 분석 사용자는 내부 웹 서비스, JSON 데이터 및 REST API를 포함하여 HTTP를 통해 액세스할 수 있는 거의 모든 데이터에 연결할 수 있습니다.

논리적 데이터 웨어하우스

선도적인 분석 조직은 Cisco, Denodo와 같은 공급 업체의 데이터 가상화 기술을 통해 모든 곳에서 유연하고 논리적이며 통합된 데이터의 차원 뷰를 제공합니다. 분석 사용자에게 논리적 데이터 웨어하우스는 관계형 데이터 웨어하우스처럼 보이고 작동합니다. Tableau 사용자는 바로 사용할 수 있는 ODBC 드라이버를 통해 이러한 도구에 연결할 수 있습니다.

데이터 가상화의 핵심 기능 중 하나는 서로 다른 쿼리를 다양한 데이터 원본 및 REST API에 원격으로 분산하여 최적화하는 것입니다. 논리적 데이터 웨어하우스는 데이터 원본 변경에서 보고 애플리케이션을 버퍼링하는 의미론적 계층으로서의 역할도 합니다. 논리 데이터 웨어하우스는 종종 엔터프라이즈 데이터 카탈로그와 함께 사용됩니다.

마스터 데이터 관리

뛰어난 분석이 가능하려면 그 기반이 되는 데이터가 우수해야 하며, 사람들은 데이터가 정확할 경우에만 올바른 결정을 내릴 수 있습니다. 해당 전문가가 상향식 접근 방식으로 데이터를 작성함에 따라 기존 데이터 품질 및 마스터 데이터 관리에 대한 관심이 새로워지고 있으며, 보고 데이터 원본이 최신으로 정리되어 일관적이고 정확하도록 보장하는 것이 다시 한번 중요해지고 있습니다.

널리 사용되는 마스터 데이터 관리 서비스의 제공업체로는 Informatica, IBM, Stibo를 들 수 있습니다. Tableau에서 가장 많이 사용되는 데이터 품질 솔루션은 Trillium, Informatica Data Quality, Talend Data Quality 6.0, Tamr Eisenhower 등입니다.

엔터프라이즈 데이터 카탈로그

또 다른 새로운 기술로 엔터프라이즈 데이터 카탈로그가 있습니다. 엔터프라이즈 데이터 카탈로그를 통해 자립형 보고 사용자는 승인된 데이터 원본으로부터 의사 결정을 위한 올바른 데이터를 쉽게 찾을 수 있습니다. 엔터프라이즈 데이터 카탈로그는 시각적 분석 솔루션 내에 존재하는데, Tableau와의 원활한 통합을 위해 독립형 서비스로도 설계되어 제공됩니다.

엔터프라이즈 데이터 카탈로그는 수집된 데이터 원본을 검사하여 테이블, 뷰 및 저장 프로시저의 메타데이터로 채워집니다. 새로운 데이터 원본, 지능적인 데이터 분류 및 교차 데이터 원본 엔터티 매핑의 자동화된 검색을 통해, 데이터 카탈로그는 기본적으로 데이터 원본 및 공통 데이터 정의에 대한 엔터프라이즈 비즈니스 용어집의 역할을 합니다. 해당 전문가는 주석, 버전 및 문서를 추가하여 카탈로그 데이터 원본 컨텍스트를 개선합니다.

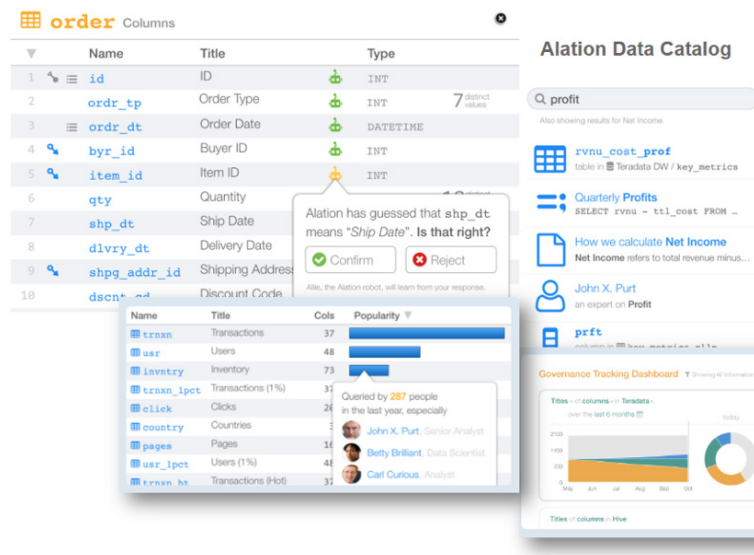


그림 4 Alation의 엔터프라이즈 데이터 카탈로그

데이터 카탈로그 솔루션은 데이터 정리 및 기존 데이터의 효율적인 재사용을 촉진합니다. 또한 선호하는 방식의 데이터 이력 정보와 한층 강화된 데이터 거버넌스, 보안, 로그 및 감사 기능을 제공합니다.

Tableau와 원만하게 통합되는 다양한 데이터 카탈로그를 보유한 공급 업체로는 Alation, Collibra, Attivio, Informatica, Waterline 등이 있습니다.

기계 학습

고급 분석을 인지, 딥 러닝, 인공 지능 분야에까지 확대하면, 기존 데이터 및 패턴을 추론하고 기존 지식 기반을 바탕으로 결론을 내린 다음 이를 다시 영구적이고 연속적인 자가 학습 루프의 지식 기반에 삽입합니다.

이러한 유형으로 분석을 사용하는 것은 통상 보고서 또는 통합 애플리케이션의 내장 API를 통해 결과물을 검토하는 것을 의미합니다. 현재 CognitiveCode, Digital Reasoning 등의 공급 업체 결과물을 시각화하는 데 Tableau가 사용됩니다.

Tableau 10.3에서 데이터에 연결하고 준비할 때 시간을 절약하기 위한 [권장 테이블 및 스마트 조인](#)을 소개했습니다. 권장 사항은 시간이 지남에 따라 데이터베이스가 더 자주 사용되면서 기계 학습을 통해 향상됩니다.

자연어

데이터 스토리텔링을 자연어 및 음성 쿼리 분야로 확장하여 모든 사람에게 데이터 검색을 새롭고 더 유연한 방식으로 제공하기 시작했습니다. 자연어는 주요 발견 사항에 대한 자동화된 문맥 설명을 받거나, 예측을 요청하거나, 텍스트 문서의 양을 분석함으로써 모든 플랫폼에서 분석에 더 쉽게 액세스할 수 있게 합니다.

오늘날, Tableau의 시각적 분석은 Yseop, Narrative Science, Automated Insight와 같은 선도적인 자연어 생성(NLG) 솔루션과 결합될 수 있습니다. 이러한 기술은 주로 Tableau 비주얼라이제이션의 컨텍스트를 해석하므로, 통합은 대개 자연어 도구 자체에서 또는 JavaScript의 확장으로서 수행됩니다. 또한, [ClearGraph의 인수](#)를 통해 더 스마트한 데이터 검색 및 분석이 Tableau에서 바로 가능해지며, 자연어를 통해 데이터와의 상호 작용이 더 쉬워질 것입니다.

권장 데이터 검색

스마트 데이터 검색은 기계 학습 알고리즘을 활용하여 훨씬 더 심층적인 분석('일어날 일')과 그러한 목적으로 서식 지정된 데이터를 통한 규범적('최적화 방법') 기능을 제공합니다. 시각적 분석이 발전함에 따라 새로운 자동화된 인사이트와 권장 사항이 추가될 것입니다. 이러한 기능은 [2016 Tableau Conference](#)의 로드맵 기조연설에서 설명된 바 있습니다. R, Python, API 및 분석적 데이터베이스 기능, 고급 분석 기능의 결과를 통합하여 Tableau에서 시각화하고 탐색할 수 있습니다.

검색

최신 분석 아키텍처를 사용하면 사용자와 전문가 모두 Google을 사용하듯 쉽게 위치에 관계 없이 데이터를 검색하고 찾을 수 있습니다. 데이터 모델링과는 달리, 분석적 검색 엔진 인덱싱 기술은 필드명, 데이터 유형 및 기계 학습 인텔리전스에 기반하여 서로 다른 데이터 원본을 자동으로 연관시킵니다. 시간이 지남에 따라, 이전 쿼리 및 보고된 사용 사례를 기반으로 동적 검색 제안이 생성됩니다. 최근 Siri, Alexa와 같은 음성 기술이 추가됨에 따라 음성 쿼리 기능과 분석적 검색 기능이 결합되기 시작했습니다. 이것의 좋은 예로, Automated Insight는 Alexa 음성 제어를 Tableau에 도입하여 Tableau Conference 2016에서 [올해의 해커톤 우승자](#)가 되었습니다.

알림

최신 분석 아키텍처에는 디지털 데이터 바다에서 유용한 신호를 지속적으로 모니터링하는, 설정 가능하고 지능적인 데이터 기반 알림 기능이 포함됩니다. 인간이 모든 중요한 값을 24시간 연중무휴로 확인하는 것은 불가능합니다. 이러한 이유로 자동화 및 알림 기능은 최신 분석 도구의 훌륭한 자산입니다.

어떤 도구는 일정한 간격으로 스냅샷을 제공합니다. 다른 도구는 실제로 로그를 추적하여 수치가 특정 임계값을 초과했는지 확인합니다. 두 방법 다 필요합니다. 어떤 대시보드는 유익한 정보를 전달하므로 단지 매일 확인할 필요가 있습니다. 다른 대시보드는 중요한 조치의 근거가 되지만, 실행 가능한 인사이트 없이 대시보드를 매일 점검하는 것은 시간을 효율적으로 사용하는 것이 아닙니다.

Tableau를 사용하면, [Tableau Server의 데이터 기반 알림](#)을 통해 비즈니스를 완벽하게 파악할 수 있습니다. 본인 또는 팀 전체에 대한 이메일 알림을 위한 임계값을 선택하기만 하면 됩니다.

스토리텔링

때로는 인사이트, 즉 '무엇'만으로는 충분하지 않습니다. 사람들은 데이터 뒤에 가려진 '이유'도 알고 싶어합니다. 매출이 왜 증가하였습니까? 웹 트래픽이 급증한 이유는 무엇입니까? 왜 우리는 의료 용품 수급에 어려움을 겪고 있습니까?

기업은 분석 기술을 텍스트, 그림, 비디오에 이르기까지 다른 형태의 소통 수단과 결합함으로써 이 문제를 해결하기 위해 오랫동안 노력해 왔습니다. 분석가는 Powerpoint를 사용하여 프레젠테이션을 작성하거나, 긴 보고서를 PDF로 작성하거나, 더 번거로운 방식으로 수십, 수백 장의 문서를 인쇄하여 바인더에 철하기도 합니다.

오늘날의 최신 분석 도구는 이러한 스토리텔링 개념에서 장점을 살려 실제로 이를 최고 수준의 기능으로 통합합니다. 대화형 대시보드를 구축하고, 새로운 데이터가 추가될 때 백그라운드에서 자동으로 업데이트되는 데이터의 특정 스냅샷을 보내고, 대화식 차트에 텍스트와 이미지가 결합된 보고서를 만들 수도 있습니다. 스토리텔링은 수치만 제공하는 게 아니라 사용자가 데이터 분석을 설명할 수 있게 해줍니다.

Tableau는 근본적으로 선택의 중요성과 개방형 표준을 중시합니다. Tableau는 분석을 빠르고 쉽게 하기 위해 연구 개발에 많은 투자를 하고 있으며, 이는 파트너와 함께 하는 혁신을 의미합니다. 이는 분석 세계가 발전하고 새로운 기술이 적용됨에 따라, 분석 리더가 항상 현재 및 미래의 데이터 기술 중 선택하여 Tableau와 통합할 수 있게 합니다.

Tableau 정보

데이터가 얼마나 큰지, 어떤 채널에서 왔는지, 어떤 데이터베이스에 저장되어 있는지와 관계없이 Tableau는 사람들이 데이터를 보고 이해할 수 있도록 지원합니다. PC에서 iPad에 이르기까지 편리한 환경에서 데이터에 신속하게 연결, 통합, 시각화 및 공유할 수 있습니다. 자동 데이터 업데이트 기능이 있는 마케팅 대시보드를 만들고 게시하고, 실시간 인사이트를 동료, 팀, 경영진, 파트너 또는 고객과 공유할 수 있으며 프로그래밍 기술이 필요하지 않습니다. 지금 바로 [무료로 체험해 보십시오.](#)

추가 리소스

[분석 문화를 구축하는 방법](#)

[분석의 정의](#)

[분석에 대한 접근 방식\(솔루션 페이지\)](#)

[클라우드 데이터 개요](#)

[Tableau를 사용한 고급 분석](#)

