



InDetail

Bloor InDetail Paper
저자 Philip Howard
발행일 2019년 1월

IBM Cloud Private for Data 1.2

“

AI를 구현하려는 기업에서
목적한 어려움 중 하나로,
관련 모델을 개발하는 데이터
과학자와 이러한 모델을
실운영에 배포하는 담당자 간
소통이 단절되는 문제가 있다는
점을 특히 강조하고 싶습니다.

”

전체 개요

A I(인공지능)와 기계학습(AI의 한 분야)은 IT 및 분석 커뮤니티만이 아니라 일반적인 비즈니스에서도 요즘 유행하는 화제입니다. 미디어 전반과 회의실 내에서도 계속해서 언급되고 있습니다. 효율성 증가 및 경쟁력 있는 차별화 요소라는 직접적인 면에서, AI와 기계학습 관련 기법 및 기술을 구현했을 때 잠재적 가치가 매우 크다는 데 전반적인 합의가 이루어진 상태입니다. 새로운 비즈니스 모델과 서비스를 도입할 수 있다는 점은 말할 것도 없습니다. 다만, 잠재적 이점이 눈에 띄는 것과 별개로 AI와 기계학습을 구현하는 일은 그다지 간단하지 않습니다. 예상에 따르면(그림 1 참조) 매우 많은 기업들이 향후 몇 년간 AI에 투자할 것으로 전망되지만, 현재까지 실제로 AI에 투자한 기업은 얼마 되지 않습니다. 여기에는 문화적 변화 사안, 보안 문제, 인재 영입 등 여러 가지 이유가 있는데, 본 보고서에서는 논의 대상이 아닙니다. 대신 본 보고서에서는 데이터 및 분석에 관하여 한정되거나 기술적 역량 없는 문제를 다루며, 특히 IBM Cloud Private for Data에 대해 논의할 예정입니다. 이 제품은 2018년 5월에 처음 출시되었으나, 본 보고서에서는 2018년 12월 기준 가장 최신 제품까지 반영합니다. 현재 해당 제품은 Cloud Native Edition(기본 지원) 및 Enterprise Edition(프리미엄 지원)과 같이 2가지 버전으로 이용 가능합니다. 또한 IBM Cloud Private Experiences를 통해서도 제품 설치 또는 별도의 비용 없이 환경을 시범 운영해 볼 수 있습니다.

IBM Cloud Private for Data는 통합 데이터 과학, 데이터 공학 및 IBM Cloud Private(ICP)을 기반으로 구축한 앱 구축 플랫폼입니다. 후자의 목적은 a) 방화벽 내에서 클라우드 컴퓨팅의 이점을 모두 누릴 수 있도록 하고, b) (퍼블릭) 클라우드 배포를 확장하는 데 필요한 디딤돌 역할을 하는 것입니다. 게다가 ICP에는 부가적인 이점이 있는 마이크로 서비스 아키텍처가 있는데, 이에 관해서는 차후 다룰 예정입니다. 이외에도 ICP for Data는 데이터 기반 처리와 운영을 구현하고 특히 AI와 기계학습 기능의 개발 및 배포를 모두 지원하기 쉽도록 만들어 주는 환경을 제공합니다. 이 마지막 포인트가 중요한 사항인데, 바로 데이터 과학자(비즈니스 부서를 위해 주로 일하는 사람)와 이러한 데이터 과학자들의 작업을 운용해야 하는 사람들(주로 IT) 간에는 의사소통이 단절되기 쉽기 때문입니다.

본 보고서에서는 먼저 ICP와 ICP for Data를 차례로 살펴보겠습니다. 그러나 둘 중 어느 것이든 배포하고자 하는 일반적인 이유를 몇 가지 간단하게 논의해 볼 필요가 있습니다. ICP의 경우, 일반적으로 클라우드 컴퓨팅이 선사하는 유연성과 규모 외에도 한두 가지 이유가 있습니다. 첫 번째는 ICP에서 제공하는 마이크로 서비스를 이용하여 이들을 리팩터링함으로써 기업 애플리케이션을 현대화할 수 있는 능력입니다. 두 번째는 새로운 클라우드 네이티브 애플리케이션 또는 퍼블릭 클라우드 서비스 등 기존의 애플리케이션과 데이터를 활용하는 클라우드 네이티브 애플리케이션을 구축하면서도 방화벽 내에서 안전하게 데이터 보안을 유지할 수 있는 능력입니다. ICP for Data의 수많은 사용 사례는 실제로 이러한 기능들을 확장시킨 것으로, 운영 애플리케이션이 의사 결정을 내릴 수 있도록 하거나 애플리케이션에 기계학습이 구축되도록 합니다. 교육은 ICP for Data 내에서 이루어지지만 사용자의 선택에 따라 직접적으로 또는 클라우드 네이티브 애플리케이션을 통해 배포하게 됩니다. 다른 사용 사례도 있습니다. 예를 들어 이러한 점들을 모두 포괄하는 데이터 카탈로그 및 데이터 가상화(아래 참조) 기능이 있는 ICP for Data를 이용하여 데이터 레이크를 손쉽게 배포할 수 있어 레이크가 격리되지 않습니다.

“
효율성 증가 및 경쟁력 있는 차별화 요소라는 직접적인 면에서, AI와 기계학습 관련 기법 및 기술을 구현했을 때 잠재적 가치가 매우 큽니다.
”

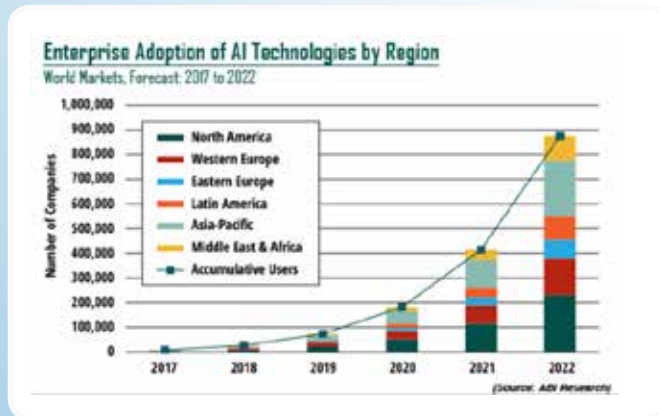
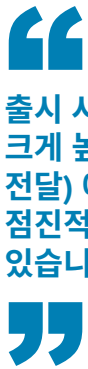


그림 1: AI 도입 예상

그러나, 본 보고서의 목적은 이들 제품의 기능에 대한 내용을 서술하는 것도 아니고 사용 사례를 자세하게 살펴보는 것도 아닙니다. 그보다는 이러한 제품들의 이용 목적인 사업적 이득을 언급하고, 이러한 환경에서 발생할 수 있는 문제점을 논의하는 데 있습니다. 이후 ICP 및 ICP for Data의 신규 출시 내용을 정리한 IBM 단기 로드맵 섹션으로 본 보고서를 마무리합니다.

IBM Cloud Private



출시 사이클의 속도를 크게 높이고(지속적인 전달) 애플리케이션을 점진적으로 개선할 수 있습니다.

ICP란 프라이빗 클라우드를 말합니다. 이는 퍼블릭 클라우드 공급업체가 서비스 제공업체 역할을 수행하는 가상 프라이빗 클라우드와는 별개의 것으로, ICP 또는 IT 부서에서 서비스 제공자 역할을 하거나 이를 관리하는 제3자를 둘 수 있습니다. 둘 중 어느 경우라도 사업부가 테넌트 역할을 합니다. 기본 원리를 설명하자면, 데이터를 방화벽 뒤에서 안전하게 보호하면서 클라우드 배포의 이점을 취할 수 있으며, 원하는 경우 데이터를 외부에서 호스팅할 수도 있다는 것입니다. ICP는 기존의 온프레미스 배포뿐만 아니라 퍼블릭 클라우드 및 가상 퍼블릭 클라우드와도 호환되므로 하이브리드 환경을 지원하며, IBM에서는 이를 향후 퍼블릭 클라우드로 가는 잠재적인 디딤돌로 보고 있습니다. ICP는 Linux 기반 x86 플랫폼에서 구현되며 IBM SoftLayer 뿐만 아니라 ICP 및 ICP for Data가 모두 Red Hat Certified Containers인 Red Hat(곧 IBM 내 별도의 법인으로 분리될 예정) OpenShift Container Platform에서 실행됩니다. AWS에서도 제품들을 이용할 수 있으며 IBM에서는 ICP 및 ICP for Data 또한 Azure

www.microservices.io에 따르면 "마이크로 서비스(마이크로 서비스 아키텍처)는 애플리케이션을 느슨하게 연결시킨 서비스로 구조화하여 비즈니스 역량을 구현하는 아키텍처 스타일을 말합니다. 마이크로 서비스 아키텍처를 통해 크고 복잡한 애플리케이션을 지속적으로 전달/배포할 수 있습니다." Microsoft는 더 나아가 (무엇보다도) 이렇게 말합니다. "어떤 점에서는 마이크로 서비스가 SOA(서비스 지향 아키텍처)의 자연스러운 발전이라 할 수 있으나, 차이점이 있습니다. 마이크로 서비스의 본질적인 의미를 보여주는 특징으로는 서비스가 소규모이고 독립적이며 느슨하게 연결되어 있다는 점, 각 서비스가 별도의 코드베이스이므로 소규모 개발팀에서 관리 가능하다는 점, 서비스를 독립적으로 배포할 수 있고 기존의 서비스를 전체 애플리케이션 재구축 및 재배포 없이도 업데이트 가능하다는 점, 다른 서비스에서는 보이지 않는 상세한 구현으로 잘 정의된 API를 이용하여 서비스 간 서로 커뮤니케이션이 이루어진다는 점 등을 들 수 있습니다.

IBM에서는 ICP 및 ICP for Data를 통해 **그림 2**에 나타낸 것과 같이 기존의 역량을 취하면서도 가능한 경우 이를 서비스로 재현했습니다. 이러한 접근법이 지니는 장점 중 일부는 위 인용문에 설명되어 있으나, 다른 장점도 있습니다. 예를 들어 호환성을 향상한다는 점이 주목할 만합니다. 예를 들어, 사용하고자 하는 서비스를 취사선택할 수 있기 때문에 데이터베이스 서비스를 취하고 이를 기계학습 서비스와 통합하는 일이 훨씬 더 수월해집니다. 또 강조할 만한 점은 바로 출시 사이클의 속도를 크게 높이고(지속적인 전달) 애플리케이션을 점진적으로 개선할 수 있다는 것입니다. 이 점이 중요한 이유는 새로운 기능을 보다 빠르게 이용할 수 있기도 하지만, 1년 또는 8개월 단위의 기존 출시 사이클에서는 차질이 발생하는 일이 많고 이것이 결국 신규 출시 제품 구현의 지연으로 이어지기 때문입니다. 또한 클라우드 기반 환경이라는 점을 감안할 때 IBM이 동일한 방식으로 새로운 역량을 점진적으로 도입할 수 있다는 의미도 됩니다.

그림 2: IBM Cloud Private for Data 구성 요소



플랫폼으로 포팅될 예정임을 발표했습니다. 또한 IBM은 ICP for Data를 어플라이언스로 이용할 수 있도록 하고, 제품을 Power and Z 중앙 컴퓨터 플랫폼에서 이용할 수 있도록 할 계획입니다. ICP for Data는 ICP 자체의 기본 기능 일체를 활용하여 관리, 보안, 기록, 모니터링 등을 수행합니다.

클라우드 기반 컴퓨팅의 장점, 특히 탄력적인 확장, 신속한 배포, 확장성 등에서 기인한 장점이 잘 알려져 있으며 본 보고서상에서는 이에 대해 반복하여 언급하지는 않을 예정입니다. 다만 ICP(및 ICP for Data)에는 마이크로 서비스 아키텍처가 있으며, 이것이 가져다 주는 이점을 논의해 볼 만합니다. Docker 및 Kubernetes를 기반으로 마이크로 서비스를 이용하는 것이 비교적 새로운 개념에 속하기 때문입니다.

IBM Cloud Private for Data

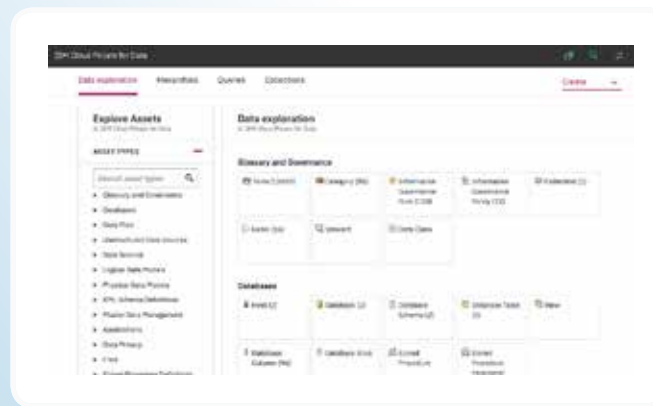
A I 이니셔티브를 지원하기 위해서는 필요한 모든 데이터를 입수하고, 이를 통제하여 신뢰성을 확보하며, 현재 다루고 있는 프로젝트에 필요한 기계학습 및 기타 알고리즘을 분석 및 구축하여 마침내 이러한 활동의 결과를 실운영에 반영할 수 있어야 합니다. 이는 결코 사소한 작업이 아닙니다. **그림 2**에서 ICP for Data의 구성 요소를 설명한 바와 같이, 이러한 활동을 담당하는 개인 및 그룹은 이질적이고 서로 소통하지 못하는 경우가 많으며, 이러한 작업을 효율적으로 하기 위해서는 협업 접근 방식이 필요합니다. 더욱이, 대부분의 소프트웨어 제공업체를 넘어서는 역량이 있어야 하며, 필요한 협업을 수행할 수 있으려면 필수적인 모든 기본적인 소프트웨어 기능 전체에 걸쳐 일관적인 사용자 인터페이스를 택해야 합니다. 잠시 후 ICP for Data에 관한 자세한 내용을 논의하겠지만, 가장 주목할 만한 점은 이 제품이 소프트웨어 스택 전체에 걸쳐 일반 사용자 경험을 제공한다는 사실입니다. 또한 이를 가능케 한 ICP for Data의 기저를 이루는 것은 바로 사용자 인터페이스와 개별 소프트웨어 구성 요소를 서로 분리한 다음 일관적인 인터페이스로 새롭게 만들어 내는 마이크로 서비스 아키텍처임을 언급할 가치가 있습니다. 그 예시는 본 보고서에 다수 설명되어 있습니다.

논의 대상이 되는 ICP for Data 아키텍처에는 여러 요소가 있습니다. 핵심은 바로 Enterprise Data Catalog로, Data and Analytics Governance 레이어의 일부를 형성합니다. 이는 기본적으로 조직에서 이용할 수 있는 상세 데이터 자산을 모두 보유하고 있는 라이브러리 카탈로그와 동일합니다. 그러나, 단순하게 이름이 붙여진 책마다 기준을 제공하는 라이브러리와는 달리, 데이터 카탈로그를 이용하면 카테고리별로 검색을 할 수 있어 판매, 고객 또는 제품과 같은 관련된 모든 자산을 찾을 수 있습니다. 즉, 다른 방식으로라면 인식하지 못했을 수도 있는 본인 역할에 맞고 가치 있는 자산을 찾을 수 있습니다. 이러한 내용은 **그림 3**에 자세히 설명되어 있습니다.

본 보고서에서 IBM에 문제를 제기하는 부분은 **그림 2**에서 제일 위 2개 박스가 서로 구분되어 있다는 것입니다. 마케터처 (marketecture) 다이어그램이라는 점을 감안해도 이 그림에는 IBM의 의도 또는 역량이 정확히 반영되어 있지 않습니다. 앱 개발자와 데이터 엔지니어가 "맞춤형 협업 팀 플랫폼"을 공유하지 않는다는 인상을

줍니다. 이러한 플랫폼에는 클라우드 소싱 권장 사항 외에도 다른 사용자도 사용할 수 있는 워크플로와 같은 보다 평범한 역량이 포함되어 있습니다. 이러한 표현 방식은 부정확하다고 봅니다. 그리고 당연히 부정확한 것이어야 합니다.

위 **그림 2**의 Data and Analytics Governance 레이어에는 3개의 박스가 있습니다. IBM에서는 이를 "수집, 구성, 분석"으로 나타냅니다. Bloor에서는 아마도 데이터 통합을 "수집" 박스에 넣고 "구성"을 "거버넌스"로 이름을 변경하겠지만, 이는 단순히 단어 차이에 불과합니다. 이들 각각에 대해서는 별도로 논의할 예정입니다.



“ AI 및 기계학습과 관련하여 어떠한 함의가 있는데 이 부분은 얼핏 명확하지 않을 수도 있습니다. ”

그림 3: Enterprise Data Catalog를 통한 데이터 자산 분석

데이터 수집

원칙적으로 "수집" 박스는 각종 데이터 소스를 이용할 수 있는 능력을 뒷받침합니다. 당연한 말이지만, 데이터가 없으면 분석할 수가 없기 때문입니다. 여기에는 "주문형 데이터베이스", "데이터 가상화"와 같이 2가지 주된 측면이 있으며 이에 대해서는 별도의 논의가 필요합니다. 가장 최신 버전에서 전자는 전체 Db2 제품군(Db2, Db2 Warehouse, Db2 Event Store 등)을 비롯하여 Hadoop, Big SQL(Hadoop 엔진에서 구동하는 IBM SQL) 및 MongoDB(이번 출시에서 신규 추가)를 아우릅니다. 특히 Db2 Event Store를 지원한다는 점에 주목해야 합니다. 센서, 액추에이터 및 기타 말단 장치에서 발생한 데이터를 저장해야 하는 IoT(사물 인터넷) 기반 환경에서 이것이 매우 중요한 부분이기 때문입니다.

ICP for Data가 처음으로 출시된 당시에는 "수집" 박스에 데이터 연함이 포함되어 있었으나 이후 데이터 가상화로 대체되었습니다. 이러한 점은 제품에 상당한 보강이 이루어졌음을 나타내는데, 그 원리를

설명해 볼 필요가 있습니다. 이 분야에서 IBM의 역량이 혁신적이며 시장에 출시된 다른 제품을 앞서기 때문입니다. 이 기술(이전 코드명 QueryPlex)이 어떻게 실현되는지 이해하기 위해서는 **그림 5**에 기술된 바와 같이 IBM의 계산 그물망을 이용하여 **그림 4**에 나타낸 바와 같이 기존의 접근법과 데이터 연합이 구분되는 방식을 비교해야 합니다.

보다시피 주요 차이점은 계산 그물망이 로컬에서만뿐만 아니라 로컬 집합체 내에서도 분석을 수행한다는 것입니다. 네트워크 전반에 걸친 데이터 이동이 기존의 데이터

그림 5에서는 IBM Data Virtualization의 원리를 개념적으로 보여 주며, 사용자의 시각에서 보이는 내용은 **그림 6**에 나타나 있습니다. 즉, **그림 5**에 설명된 바와 같이 말단 장치는 스스로 구성되며, 사용자는 이러한 과정이 어떻게 진행되는지 알 필요가 없습니다. 여기에서는 "말단 장치"라고 하며 이 표현은 IoT 환경에서 확실히 적합하지만, 해당 내용은 데이터베이스 및 말단에 있을 필요가 없는 기타 소스에도 적용됩니다.

데이터 구성 및 분석

"구성"에 관한 한, ICP for Data는 데이터 정화, 데이터 암호화, 거버넌스 등을 지원하는 InfoSphere 브랜드의 친숙한 기술을 활용합니다. 이러한 맥락에서 볼 때 언급할 만한 점은, AI 및 기계학습과 관련하여 어떠한 함의가 있는데 이 부분은 얼핏 명확하지 않을 수도 있다는 것입니다. 예를 들어, 자동 의사 결정을 생각해 보십시오. 즉, 일부 전산화된 알고리즘 종류가 관련 프로세스의 일환으로서 수신 정보를 바탕으로 자동으로 의사 결정을 내리는 경우를 말합니다. 해당 알고리즘에서는 절대적으로 신뢰할 수 있으며 편파적이지 않은 데이터로 프로세스를 진행해야 하며, 그렇지 않으면 비즈니스상 치명적인 결과를 야기할 수 있을 정도로 의사 결정에 결함이 발생할 수 있습니다. 즉, 데이터 품질(주로 이름 및 주소 중복 제거 및 정화만으로 여겨지는 경우가 많음)은 이러한 맥락에서 볼 때 근본적으로 중요한 사항이며, 구조화된 데이터뿐만 아니라 센서 판독값과 텍스트와 같은 반구조 및 비구조적 데이터에도 적용됩니다. IoT 환경에서는 비순서, 누락 및 중복 판독값과 같이 각각 대응해야 하는 문제점이 있을 수 있다는 점에도 유의하십시오. 또한 최대 및 최소 센서 판독값이 서서히 증가 또는 감소(주로 극한의 환경 조건 때문)하는 "센서 드리프트", 그리고 변칙적인 급증(관심의 문제인가, 아니면 느슨한 관계가 원인인가?)과 같은 상황을 맞이할 수 있습니다.

AI를 구현하려는 기업에서 목격할 어려움 중 하나로, 관련 모델을 개발하는 데이터 과학자와 이러한 모델을 실운영에 배포하는 담당자 간 소통이 단절되는 문제가 있다는 점을 특히 강조하고 싶습니다. 분석할 데이터를 준비하는 사람이 데이터 엔지니어, 실제로 과학적인 부분을 처리하는 사람이 데이터 과학자라고 한다면, 해당 내용은 데이터 엔지니어와 데이터 과학자에게도 적용됩니다. 따라서 여기서 필요한 것은 DevOps와 비슷한 "AnalyticOps"입니다. 이는 분석과 운영 배치 간 차이를 좁히는 역할을 하며, 이것이 바로 ICP for Data의 협업 환경이 지닌 목표입니다. 또한 이러한 역할 및 역량의 전 분야에 걸쳐 협업을 제공할 수 있는 능력은

그림 4: 기존의 데이터 연합의 원리 Vs. IBM Data Virtualization의 원리

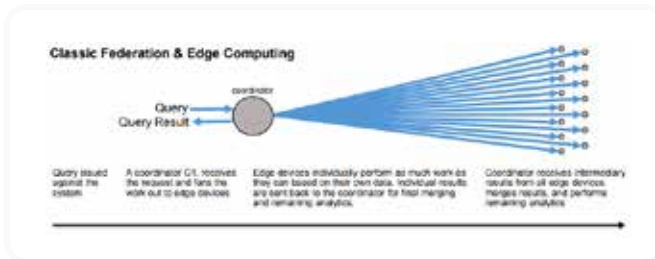
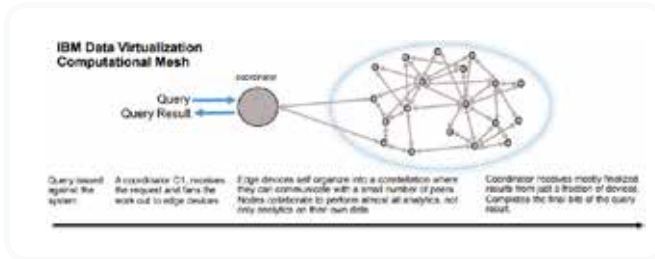


그림 5: 기존 데이터 연합의 원리 Vs. IBM Data Virtualization의 원리



연합 기술에서 가장 까다로운 문제인 점을 고려하면 상당한 수준으로 성능이 개선될 것임을 예상할 수 있습니다. 나아가 비용에 미칠 효과도 크다고 볼 수 있습니다. 더 적은 인프라로 동일하거나 더 나은 수준의 성능을 실현할 수 있기 때문입니다. 계산 그물망이 지원하는 데이터 소스로는 Db2 제품군(위 참조), Netezza, BigSQL, Informix, Derby, Oracle, SQL 서버, MySQL, PostgreSQL, Hive, Impala가 있습니다. 추가 지원 사항이 준비 중이며 자세한 내용은 아래 로드맵에 논의되어 있습니다. Data Virtualization이 Enterprise Data Catalog와 통합된다는 점에 유의하십시오.

그림 6: IBM Data Virtualization 사용자 보기



IBM의 역량을 통해서만 갖출 수 있습니다. 이를 더욱 손쉽게 가능케 하기 위해 IBM AI OpenScale이 ICP for Data를 기반으로 구축되었으며, 이는 별도의 제품으로 이용 가능합니다.

AI OpenScale은 전체 수명 주기에 걸쳐 기업이 시를 운영하고 자동화할 수 있도록 다양한 역량을 보유하고 있으며, 엔터프라이즈급 AI에 보다 손쉽게 접근할 수 있도록 하는 것을 목표로 합니다. 우선, 이 제품을 이용하면 AI 배포가 비즈니스 성과에 어떠한 영향을 미치는지, 긍정적인지 부정적인지 등을 모델 개발 또는 배포 여부와는 상관없이 모니터링할 수 있습니다.

그림 7에는 사용자 정의 조건을 기반으로 불공정하거나 부정확한 결과를 자동으로 감지하는 데 사용되는 대시보드가 나타나 있습니다. 이러한 예시에서 볼 수 있듯이 "드라이버 플랫폼"에는 2가지 문제점이 있는데, 부정확하다는 점(즉, 좋지 않은 성과)과 공정하지 못한 결과를 낸다는 점입니다. 종합적으로는 3가지 정확도 경고와 6가지 편파성 경고가 있습니다.

AI OpenScale의 두 번째 기능은 편파성을 검출하고 자동으로 마이그레이션합니다. 그림 8을 참조하십시오. 여러 기업에서 교육을 목적으로 이러한 역량을 발표했으나, 알려진 바와는 편파성을 제거한 권장 모델로 배포한 모델의 성능을 비교했을 때 런타임으로도 이용 가능한 제품으로는 AI OpenScale이 유일합니다. 재교육이 불필요합니다. 대신 사용자를 대상으로 편파성이 제거된 모델을 실운영에 배포하기에 앞서 이를 다운로드 및 점검할 수 있는 방법이 제시됩니다.

세 번째로, 모든 의사 결정 뒤에 숨겨진 권장 사항을 추적할 수 있는 감사 역량 외에도 AI OpenScale에는 의사 결정과 관련한 요인, 이러한 요인이 의사 결정에 영향을 미친 정도, 결과를 바꾸었을 수도 있는 요인의 차이 정도를 살펴볼 수 있는 설명 가능성 기능이 있습니다. 원하는 트랜잭션의 신뢰 수준을 확인할 수 있으며, 고객, 규제 당국 또는 타사를 대상으로 의사 결정 내용을 설명할 때 기업이 도움을 받을 수 있는 시설도 있습니다.

또한 이 제품은 페이로드 기록 기능을 제공합니다. 이 기능을 이용하면 평가 대상인 모든 데이터가 데이터베이스에 기록 및 영구 저장되며 AI OpenScale의 다양한 대시보드 및 기타 기능에서 이러한 데이터를 기초로 삼습니다. IBM Cognos 또는 제3자 도구를 이용하여 데이터베이스를 쿼리하여 성과 지표를 얻을 수 있습니다. 또한 이러한 지표를 내보내고 애플리케이션 지표와 결합하여 비즈니스 성과를 측정할 수도 있습니다.

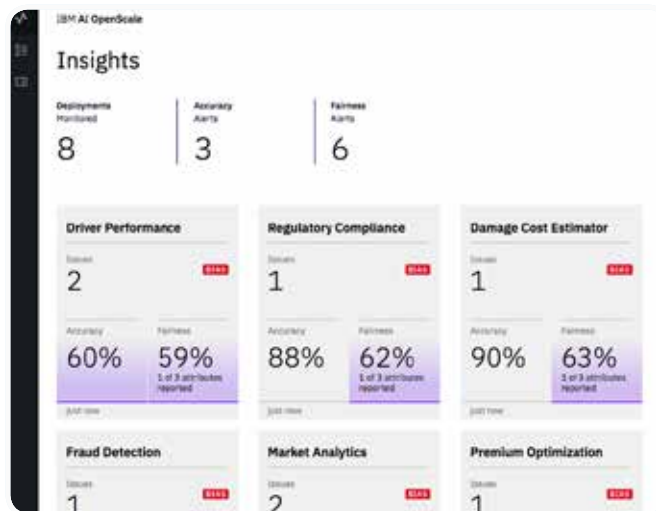


그림 7: AI OpenScale의 정확성 및 공정성



그림 8: AI OpenScale의 편파성 검출

마지막으로, AI OpenScale은 현재 베타 버전으로 이용 가능한 NeuNets(신경망 합성)를 통해 AI 수명 주기의 다른 측면을 자동화합니다. 이는 모든 런타임 환경에서 AI 배포를 위한 맞춤형 신경망을 자동으로 생성 및 권장하여, AI 구축 및 재교육을 담당하는 데이터 과학팀에 상당한 생산성 향상 효과를 가져다 줍니다.

부가적인 분석 역량 측면에서 ICP for Data는 SPSS Modeler, Watson Explorer,

Watson Studio 등 기계학습 모델을 교육, 유지 및 평가하는 데 특정한 역량을 지닌 기존의 IBM 기술을 활용합니다. 그러나, 신뢰할 수 있는 데이터와 기계학습 모델 간에는 차이가 존재합니다. 최신 버전의 ICP for Data에서는 IBM ILOG CPLEX Optimization Studio가 애드온으로 지원됩니다. 이 규범적인 분석 솔루션을 이용하면 수학 및 제약적 프로세싱을 이용한 의사 결정 최적화 모델을 개발하고 배포할 수 있습니다.

마지막으로, ICP for Data에서 모델 관리를 지원한다는 점도 빼놓을 수 없습니다. 오늘의 최적 모델이 내일의 최적 모델이 아닐 수도 있기 때문에 중요한 부분입니다. 여기에는 2가지 이유가 있는데, 첫 번째는 미래에 다루어야 하는 실제 데이터가 더 늘어나기 때문이고, 두 번째는 조건과 추세가 시간이 흐를수록 변하기 때문입니다. 따라서 정기적으로 알고리즘이나 모델을 다른 것으로 대체해야 할 수도 있습니다. 이상적으로는 이 작업이 가동 중단 없이 "핫 스왑" 방식으로 이루어져야 합니다. 어찌 되었든 모델 성과는 필요시 모니터링하고 변경해야 하며 따라서 모델 관리가 필요합니다.

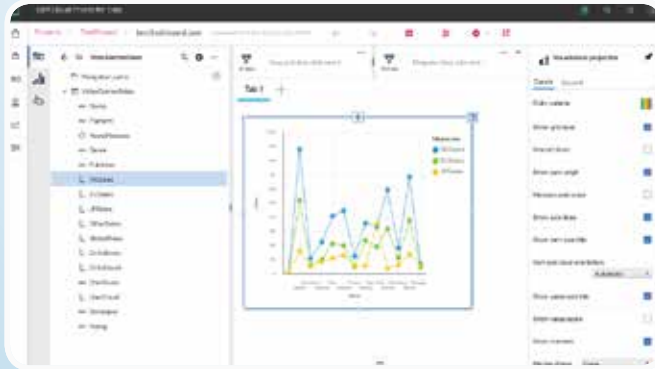


그림 9: ICP for Data의 데이터 시각화용 드래그 앤드 드롭 인터페이스

로드맵

언 급할 만한 주요 계획은 3가지입니다. 첫 번째는 ICP for Data의 "주문형 데이터베이스" 기능을 관리하는 데 사용하는 Unified Console입니다. 이를 상당 부분 향상시키려는 계획이 있습니다. 두 번째, 제품 "데이터 분석" 기능의 일환으로 Cognos Analytics가 지원될 예정입니다. 세 번째, Data Virtualization에 Excel, CSV 및 텍스트 파일, MongoDB, SAP HANA, SAS, MariaDB, CouchDB, Cloudant, 다양한 Amazon 및 Azure 데이터베이스, 다양한 스트리밍 제품, 다수의 중앙 컴퓨터 환경(IBM 및 타사), 일반 JDBC 액세스 및 여러 제3자 데이터 웨어하우징 데이터베이스 지원이 포함될 계획입니다.

결론

2

018년 4월 베를린에서 열린 Data Works 정상 회담에서 관중을 대상으로 얼마나 많은 수의 참가자 회사에서 데이터 및 분석 내용을 클라우드에 저장할 계획인지 묻는 여론 조사가 실시되었습니다. 놀랍게도 400여 명에 달하는 투표자 중 34%가 본인의 회사에는 그러한 계획이 없다고 응답했습니다. 얼마나 광고가 인상적이었던 상관없이, 실상은 수많은 조직에서 다양한 이유를 들어 이러한 단계에 발을 들이기를 주저합니다. 클라우드 기반 컴퓨팅의 이점을 깨닫지 못해서가 아니라, 현재로서는 너무 벅찬 일이라 인식하는 것입니다. ICP for Data 제품은 데이터를 방화벽 외부로 이동시키는 위험 없이 클라우드 컴퓨팅의 이점을 취할 수 있도록 하는 중간자적 위치에 있습니다.

그러나 ICP for Data가 제공하는 내용은 이것으로 끝이 아닙니다. 거의 모든 경우에서 그렇듯이 기계학습을 배포하려면 해당 기계학습이 가능한 환경이 필요합니다. IBM 에서는 이를 두고 정보 아키텍처 없이는 인공지능을 가질 수 없다고 표현합니다 (*"AI에는 IA가 필요하다"*). 정보 아키텍처를 구축할 때의 문제점은 수많은 구동 부품, 수많은 소프트웨어 요건 및 수많은 인력이 수반된다는 것입니다. 이를 위해서는 기업이 원칙적으로 AnalyticOps를 도입해야 하며, 이는 단순히 다양한 기반 기능뿐만 아니라 모든 관련자들을 아우르는 협업 지원을 필요로 합니다. ICP for Data는 지금도 발전 중이지만, 이것이야말로 해당 제품이 지향하는 방향에 해당합니다. 여러 판매업체의 서로 다른 제품으로는 확실히 더 달성하기 어려운 일일 것입니다.



ICP for Data는 초기 제품이지만 이것이야말로 해당 제품이 지향하는 방향에 해당합니다. 여러 판매업체의 서로 다른 제품으로는 확실히 더 달성하기 어려운 일일 것입니다.



추가 정보

해당 주제와 관련한 자세한 정보는 다음 웹사이트에서 확인할 수 있습니다.

www.BloorResearch.com



저자 소개

PHILIP HOWARD

연구책임자 / 정보 관리

저자 Philip Howard는 1973년 컴퓨터 산업에 입문하여 시스템 분석가, 프로그래머 및 영업 사원 등 다양한 직종에서 근무했을 뿐 아니라 GEC Marconi, GPT, Philips Data Systems, Raytheon, NCR 등 여러 회사의 마케팅 및 제품 관리 분야에서 경력을 쌓았습니다.

Philip은 약 25년간 회사에서 근무한 후 1992년 본인 소유의 회사를 설립하였으며, 첫 고객이 바로 Philip이 선임분석가로 일했던 Bloor Research(당시 ButlerBloor)였습니다. Bloor Research와의 인연이 그 후 계속 이어져 이제는 정보 관리에 주력하는 연구책임자직을 맡고 있습니다.

정보 관리는 관리, 이동, 거버넌스 및 데이터 저장뿐만 아니라 그러한 데이터에 대한 액세스 및 분석을 일컫는 것이라면 무엇이든 해당됩니다. 데이터베이스 및 데이터 웨어하우징, 데이터 통합, 데이터 품질, 마스터 데이터 관리, 데이터 통합 관리,

데이터 마이그레이션, 메타데이터 관리, 데이터 준비 및 분석 등을 포함하나 이에 국한되지 않는 다양한 기술을 수반합니다.

Philip은 Bloor Research를 대신하여 수많은 보고서를 작성했을 뿐 아니라 **IT-Director.com** 및 **IT-Analysis.com**에도 정기적으로 기고하고 있으며, 이전에는 Cambridge Market Intelligence를 대신하여 **Application Development News** 및 **Operating System News**의 편집장을 역임하기도 했습니다. 다양한 잡지에도 글을 실었고, CMI 및 The Financial Times와 같은 기업에서 발행하는 수많은 보고서를 작성한 바 있습니다. Philip은 유럽과 북미 지역에서 컨퍼런스 및 기타 이벤트에 참석하여 정기적으로 강연을 합니다.

업무 이외에 즐기는 여가 활동으로는 운하용 보트 타기, 스키, 브리지 게임(라이프 마스터 레벨) 및 외식이 있습니다.

Bloor 개요

기술은 비즈니스 진화를 빠르게 합니다. 기회는 어마어마하지만 상황에 적응하지 않으면 살아남을 수 없습니다. 따라서 격변하는 비즈니스의 시대에서 진화는 성공하는 데 필수 요소입니다.

미래를 제시하고 그러한 미래를 맞이할 수 있도록 도와드리겠습니다.

Bloor는 신선한 기술적 사고를 통해 복잡한 비즈니스 상황을 헤쳐 나갈 수 있도록 도움을 드리며 도전 과제를 진정한 성장, 수익성 및 영향력의 새로운 기회로 바꾸어 드립니다.

혁신적이고 독립적인 기술 연구, 자문 및 컨설팅 서비스를 통해 실행 가능한 전략적 통찰을 제공합니다. Bloor는 기업들이 변화의 여정을 겪는 동안 맞대응할 수 있도록 지원을 아끼지 않으면서, 복잡한 비즈니스 상황에 신선한 사고력을 불러오고 도전 과제를 진정한 성장과 수익성의 새로운 기회로 바꾸어 드립니다.

지난 25년간 Bloor는 기업들의 전략에 맞추고 최선의 성과를 달성하고자 기술을 수용함으로써 기업들이 현명하게 발전할 수 있도록 도움이 되어 왔습니다. Bloor에서는 도전 시도가 지속적인 개선과 성공으로 이어지도록 도움을 드립니다.

저작권 및 고지 사항

본 문서의 저작권은 © 2019 Bloor에 있습니다. 본 발간물의 내용 일체는 Bloor Research의 사전 동의 없이는 그 어떠한 방식으로든 재생산될 수 없습니다.

본 자료의 특성상 수많은 하드웨어 및 소프트웨어 제품의 이름이 언급되어 있습니다. 전부는 아니더라도 대부분의 경우 이러한 제품명은 해당 제품을 제조하는 기업의 상표입니다. Bloor Research에서는 해당 제품명 또는 상표를 Bloor Research의 것으로 주장하려는 의도를 가지고 있지 않습니다. 이와 마찬가지로, 기업 로고, 그래픽 또는 스크린샷은 해당 소유권자의 동의하에 재생산되었으며 이는 해당 소유권자의 저작권에 해당됩니다.

발행인은 본 문서를 작성하는 데 정확한 정보를 전달하려 최선의 노력을 다했으나, 오류 또는 누락 사항에 대해서는 책임지지 않습니다.



Bloor Research International Ltd
20-22 Wenlock Road
LONDON N1 7GU
United Kingdom

전화: +44 (0)20 7043 9750
웹사이트: www.Bloorresearch.com
이메일: info@Bloor.eu