

MARKET PERSPECTIVE

Korea Big Data and Analytics Competitive Landscape and Vendor Analysis, 2018

Sang-Hyuk Doh

EXECUTIVE SNAPSHOT

FIGURE 1

Executive Snapshot: Korea Big Data and Analytics Competitive Landscape and Vendor Analysis

Korea Big Data and Analytics Competitive Landscape and Vendor Analysis 보고서에서는 국내 빅데이터 분석 시장의 주요 솔루션 벤더 및 서비스 사업자의 비즈니스 현황과 제품 그리고 시장 전략을 살펴본다. 그리고, 국내 빅데이터 분석 시장의 주요 벤더간 경쟁 관계에 대해 빅데이터 분석 벤더의 인공지능 솔루션 확대, 대형 SI 업체의 빅데이터 솔루션 개발 및 공급, 클라우드 기반의 빅데이터 분석 및 인공지능 벤더의 시장 선도, 데이터 가속화 솔루션을 공급하는 벤더 부각 등의 세부 주제로 분석하였다.

Key Takeaways

- 지속적으로 변화하는 비즈니스 환경 때문에 데이터를 실시간으로 수집하고 분석하는 것이 과거보다 중요해졌다.
- 빅데이터 분석 벤더는 솔루션의 영역을 확대하여 머신러닝 및 딥러닝 분석 솔루션을 제품에 추가하고, 데이터 분석 과정을 개선할 수 있는 데이터 사이언티스트 도구를 발표해왔다.
- 대형 SI 업체의 빅데이터 솔루션의 시장 공략 전략은 관계사를 중심으로 수요 확대에 나서는 한편, 공공 빅데이터 및 인공지능 서비스 시장에도 적극적으로 참여하는 것으로 보인다.
- 퍼블릭 클라우드 사업자는 빅데이터 영역에서도 시장 선도적인 기술을 공개하면서 인공지능 솔루션 시장과 빅데이터 분석 시장의 발전을 주도하고 있다.

Recommended Actions

- 비즈니스 인텔리전스 도구의 유용성을 높이는 가장 좋은 방법은 분석 워크 플로우의 모든 단계에 기계 학습을 적용하여 자동화 수준을 향상시키는 것이다. 시각화 또는 고급 분석 기술을 사용하는 모든 단계에서 데이터 분석 제안 기능을 제공하면 사용자 분석 환경을 개선할 수 있다.
- 고객이 위치한 지역의 데이터에 관한 규제 및 규정을 준수해야 하는 의무도 높아지고 있다. 자체 데이터센터나 클라우드 사업자의 데이터센터에 상관없이 기업은 원활한 데이터 접근이 가능하기를 원하고 있다.
- 인지/인공지능 소프트웨어 플랫폼 벤더는 인공지능 툴과 플랫폼을 개발에 이용하기 쉽고, 확장하기 쉬운 형태로 공급해야 하고, 폭넓게 이용되고 있는 오픈소스 소프트웨어 기반의 딥러닝과 머신러닝 프레임워크, 서비스를 수용하여 개발에 반영해야 한다.

Source: IDC, 2018

NEW MARKET DEVELOPMENTS AND DYNAMICS

디지털 트랜스포메이션은 비즈니스의 형태를 변화시키며, 모든 산업의 기업과 소비자에 폭넓게 영향을 미치고 있다. 지능형 데이터를 기존 비즈니스에 통합하는 것은 디지털 트랜스포메이션의 주요 과제 중의 하나이다. 데이터 기반 비즈니스를 효율적으로 운영하기 위해서는 데이터를 수집하고, 모니터링하며, 수집한 데이터를 학습시키는 것이 중요하다. IDC는 디지털화된 콘텐츠를 생성, 캡처, 복사하는 등의 과정에서 생성되는 모든 데이터를 Global Datasphere 라고 한다. IDC는 Global Digital Sphere가 2018년에 33 제타바이트로 증가하고, 2025년에는 175 제타바이트까지 확대될 것으로 예측하고 있다.

비즈니스 환경이 실시간으로 변화하기 때문에 데이터 흐름을 실시간으로 수집하고 분석하는 것이 과거보다 중요해졌다. 데이터를 실시간으로 통합하는 솔루션과 연동하여 고급 분석, 기계 학습 기반의 분석을 실시하고 이를 바탕으로 의사 결정을 실시간으로 수행 할 수 있다. 아직도 많은 데이터 통합 및 분석 과정이 배치작업을 통해 수행하고 있지만 오늘날의 비즈니스 및 데이터 환경에서는 데이터가 준실시간으로 처리되지 않으면 이것으로 인해 전체 의사 결정 과정이 늦어질 수 있다. 배치 프로세싱 중심에서 실시간 데이터 분석으로 전환하기 위해서는 중요한 아키텍처 및 프로그래밍 방식의 변경도 중요하지만, 보다 구체적인 데이터 통합도 꾸준히 이루어져야 한다. 이미 저장된 배치 데이터는 병렬 스트림 데이터와 통합해야 하며, 데이터 품질을 높이기 위하여 스트리밍 데이터를 포함한 전체 데이터를 정제하고 필터링해야 한다.

최근 인공지능 솔루션 시장의 확대는 분석 대상이 되는 데이터의 증가와 다양한 오픈소스 기반 인공지능 개발 소프트웨어 프레임워크의 발전, GPU 기반의 가속화 서버의 성능 증가와 같은 요소가 전반적으로 기술적인 뒷받침이 되었기 때문이다. 국내 인공지능 시장을 세부 기술에 따라 서버, 스토리지를 포함한 하드웨어 시장과 IT 서비스 시장, 소프트웨어 시장으로 구분할 수 있다. 서버 시장에서는 엔비디아와 GPU와 함께 NVLink 2.0 기술을 적용하여 CPU와의 데이터 통신 속도를 개선한 가속화 서버가 공급되고 있다. 가속화 서버 시스템은 각종 오픈소스 기반 딥러닝 프레임워크에 최적화하였고, 기업에서 쉽게 개발할 수 있도록 인공지능 개발툴도 함께 제공되고 있다. 국내 스토리지 시장에서도 엔비디아가 함께 설계한 데이터 아키텍처 환경을 통해 데이터 사이언티스트가 기업 환경에 맞는 인공지능 스토리지 인프라를 구성하고 운영할 수 있는 환경을 제공되고 있다. IT 서비스 업계에서는 SI 기반 비즈니스를 통해 쌓은 경험을 바탕으로 인공지능 시스템 구축에서도 고객의 요구 사항을 수용하여 이에 맞춘 최적화 솔루션을 제공할 수 있다는 강점을 기반으로 시장 공략에 나서고 있다. 또한 GPU 서버, 음성 인식, 비전 검사 등의 분야에 강점을 가지고 있는 솔루션 벤더와 에코시스템 강화에도 나서고 있다. 인공지능 솔루션을 제공하고 있는 소프트웨어 벤더는 클라우드 기반의 인공지능 플랫폼을 제공하며 음성, 비전, 딥러닝 관련 API를 제공하며 시장을 확대하고 있다.

빅데이터 분석 벤더의 인공지능 솔루션 확대

빅데이터 분석 솔루션 전문 벤더는 데이터 수집, 이동, 처리, 분석 등의 기존 빅데이터 분석 솔루션의 영역을 확대하여 머신러닝 및 딥러닝 분석 솔루션을 추가하고, 데이터 분석 라이프 사이클을 통합 관리하고, 데이터 분석 과정을 개선할 수 있는 데이터 사이언티스트 도구를 지속적으로 시장에 발표해왔다. 한편 기업용 애플리케이션 벤더는 솔루션에 저장된 데이터의 활용도를 높이기 위하여 애플리케이션의 데이터를 인공지능 기반의 분석 모델과 연동하려는 노력을 지속하였다.

SAS 코리아는 2018년 9월 인공지능 해석력과 투명성을 강화한 SAS Viya의 최신 버전을 공개하였다. SAS Viya 최신 버전은 투명한 인공지능 의사결정, 인공지능 기반 지능형 태깅 자동화 기술을 통해 강화된 개인정보보호, 최신 인공지능 기술 통합 등의 기능을 제공한다. 데이터의 성격을 찾아내는 지능형 태깅 기술은 개인정보를 자동으로 식별하고 개인정보보호 정책에 맞도록 분류해 유럽 GDPR과 같은 관련 규제에 효율적으로 대응할 수 있도록 해준다. 파이썬, R과 같은 오픈소스 기반의 통계 소프트웨어에 만든 모델을 SAS Viya와 연계할 수 있다.

SAS 코리아는 SAS 비주얼 데이터 마이닝 엔드 머신을 NH 농협은행에 공급했고, 이를 통해 개인 및 기업 고객의 데이터를 효과적으로 탐색하고, 고객별 상품 추천, 고객 이탈 가능성 등을 사전에 관리할 수 있는 빅데이터 분석 플랫폼을 구축하였다. KB 손해보험은 SAS Viya의 인공지능 기술을 활용하여 고객 관리, 고객 맞춤형 상품 개발, 보험 사기 탐지 등에 필요한 빅데이터 분석 역량을 향상시켰다.

2009년 하둡이 생겨나고, 하둡 창시자인 더그 커팅이 최고 아키텍트로 참여했던 클라우드라가 2009년 설립되었으며, 2011년에는 하둡 기술 개발에 많은 기여를 했던 야후의 데이터 분석 조직이 분사하여 호튼웍스가 설립되었다. 함께 하둡 에코시스템을 확대하며 솔루션 개발에 기여했고, 벤처캐피탈 사이에서 주목 받던 유니콘 기업 중에 하나였던 두 회사는 호튼웍스가 2015년에 그리고, 클라우드라가 2017년에 나스닥에 상장했다. 2018년 10월 하둡 전문 기업이었던 두 회사는 합병을 발표했다. 클라우드라는 데이터웨어하우스 부문의 중요한 기업 도입 사례를 가지고 있으며, 인공지능 및 머신러닝에 대한 솔루션을 보유하고 있고, 호튼웍스는 클라우드라에 비해 IoT 엣지 컴퓨팅에 대해 많은 투자를 해 왔기 때문에 이러한 서로 다른 장점을 보유한 두 기업의 합병은 충분한 시너지가 예상되지만, 오랜 기간동안 빅데이터 분석 부문의 경쟁 관계로 공통 부문에 대한 투자를 정리하는 것은 쉽지 않을 것으로 예상된다. 머신러닝과 관련하여 클라우드라는 CDSW(Cloudera Data Science Workbench)를 보유하고 있다. CDSW는 기업의 머신러닝 관련 전체 라이프사이클을 통합 관리할 수 있는 솔루션으로 데이터 탐색, 분석, 모델링, 시각화 등을 관리할 수 있다. 클라우드라 코리아는 클라우드라 세션 2018 행사에서 국내의 카카오뱅크, 케이뱅크, 한화생명 등 금융권 고객과 LG 유플러스 등의 통신 기업의 도입 사례와 함께 빅데이터 인재 양성 프로그램에 대해서 소개했다.

엠로는 구매 애플리케이션을 공급해 온 국내 소프트웨어 벤더로서 빅데이터 분석 인공지능 영역을 확대하기 위하여 인공지능 및 머신러닝 솔루션 전문 기업인 에이아이더를 자회사로 설립하였다. 에이아이더는 자동화 머신러닝(AutoML), 인공지능 추천 엔진, 머신러닝과 추론 엔진 기반의 BRMS 솔루션과 같은 머신러닝 솔루션 공급과 산업 특화된 인공지능 시스템 구축을 위한 컨설팅 서비스, 신용카드 결제 데이터 등을 가공하여 판매하는 데이터 재판매를 주요 사업 영역으로 하고 있다. 국내 기업은 인공지능 및 빅데이터 분석 사업을 진행할 때 데이터 부족, 데이터 분석 인력 부족, 프로젝트 구축 경험 부족 등으로 어려움을 겪고 있기 때문에 에이아이더는 이를 해결하기 위해 자동화된 분석 모델링 기법을 이용한 솔루션을 공급하고 있다. 에이아이더의 Smart AutoML(Automated Machine Learning)은 분석 과정을 자동화하는 솔루션으로 많은 시간이 소요되는 데이터 처리 및 모델 튜닝 과정을 자동으로 처리해 주기 때문에 코딩 작업 없이 현업 사용자도 이용이 가능하다. 해외에서도 기업용 애플리케이션 전문 업체의 인공지능 솔루션에 대한 투자는 활발하게 이뤄지고 있다. 세일즈포스는 오랜기간 인공지능 솔루션인 아인슈타인의 개발에 힘써 왔으며, 메타마인드, 릴레이 IQ, 비온드코어와 같은 인공지능 전문 기업도 인수해왔다. 세일즈포스 제품을 통해 수집된 데이터에 아인슈타인의 예측 모델을 적용하는 것처럼 엠로도 기존 구매 애플리케이션 고객의 데이터에 인공지능 기반의 예측 모델을 적용할 것으로 전망된다.

대형 SI의 빅데이터 솔루션 개발 및 공급

대형 SI 업체는 새로운 매출을 창출하기 위한 새로운 솔루션 영역 중의 하나로 빅데이터 및 인공지능 기반 분석 솔루션으로 잡고 관련 제품을 지속적으로 개발해왔다. 대형 SI 업체의 빅데이터 솔루션의 시장 공략 전략은 관계사를 중심으로 수요 확대에 나서는 한편, 공공 시장에서의 빅데이터 및 인공지능 서비스 시장에도 적극적으로 참여하는 것으로 보인다.

삼성 SDS는 브라이틱스라는 인공지능 솔루션 브랜드를 중심으로 분석형, 대화형, 시각형 솔루션이 세부적으로 나누어져 있고, 대화형 인공지능 솔루션은 브리티웍스(Brity Works) 플랫폼이라는 이름으로 출시했다. 2011년 제조 관계사 중심의 대용량 데이터 분석을 위해 시작된 솔루션 개발이 현재 브라이틱스 3.0 단계로 진화하여 데이터 시각화 분석에서 머신러닝을 통한 분석자동화 단계까지 도달해 있다. 주로 제조 관계사를 고객사로 확보하며, 브라이틱스와 진보된 빅데이터 사업을 통해 관련 매출을 확보하고 있다. 제조 관계사의 대용량 데이터 분석 시간을

1/10 로 단축하는 등의 성공 사례를 바탕으로 스마트팩토리 등의 제조 산업에서 우선적인 시장 확보에 집중하고 있다.

LG CNS 는 고급분석과 빅데이터 기술 접목을 통한 기술 개발을 기반으로 자체 AI 빅데이터 플랫폼인 DAP 를 출시하였다. DAP 은 제조산업의 비전검사(Vision Inspection)와 금융에서의 챗봇, 음성인식봇을 통한 스마트컨택센터, CCTV 나 산업안전에 적용할 수 있는 동영상인식, 자연어 Q&A 등의 기능을 제공하고 있다. 현재까지는 그룹사를 주 대상으로 직접 구축 및 솔루션 제공을 하고 있으며, 제조산업에서 품질 검수를 위한 비전검사와 금융권 대상 챗봇서비스가 상용화되어 제공되고 있다. 동영상인식 부문에서는 공공 부문 CCTV 영상분석이나 무인점포 구축을 통해 유통 산업에서의 활용 사례를 확보하고 있다.

LG 히다찌는 자체 AI 솔루션 없이 고객사에 맞는 머신러닝 및 인공지능 서비스를 개발 제공하고 있다. 추가로 일본 히다찌(Hitachi)사의 다양한 솔루션(클라우드 기반 AT/H 와 온프레미스 기반 Irene 등)의 국내 판매 및 적용을 검토중이다. 2018 년 고객 맞춤 서비스형 AI 사업 모델로 기반으로 공공 부문에서 AI 기반 차세대 보안시스템의 1 차년도 구축 사업을 수주하였고, 컨설팅 회사들과의 협업을 통해 금융부문(은행권), 의료부문 사업 발굴에 나서고 있다. LG 히다찌는 직접 영업을 통해 인공지능 부문에 대한 고객을 발굴하고, 서비스 공급도 직접하고 있다.

클라우드 기반의 빅데이터 분석 및 인공지능 벤더의 시장 선도

퍼블릭 클라우드 영역은 새로운 기술을 적용하고, 배포하는 중요한 플랫폼 중의 하나가 되어 왔다. 퍼블릭 클라우드 사업자는 빅데이터 영역에서도 시장 선도적인 기술을 공개하면서 인공지능 솔루션 시장과 빅데이터 분석 시장의 발전을 주도하고 있다. 인터넷 영역에서 발생하는 스마트폰 로그, 동영상, 사진 등의 데이터는 인터넷과 밀접하게 연결되어 있는 퍼블릭 클라우드에서 직접 데이터를 저장하고 분석하는 것이 유리하다. 따라서 퍼블릭 클라우드 영역에 저장되는 데이터는 지속적으로 증가하고, 이러한 데이터를 분석하기 위한 솔루션도 클라우드에 먼저 적용되는 사례가 많아질 것으로 전망된다.

아마존웹서비스는 매년 AWS 리인벤트 행사에서 새로운 클라우드 서비스를 공개해왔다. 2018 년 11 월 리인벤트 행사에서는 아마존 오로라와 다이노미 DB 의 새로운 기능과 새로운 종류의 데이터베이스를 발표했다. 아마존 오로라에는 글로벌 데이터베이스 기능이 추가되어서 단일 리전의 데이터베이스를 가용성을 높이기 위해 다른 AWS 리전으로 확장할 수 있는 기능이 추가됐다. 다이노미 DB 는 트랜잭션 기능을 지원하여 ACID(Atomic, consistency, isolation, durability)를 보장하는 거래를 처리할 수 있게되었고, 온디맨드 기능을 제공하여 프로비저닝이나 용량 계획을 세울 필요 없이 읽기, 쓰기 요청과 스토리지를 이용한 비용한 지불하면 되도록 설계되어 있다.

머신 러닝과 관련해서는 개발자들이 보다 쉽게 저비용 자동 데이터 라벨링 및 강화 학습과 같은 머신 러닝을 구축, 배포할 수 있도록 아마존 세이지 메이커의 기능을 고도화하였다. 인공지능 학습에 이용할 수 있도록 엔비디아 V100 GPU 8 개와 빠른 NVMe 스토리지, 100Gbps 네트워크를 갖춘 새로운 P3dn.24el 인스턴스를 공개하여 더 빠른 시간내에 더 많은 데이터를 학습 모델에 적용시킬 수 있게 되었다. 2017 년 리인벤트 행사에는 삼성전자의 다이노미 DB 사용 사례가 공개되었다. 삼성전자는 갤럭시 스마트폰 사용자의 사진, 데이터, 디바이스 설정을 안전하게 보관하기 위해 아마존 다이노미 DB 를 사용하고 있다. 삼성전자 클라우드를 사용하는 계정 수는 약 3 억개이고, 다이노미 DB 스토리지에 저장된 용량은 약 800 테라바이트이며, 데이터는 해당 국가의 아마존웹서비스 리전에 저장되고 있는 것으로 알려져 있다.

네이버 클라우드 플랫폼은 빅데이터, 인공지능 등 부가가치가 높은 서비스를 자사의 클라우드 서비스에 지속적으로 추가하며 전체적인 클라우드 서비스 포트폴리오를 확대하고 있다. 구체적으로 네이버 클라우드 플랫폼은 클로바라는 브랜드로 음성 합성, 번역, 음성 인식 등의 기능을 제공하고 있다. 또한 퍼블릭 클라우드 기반 인프라 서비스의 경우 고성능 컴퓨팅 서비스, 베어 메탈 서버, GPU 서버, 텐서플로우 서버 등 복잡한 연산을 위해 필요한 가속화 서버 관련

서비스도 보유하고 있다. 네이버 리테일 부문의 실시간 쇼핑 문의 서비스에 적용하여 단순, 반복적인 문의를 해결하는데 챗봇 솔루션을 도입하였다. 유사한 서비스는 쇼핑몰 뿐만 아니라 병원 등에서도 이용되고 있다. 2018년 8월에는 빅데이터 분석과 관련하여 빅데이터를 쉽고 빠르게 처리할 수 있는 오픈소스 기반 분석 서비스인 하둡, 컴퓨팅 노드의 수를 손쉽게 확장할 수 있고, 대용량 데이터를 저장할 수 있는 오브젝트 스토리지, 복잡한 계층의 프로그램을 모니터링하며 성능을 분석하여 문제 발생 지점과 병목 구간을 진단해 시스템 품질을 개선할 수 있도록 도와주는 핀포인트 서비스, 계정 활동 로그를 자동으로 수집하는 클라우드 액티비티 트레이서 등의 신규 서비스를 공개했다. 네이버 클라우드 플랫폼은 한국철도공사, 한국재정정보원, 한국은행, 미래에셋대우증권, 건국대병원 등에서 도입하여 사용하고 있다고 공개하였다.

한국 마이크로소프트는 퍼블릭 클라우드 서비스인 애저를 기반으로 인공지능 관련 서비스를 제공하고 있다. 마이크로소프트는 기존 윈도우 솔루션 판매보다는 애저 솔루션 판매에 집중하고 있으며 이를 통해 인공지능 솔루션 판매도 증가할 것으로 기대된다. 제조 부문에서 외부에 설치된 디지털 사이니지 장비를 관리하기 위하여 디스플레이 장비 관련 로그를 퍼블릭 클라우드를 통해 수집하고, 인공지능 기능을 통하여 장애 시점을 예측하며 이러한 데이터를 활용하여 장비가 장애가 나기 전에 사전 정비를 수행하는데 이용되고 있다. 한국 마이크로소프트는 2018년 3월 디지털 트랜스포메이션 서밋에서 SQL PaaS, 머신러닝, AI, IoT, 빅데이터 등을 포함한 애저 데이터 서비스 매출이 큰 폭으로 성장하였다고 밝히고, 애저 클라우드의 주요 도입 사례로 삼성전자 클라우드 실시간 빌딩 에너지 분석 시스템, LG CNS 스마트팩토리 솔루션의 생산관리 및 예측 분석 서비스, NH투자증권의 알고리즘 기반의 예측 기법, 현대자동차의 코나 챗봇 서비스, 아시아나 항공의 챗봇 서비스 등을 소개하였다.

데이터 가속화 솔루션을 공급하는 하드웨어 벤더의 부각

데이터 증가와 함께 고성능 컴퓨팅이 필요한 연산이 많아지면서 GPU 기반의 데이터 가속화 솔루션의 수요가 늘어나고 있다. 새로운 인공지능 학습을 수행하는 과정에는 다중 코어를 지원하는 GPU를 이용하고, 학습 과정을 통해 만들어진 모델을 이용하여 추론을 수행할 때는 FPGA, ASIC 등을 이용하여 컴퓨팅 성능보다는 빠른 응답 속도에 집중할 것이다. 물론 이러한 하드웨어 성능을 최대한 이용하기 위해서는 연산 과정을 병렬 처리할 수 있는 인공지능 프레임워크와 같은 오픈소스 기반의 소프트웨어가 중요한 역할을 했다.

엔비디아와 GPU 컴퓨팅이 지난 5년간 부상한 이유는 딥러닝과 인공지능 때문이다. 이는 인공지능 연산 과정에서 고성능 컴퓨팅이 필요하기 때문에 GPU 다중 코어를 이용한 분산 컴퓨팅이 폭넓게 이용되었고, 이를 뒷받침하기 위한 CUDA와 같은 엔비디아의 인공지능 플랫폼 개발이 있었기 때문에 가능하였다. 엔비디아 서버의 경우 Tesla V100 GPU를 추가로 16GB 메모리를 추가하여 32GB 메모리로 업그레이드하였으며, 시스템 공급 업체가 시스템 당 최대 16개의 Tesla V100 GPU를 상호 연결할 수 있는 NVSwitch도 발표하였다. 엔비디아는 빅데이터 분석과 머신러닝 용으로 개발된 래피즈 GPU 가속 플랫폼도 공개하였다. 래피즈 GPU 가속 플랫폼은 CUDA 및 관련 오픈소스와 긴밀한 협력을 통해 완성되었다. 래피즈 GPU 가속 플랫폼은 데이터 사이언스 라이브러리와 워크플로우에 원활히 통합되어 머신러닝 속도를 향상시킨다. 엔비디아는 GPU 가속 분석, 머신러닝, 그리고 향후에 제공될 데이터 시각화 기능을 위한 라이브러리를 오픈소스 형태로 제공하여 오픈소스 커뮤니티와 함께 긴밀히 협력하였다. 래피즈 GPU 가속 플랫폼은 파이썬(Python) 데이터 사이언스 툴체인에 GPU 가속을 추가하고, 아파치 애로우(Apache Arrow), 팬더스(Pandas), 사이킷-런(scikit-learn)과 같은 주요 오픈소스 프로젝트를 기반으로 구축되었다.

한국 HPE는 기업이 HPE 솔루션을 도입하여 쉽게 인공지능 프로젝트를 시작할 수 있도록 활용하는 방법을 안내하여 도입 문턱을 낮추는 전략으로 시장을 공략하고 있다. HPE의 대표적인 가속 서버는 HPE 아폴로 6500 Gen 10으로 해당 제품은 GPU 워크로드 성능 가속에 초점을 맞춰 125 테라플롭스의 성능을 낼 수 있고, GPU 클러스터 연결 기술인 NVLink2.0을 이용하여

고대역의 데이터 연결을 제공한다. 서버당 8 개의 GPU 를 장착할 수 있고, PCIe 인터페이스 기반의 GPU-CPU 조합을 4 대 1 또는 8 대 1 로 설정 가능하다. HPE 에서는 금융권에서 요구하는 다양한 서비스를 수용하기 위해서 기존의 획일적인 솔루션 제공에서 벗어나 유연하고 확장 가능한 인프라 수요에도 대응하고 있다.

한국 IBM 은 2018 년 파워 9 프로세스 기반의 차세대 서버를 공개했다. 파워 9 기반 AC922 시스템은 엔비디아 테슬라 V100 GPU 와 파워 CPU 를 통합해 대규모의 인공지능 워크로드 지원에 효율적이다. IBM 은 기존 인텔 제온 E5-2640 v4 보다 딥러닝 프레임워크 학습 시간을 개선할 수 있다고 주장하고 있다. 새로운 파워 시스템도 엔비디아와 함께 개발한 NVLink 2.0 기술을 적용하여 CPU 와의 데이터 통신 속도를 개선하였다. IBM 은 파워 시스템에 각종 오픈소스 기반 딥러닝 프레임워크를 최적화하여 기업이 IBM 파워시스템 기반으로 손쉽게 개발할 수 있도록 하였다. IBM 은 범용 인공지능을 제공하는 왓슨 제품과 중앙 진단 등에 최적화된 왓슨 헬스케어 제품을 보유하고 있다. 국내에서는 인공지능 기반의 어시스턴트 제품과 지식 스튜디오 관련 도입 사례를 확보하고 있다.

퓨어스토리지는 2018 년 2 분기 엔비디아 기반의 인공지능 인프라 아키텍처를 제공하는 에이리(AIRI)를 국내에 소개했다. 퓨어스토리지와 엔비디아가 함께 에이리를 설계하여 데이터 아키텍트 및 데이터 사이언티스트가 기업 환경에 맞는 AI 인프라를 구성하고 운영할 수 있는 환경을 제공하고 있다. 에이리는 퓨어스토리지 플래시블레이드와 4 대의 엔비디아 DGX-1 서버로 구성된다. 이중 플래시블레이드는 데이터 분석 및 인공지능 연산을 위해 설계된 스토리지 플랫폼이다. 퓨어스토리지는 국내 반도체 기업에서 공정을 개선하기 위해 스토리지를 도입한 사례를 가지고 있다. 넷앱은 2018 년 4 분기 데이터 패브릭을 통해 기업 환경에서 인공지능 데이터 연산을 효과적으로 지원하기 위한 스토리지 서비스 플랫폼을 출시했다. 넷앱이 제공하는 플래시 성능 보장 서비스를 이용하여 넷앱 데이터 패브릭을 엣지, 코어, 클라우드로 확장하고 이를 통해 기업이 낮은 지연 시간을 통해 인공지능 애플리케이션을 운영할 때 서비스 응답 시간을 개선할 수 있을 것으로 보인다. 제조 부문에서 산업 현장에서 발생하는 수많은 데이터를 수집하고 분석하여 제조 공정의 효율화를 향상하기 위해 넷앱의 스토리지를 도입하고 있는 것으로 판단된다.

ADVICE FOR THE TECHNOLOGY SUPPLIER

국내에서도 빅데이터 분석이 모든 산업에서 상당한 영향을 끼치고 있는 것을 산업별 주요 사례로 알 수 있다. 기업은 빅데이터 및 분석 솔루션 도입을 통해 수익과 성과를 개선하는 동시에 고객 대응 능력과 잠재 고객 확보에 큰 효과가 있기를 기대한다. 이러한 기업의 기대를 충족하는 것도 중요하지만 빅데이터 분석 솔루션 벤더들은 데이터 개인 정보 보호 및 보안 문제에 대해 사전 예방적인 접근법을 취해야 한다. 리스크에 대비한 계획 수립, 시나리오 매핑 및 보안 사고에 대비한 사전 훈련을 주요 위험 업무 전반에 걸쳐 실시하여야 한다. 빅데이터 분석 솔루션 벤더는 데이터 사이언티스트와 같은 기술적 마인드, 넓은 맥락을 창조적으로 포착하는 능력과 비선형적 사고를 가진 사람들로 구성된 데이터 사이언티스트 팀을 갖추는 것이 중요하다.

비즈니스 인텔리전스 도구의 유용성을 높이는 가장 좋은 방법은 분석 워크 플로우의 모든 단계에 기계 학습을 사용하여 자동화 수준을 향상시키는 것이다. 시각화 또는 고급 분석 기술을 사용하는 모든 단계에서 데이터 분석 제안 기능을 제공하면 사용자 분석 환경을 개선 할 수 있다. IT 주도의 데이터 분석 환경을 개선하기 위하여 비즈니스 분석가가 IT 승인 없이도 작업 할 수 있는 셀프 서비스 도구 및 오픈 소스 도구가 생겨났다. 빅데이터 솔루션 벤더는 오픈 소스 툴을 통합하고 API 를 제공하여 다른 툴과 통합할 수 있는 기능을 제공해야 한다. 오픈소스 분석 도구를 이용하는 사용자가 엔터프라이즈 도구로 전환하도록 강요하는 대신 셀프 서비스 도구를 채택한 경우에는 이를 중앙 IT 조직에서 관리할 수 기능을 사용할 수 있도록 해야 한다.

클라우드 사업자는 기업 환경의 데이터 저장 및 데이터 서비스에 있어서 중요한 경쟁자로 부상했다. 대형 클라우드 사업자는 대부분의 기업의 데이터 센터보다 더 많은 자원, 규모, 보안 및

성능을 제공하고 있다. 클라우드 사업자의 글로벌 네트워크를 통해 기업은 전 세계에 서비스를 제공할 수 있고, 클라우드 서비스에 저장된 데이터를 기업의 다양한 리소스가 활용할 수 있으며, 향후에는 인공지능 기능에도 활용할 수 있을 것이다. 서비스 응답 속도를 개선하고, 고객 만족도를 향상시키기 위해서는 데이터에 대한 글로벌 접근 방식을 개선하는 것이 중요하다. 이때 고객이 위치한 지역의 데이터에 관한 규제 및 규정을 준수해야 하는 의무도 높아지고 있다. 자체 데이터센터나 클라우드 사업자의 데이터센터이든 상관없이 기업은 원활한 데이터 접근이 가능하기를 원하고 있다.

기업의 비즈니스가 이동 중에도 실시간으로 구동할 수 있도록 지능형 데이터가 구현되어야 한다. 실시간 의사 결정을 내리기에는 엔드포인트에서 코어까지 데이터가 이동하는데 걸리는 시간을 단축하는 것이 중요하다. 따라서 엔터프라이즈 엣지에서 데이터를 처리하면 이러한 소요 시간을 단축하는데 도움이 된다. 향후에는 엣지에서 데이터 분석, 지능형 데이터 저장 등의 중요한 작업을 처리하게 될 것이다. 기업은 데이터 준비 상태를 개선할 때 경영진이 프로젝트를 주도하고, 관련 예산 책정, 데이터 분석에 익숙한 직원 채용 등을 프로젝트 팀에서 직접 실행할 수 있도록 권한을 부여해야 한다. 기업 내부의 많은 IT 조직이 물리적 보안 뿐만 아니라 제조 현장, 운영 데이터의 보안을 책임지고, 별도로 보관되고 있는 데이터 저장소를 통합하는 업무도 책임지고 있다. 디지털 트랜스포메이션을 이루기 위해서는 새로운 기술 뿐만 아니라 새로운 역량을 확보하고, 경영진 및 비즈니스 인력과의 원활한 관계 유지가 필요하다.

인지/인공지능 소프트웨어 플랫폼 벤더는 인공지능 툴과 플랫폼을 분석 모델 개발에 이용하기 쉽고, 확장하기 쉬운 형태로 공급해야 한다. 기업 환경에서는 아마존의 세이지메이커나 마이크로소프트의 애저 머신 러닝 스튜디오와 같은 인공지능 기반 애플리케이션을 만들 수 있는 개발 스튜디오와 관련 인공지능 API 서비스를 필요로 하고 있다. 이러한 인공지능 개발 스튜디오는 제품 추천, 가격 최적화, 예지정비 등과 같은 인공지능 기반 애플리케이션을 만들 수 있도록 도움을 준다. 인공지능 솔루션 및 플랫폼을 제공하는 벤더는 이미지 인식과 분류, 언어 인식, 녹음기, 기계학습 기반 텍스트 받아적기, 기계 학습 기반 텍스트 분석 기능과 AI 프레임워크 등을 제공하고 있다. 이러한 인공지능 관련 툴을 이용하면 광범위한 데이터 학습이나 개발없이 인공지능 관련 기능을 기존 애플리케이션에 쉽게 연결할 수 있다.

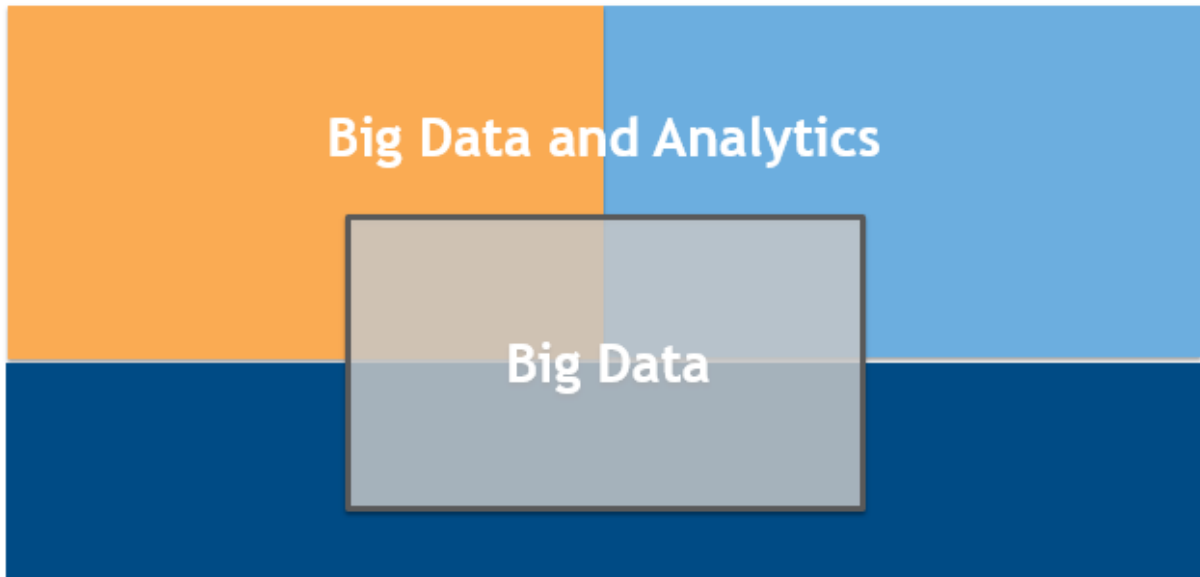
인공지능 솔루션 벤더는 시장에서 폭넓게 이용되고 있는 오픈소스 소프트웨어 기반의 딥러닝과 머신러닝 프레임워크, 서비스를 수용하여 솔루션에 개발에 반영해야 한다. 많은 기업은 텐서플로우, Caffe 와 같은 인공지능 프레임워크를 기업의 인공지능 애플리케이션 개발을 위해 사용한다. 그러나 기업은 오픈소스에서 제공되지 않는 기업환경에서 필요한 추가적인 기능과 기술을 필요로 하기 때문에 다양한 추가적인 기술을 제공하는 인지/인공지능 플랫폼이 시장에서 서비스되어야 한다. 오픈소스 모델을 추가하기 위해 인공지능 솔루션 벤더는 확장 가능한 프레임워크를 설계하고, 이 인공지능 프레임워크를 이용하는 개발자가 유연하게 개발할 수 있도록 환경을 제공해 주어야 한다 .

MARKET DEFINITION

IDC 는 빅데이터 기술을 다양한 형태(Variety)의 대용량(Volume) 데이터를 고속(Velocity)으로 수집, 처리, 분석함으로써 가치(Value)를 비용효율적으로 도출해낼 수 있도록 설계된 새로운 세대의 기술 및 아키텍처로 정의한다. 빅데이터는 개별 기술이 아닌 특정 유스 케이스의 워크로드로 정의된다. 예를 들어 빅데이터 기술에는 하둡 또는 NoSQL 데이터베이스 뿐만아니라 MPP 형태의 RDBMS 도 포함된다. 이러한 기술이 빅데이터 분석 워크로드를 처리하는데 적용이 가능하다면 해당 소프트웨어 기술은 빅데이터 시장에 포함된다. 그림 3 과 같이 빅데이터 및 분석(Big Data and Analytics, BDA) 시장은 빅데이터 시장을 포함한다. 빅데이터 시장은 넓은 의미의 빅데이터 및 분석 시장의 일부라고 할 수 있다.

FIGURE 3

Big Data as a Subset of the Big Data and Analytics Market



Source: IDC, 2018

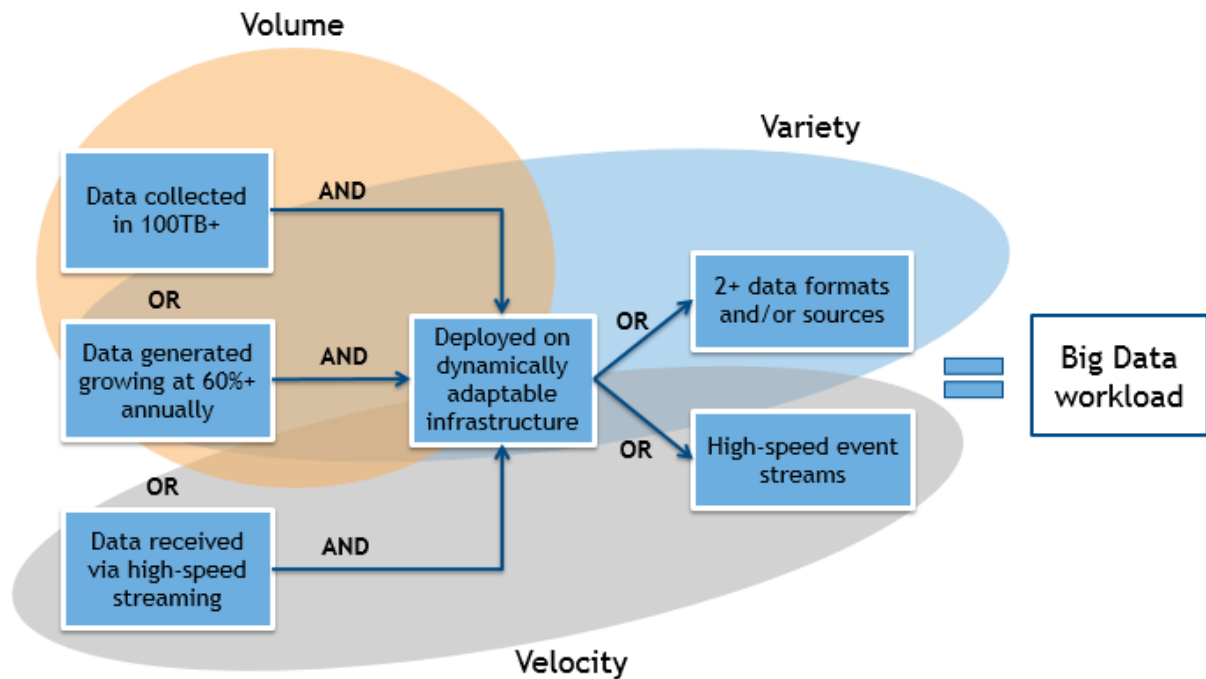
IDC 는 데이터 규모, 다양성, 데이터 속도와 같은 3 가지 데이터의 속성을 연결하여 빅데이터를 정의한다. 그림 4 는 유스 케이스 및 관련 기술과 서비스가 빅데이터 시장에 포함되기 위한 IDC 의 기준이다. 유스 케이스를 평가하기 위한 기준과 단계는 다음과 같다.

- 1 단계는 세 가지 시나리오를 기반으로 유스 케이스를 평가한다.
 - 수집된 데이터가 100TB 를 초과하는 경우(데이터를 디스크에 저장하기보다는 수집하여 인메모리를 기술을 이용하기 위하여 수집하는 경우도 포함)
 - 실시간 스트리밍 데이터 통합 및 분석을 위한 고속 메시징 기술 이용(현재 IDC 는 60Gbps 또는 그 이상의 속도로 정의하며, 이러한 경우를 데이터 인 모션(Data in Motion)이라고 하며 저장되어 있는 데이터(data in use)와 구분된다.
 - 현재 데이터 규모가 크지 않지만, 매년 60% 이상의 속도로 매우 빠르게 성장하는 경우

- 2 단계는 1 단계의 3 가지 시나리오 각각에 대해 기술이 동적으로 적용 가능한 인프라에 배포되지 여부를 평가한다.
 - IDC 는 빅데이터 인프라의 기준을 일부 수정하여 스케일 아웃 형태의 확장형 인프라를 필수 요소에서 제외하고 대신 동적 적응 인프라를 사용하는 경우를 주요한 요소로 포함시켰다.
- 3 단계에는 다음과 같은 두 가지 시나리오를 포함한다.
 - 두 개의 이상의 데이터 형식 또는 데이터 소스를 이용하는 경우
 - 클릭 스트림 트래킹 또는 머신 데이터를 모니터링 하는 경우와 같은 고속으로 발생하는 데이터인 경우

FIGURE 4

IDC's Big Data Criteria



Source: IDC, 2017

Technology View of Big Data and Analytics

IDC 가 빅데이터 및 분석 시장에 대해서 다루고 있는 기술 그룹과 카테고리는 아래 표와 같다.

TABLE 3

Big Data and Analytics Technology Definitions

| Technology Group | Technology Category | Description |
|------------------|---|---|
| Hardware | Server and storage | <p>A server is a computer or device on a network that manages network resources. For example, a file server is a computer and storage device dedicated to storing files. Any user on the network can store files on the server. A print server is a computer that manages one or more printers, and a network server is a computer that manages network traffic. A database server is a computer system that processes database queries.</p> <p>Storage is the part of a computer system, connected system, or peripheral device that stores information for subsequent use or retrieval. It can take the form of storage, which is an integral component of functional computer systems, or additional systems and devices. IDC include incremental spending on external storage systems as internal storage is already counted in the server market values.</p> |
| Services | IT services | IT services include IT consulting, systems and network implementations, IT outsourcing, application development, IT deployment and support, and IT education and training related to BDA software and infrastructure spending. IT services also involve helping buyers create the IT strategy of their overarching BDA journey. |
| | Business services | Business services comprise business consulting and business process outsourcing related to BDA software and hardware. |
| Software | Business Intelligence and Analytics Tools and Platforms | This software includes five segments. Query, reporting, and analysis tools include ad hoc query and multidimensional analysis tools. Advanced and predictive analytics tools include data mining and statistical software. Content analytics |

TABLE 3**Big Data and Analytics Technology Definitions**

| Technology Group | Technology Category | Description |
|------------------|---|--|
| | | systems provide tools for recognizing, understanding. Search systems include departmental, enterprise, and task-based search and discovery systems. Spatial information and location analytics is a segment of the broader spatial information management (SIM) software. Cognitive/AI software platforms. |
| | Analytic and Performance Management Applications | This software represents prepackaged analytic applications. IDC defines these applications as software that must meet each of the following three conditions: Business process support, Separation of function, Time-oriented, integrated data from multiple sources |
| | Analytic Data Management and Integration Platform | IDC defines analytic data management and integration platform as a set of tools for the creation and ongoing management of data repositories to support analytic workloads. |

Source: IDC, 2018

Related Research

- IDC's Worldwide Big Data and Analytics Software Taxonomy, 2017 (DOC #US42353216, MAR. 2017)
- Korea Big Data and Analytics 2017-2021 Forecast, 2017 (DOC # AP43433618, Jan. 2018)
- Korea Artificial Intelligence Forecast, 2018-2022 (DOC # AP42668218, Nov. 2018)

Synopsis

Korea Big Data and Analytics Competitive Landscape and Vendor Analysis 보고서에서는 국내 빅데이터 분석 시장의 주요 솔루션 벤더 및 서비스 사업자의 비즈니스 현황과 제품 그리고 시장 전략을 살펴본다. 그리고, 국내 빅데이터 분석 시장의 주요 벤더간 경쟁 관계에 대해 빅데이터 분석 벤더의 인공지능 솔루션 확대, 대형 SI 업체의 빅데이터 솔루션 개발 및 공급, 클라우드 기반의 빅데이터 분석 및 인공지능 벤더의 시장 선도, 데이터 가속화 솔루션을 공급하는 하드웨어 벤더 부각 등의 세부 주제로 분석하였다.

IDC 에서 국내 빅데이터 분석 및 인공지능 시장 리서치를 담당하는 도상혁 책임 연구원은 “지속적으로 변화하는 비즈니스 환경에서는 빠른 의사 결정을 내리는 것이 중요하기 때문에 데이터를 실시간으로 수집하고 분석하는 것이 과거보다 중요해졌다. 빅데이터 분석 솔루션을 제공하는 벤더는 새로운 인사이트를 제공하고, 데이터 정확도를 높이기 위하여 딥러닝 및 머신러닝 기반의 분석 솔루션을 제공하고 있다. “ 그러나, 기업은 딥러닝 및 머신러닝 기반 분석을 수행하거나 관련 솔루션을 직접 개발할 인력은 부족한 상태라고 말하며, “빅데이터 분석 및 인공지능 솔루션을 제공하는 벤더는 데이터 분석 과정을 개선하기 위해 데이터사이언티스트 토크를 제공하고, 데이터 분석을 대행할 수 있는 데이터 사이언티스트 팀을 운영해야 한다”고 언급했다.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC Korea

Suite 1404, Korea Trade Center
159-1, Samsung-Dong, Kangnam-Ku
Seoul, Korea, 135-729
822.551.4380
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or web rights.

Copyright 2018 IDC. Reproduction is forbidden unless authorized. All rights reserved.

