

4차 산업혁명 시대의 핵심 ICT 기술: 빅데이터, 인공지능, 클라우드 기술 동향

김대엽·김영배 (펜타시스템테크놀로지)

목 차	1. 서 론
	2. 빅데이터 기술동향
	3. 인공지능 기술동향
	4. 클라우드 기술동향
	5. 결 론

1. 서 론

4차 산업혁명은 기술 융합의 시대이며, 기술의 융합은 실제 다양한 산업 분야에서 예상치 못한 혁신적 결과를 불러일으키고 있다. 그 가운데 빅데이터, 인공지능, 클라우드 등의 최신 ICT 기술들은 가히 비약적으로 발전하였으며, 4차 산업혁명의 시대를 이끄는 핵심적인 요소 기술로 인식되고 있다.

빅데이터는 소셜 미디어의 활성화로 인해 이미 오래 전부터 그 개념이 등장하였고, 말 그대로 거대 데이터를 다루고 그로부터 유의미한 가치를 도출하기 위한 다양한 기술적 노력들이 진행되어왔다. 이제는 소셜 미디어 데이터뿐만 아니라 전에는 활용되지 않던 다양한 유형의 데이터들이 빅데이터라는 범주 안에 포함되어 새로운 가치를 생산해내고 있다. 수년 전까지만 해도 빅데이터는 최신 ICT 기술의 흐름을 대표하는

요소로서 인식되었으나 현재는 최신 기술의 중심에 자리하지 않는다. 빅데이터는 이제 시대를 대표하는 핵심 기술이 아니라 다른 기술들과 융합되어 더 큰 가치를 선사하는 근간이자 기반 기술로서 인식되고 있다.

인공지능은 현재 가장 뜨겁게 주목받는 기술 중 하나로 이미 여러 산업 분야에서 다양한 영향력을 행사하고 있다. 각종 지능형 서비스를 창출하고 있는 인공지능은 높은 처리성과 대규모의 학습 데이터를 필요로 하는 기술적 특성 상 과거에는 현실에 적용하기 어려운 기술이었으나, 현재는 고성능 컴퓨터와 빅데이터의 등장으로 인해 전혀 새로운 국면을 맞이하고 다양한 형태로 발전하고 있다. 음성 및 영상 인식, 챗봇과 같은 질의응답 서비스, 자율주행 등과 같은 여러 응용 사례들이 인공지능 기술을 통해 등장하고 있으며, 인공지능 기술을 기반으로 한 지능형 서비스들의 품질은 날로 향상되고 있다.

클라우드 기술 역시 최근 들어 새롭게 등장한 기술은 아니다. 클라우드 기술은 인프라, 플랫폼, 소프트웨어 등의 자원을 가장 효율적으로 활용하는 방법을 제시함으로써 개발자들뿐만 아니라 일반 사용자들에게까지 다양한 가치의 서비스를 제공하는데 활용되고 있다. 아마존 웹 서비스(AWS), 마이크로소프트, 구글 등의 글로벌 기업들은 이미 오래 전부터 클라우드 서비스의 발전과 활용 가능성을 인식하고 다양한 서비스를 구축 및 제공하고 있다.

본고에서는 4차 산업혁명 시대의 핵심으로 손꼽히는 빅데이터, 인공지능, 클라우드 각각의 기술들에 대해서 그 발전 동향을 소개하고자 한다.

2. 빅데이터 기술동향

빅데이터의 기술동향은 크게 하둡을 중심으로 한 인프라 관련 기술과 데이터 분석 관련 기술의 두 가지 영역으로 나누어 설명할 수 있다.

먼저 빅데이터 인프라 기술의 동향에서 하둡의 변화를 살펴보고자 한다. 하둡은 분산 환경에서 빅데이터를 저장하고 처리할 수 있는 자바 기반의 오픈소스 프레임워크이다. 하둡은 데이터 저장을 위한 하둡 분산 파일시스템 (Hadoop Distributed File System, HDFS)과 대용량 데이

터의 처리를 위한 분산 프로그래밍 모델인 맵리듀스(Map-Reduce)로 구성된다. 이러한 하둡의 핵심 영역과 다양한 하둡 에코 시스템이 조합되어 하나의 빅데이터 플랫폼을 구성한다. 2006년 하둡이 처음 발표된 이후 하둡과 하둡 에코시스템에 대한 프로젝트들이 지속적으로 개발 및 발전되고 있다.

하둡 프로젝트에서 최근 주목할 만한 사실은 하둡 3.x 버전의 등장이다. 하둡 3.x 버전은 지난 2016년 9월 3일에 발표된 하둡 3.0 알파 버전이 공개된 이후 2018년 4월 6일 하둡 3.1.0 정식 버전이 배포되었으며, 현재 서브 버전들이 계속해서 개발되고 있다. 하둡 3.x 버전에서 가장 주목할 만한 특징은 이레이저 코딩(erasure coding) 기술과 네임노드 다중화 구성이라고 볼 수 있다. 이레이저 코딩은 기존 HDFS의 복제방식에 따라 3배의 저장 공간이 필요했던 부분을 개선하여 약 1.4배의 저장 공간만을 사용하는 방식으로 최적화 한 것이다. <표 1>은 하둡 3.x 버전의 주요 개선사항을 요약한 것이다[1].

빅데이터 인프라 관점에서 주목할 만한 또 다른 동향은 클라우드 환경에 빅데이터 시스템을 배치하는 것이다. 2018년까지 기업의 40%가 빅데이터 어플리케이션용 스토리지를 퍼블릭 클라우드 스토리지 형태로 구축할 예정이라는 조사

<표 1> 하둡 3.x 버전의 주요 추가 · 개선사항

개선사항	설명
Erasure Coding 기술 적용	데이터 저장 시 3배의 저장 공간이 필요한 복제방식(replica 3)에서 약 1.4배의 저장 공간으로 최적화하는 기술
YARN timeline service v.2	v.1과 비교하여 데이터의 읽기/쓰기를 HBase를 이용하여 분산 처리함으로써 확장성과 신뢰성 확보
네임노드 다중화 구성 지원	하둡 2.x 버전에서 HDFS 네임노드는 하나의 활성화된(active) 노드와 대기 상태(stand by) 노드로 구성되었지만, 하둡 3.x에서는 두 개 이상의 활성화된(active/passive) 네임노드로 운영 가능하여 fail over 시 down time 최소화 가능

결과가 보고되었으며, 조사 전문기관인 포레스터는 클라우드 기반 빅데이터 관련 제품의 수요가 기존 온프레미스(On-premise) 방식에 비해 약 7.5배 빠르게 증가할 것으로 예상하였다[2,3].

빅데이터 활용을 위해 클라우드 서비스를 도입함으로써 얻을 수 있는 장점은 일반적으로 데이터센터에서 클라우드 서비스 도입을 통해 얻을 수 있는 장점과 유사하다. 사용자 요구사항에 따라 활용 자원의 유연한 확장과 축소가 가능하다는 점, 다양한 서비스 비용 모델의 활용을 통해 온프레미스 방식에 비해 비용 절감이 가능하다는 점, 클라우드 서비스 인프라의 분산 배치를 통해 재해 복구가 유리하다는 점 등 외에도 신속성, 보안성 등 다양한 장점이 존재한다.

현재 빅데이터 관련 클라우드 서비스를 제공하는 대표적인 기업으로는 아마존 웹 서비스(AWS), 마이크로소프트(Microsoft Azure), 구글(Google Cloud Platform, GCP) 등이다. 이들은 하둡 기반의 데이터 저장, 실시간 데이터 처리 및 분석, 시각화 등을 위한 다양한 제품군을 보유하고 있으며, 고객들의 사용 편의와 활용을 보장하기 위해 각각의 고유 전략을 개발하고 있다.

클라우드 환경에서 빅데이터 활용을 고려하는 기관이나 기업은 기존 클라우드 서비스 제공업체의 제품을 사용하기도 하지만, 경우에 따라서 자체 클라우드 플랫폼을 구축하여 빅데이터 활용기반을 마련하는 경우도 있다. 자체 클라우드 환경 구축 시, 가장 많이 활용되는 플랫폼으로는 Red Hat의 오픈스택(OpenStack)이 있다. 오픈스택은 매년 활용 비중이 증가하고 있으며, 미국의 경우 100대 기업의 절반 이상이 오픈스택을 이용하여 클라우드 환경을 구축하고 있다.

빅데이터 분석 기술 동향을 살펴보면 최근 인공지능 등 기술의 발전 및 융합으로 함께 진화하고 있다. 기존의 데이터 분석기법(통계분석, 데이

터마이닝, 소셜 네트워크 분석 등)에 머신러닝, 딥러닝 등 인공지능을 활용한 분석기법이 많이 사용되고 있다. 이러한 분석을 지원하기 위한 다양한 툴도 등장하고 있다[4].

기존 분석 기법은 한두 가지 알고리즘으로 분석을 수행했으나, 최근에는 여러 분석 기법을 적용·경쟁하여 가장 우수한 모델을 선정하는 챔피언 모델이나 다양한 분석 기법을 혼합하여 최적의 모델을 만들어내는 앙상블(블렌딩) 모델의 적용이 점차 보편화되고 있다[5].

중요한 점은 빅데이터 분석 기술의 동향이 알고리즘과 같은 분석 기법에만 국한된 것은 아니라는 것이다. 이제는 분석 툴이나 인프라와 같은 분석 환경과 분석 결과의 활용전략 등 복합적인 관점에서 접근해야 할 때이다. <표 2>는 디지털 기술 전문매체에서 빅데이터 분석과 관련하여 주목할 만한 기술이나 활용 동향을 전망한 것이다[4,6,7].

<표 2>에서 주목할 만한 내용은 빅데이터 분석에서의 머신러닝 등 인공지능 기술 활용, 클라우드 기반의 빅데이터 인프라 구축이 앞으로 크게 확대될 것이라는 전망이다. 인공지능과 클라우드가 각각 빅데이터 분석 기술과 인프라 구축 기술에 관련이 있다면, 빅데이터 서비스 차원에서는 서비스로서의 인사이트(Insight as a Service)에 대한 시장의 확대를 주목할 만하다. 빅데이터 분석을 통한 서비스의 수준이 기존에는 분석 결과의 제공에 그쳤다면 이제는 분석 결과를 통한 통찰력(Insight) 및 활용계획(Action Plan)에 대한 부분까지 제공하는 개념으로 확대되고 있다는 것이다. 빅데이터를 둘러싼 이러한 기술 및 시장의 동향은 향후 민간과 공공 전 분야에 걸쳐 영향을 미칠 것으로 예상된다[8].

<표 2>에서 강조한 빅데이터 관련 세 가지 트렌드 가운데 클라우드 관련 내용은 앞 절에서 인

〈표 2〉 해외 주요 매체에서 전망한 2018 빅데이터 분석 최신동향

Forbes	CIO	TDS
<ul style="list-style-type: none"> - 대화형 인터페이스 - 인공지능 기반 의사결정 - 비정형 데이터 분석 확대 - 빅데이터 분석을 위한 클라우드 우선 전략 채택 - 인사이트 전문조직 구성 - 데이터 엔지니어들의 위상 증가 - 서비스로서의 인사이트(Insight as a Service) 시장 확대 	<ul style="list-style-type: none"> - 셀프서비스 BI도구 활용 확대 - 모바일 대쉬보드 활용 증가 - DNN 알고리즘 활용 확대 - 텐서플로우, MxNet, Cognitive Toolkit 2.0, scikit learn 등의 활용도 증가 - 클라우드 기반 저장 및 분석 활용 증가 - R, 파이썬 등 분석언어 활용 증가 	<ul style="list-style-type: none"> - 클라우드 기반 저장 및 컴퓨팅 - 신경망(NN) 기반 분석 - 분석 기반 마이크로서비스 - 데이터 분석가 및 개발자를 위한 개선된 인터페이스(R, 주피터 노트북) - 신경망 분석을 위한 도구 개선(텐서플로우, MxNet, Cognitive Toolkit 2.0, scikit learn)

3개 매체에서 전망한 최신기술 동향에서 공통적으로 주목하고 있는 것은 1) 인공지능 기술의 활용, 2) 클라우드 기반 빅데이터 활용으로 요약할 수 있으며, 서비스로서의 인사이트는 타 매체들에서 많이 언급하고 있는 최신 동향으로 주목할 만함

프라 관점으로 이미 설명한 바 있다. 그렇다면 빅데이터와 인공지능 그리고 서비스로서의 인사이트에 대해 좀 더 살펴보고자 한다.

빅데이터 분석과 활용이 전 세계 민간-공공 분야에서 보편화되어 가고 있는 상황에서 분석에 사용되는 데이터의 형태도 더욱 복잡해지고 있다. 이러한 가운데 기계학습과 같은 인공지능 기술은 보다 성공적인 분석 결과 도출을 위해 필수적인 기술로 인식되고 있다.

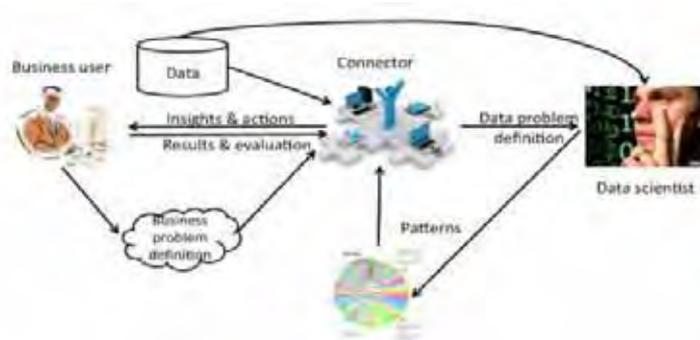
빅데이터의 활용 가치를 높이기 위해 기계학습, 딥러닝과 같은 인공지능 기술이 필요해졌다면, 인공지능 기술의 성공적 활용사례 발굴을 위해서는 충분한 데이터의 확보가 관건이다. 즉, 빅데이터와 인공지능은 상호 필수불가결한 요소로 함께 발전해 나갈 것이다.

최근 인공지능 기술은 딥러닝(심층 학습) 알고리즘을 중심으로 발전해 나가고 있다. 그 중 가장 강력한 딥러닝 알고리즘은 다층 구성의 신경망(Neural Network, NN)을 기반으로 한 DNN(Deep Neural Network) 알고리즘이다. DNN에서 신경망 계층(layer)이 많을수록 계산에 더 많은 시간이 소요되며 학습 과정도 어려워진다. 최근에는 DNN을 비교적 쉽게 활용할 수 있는 다양한 패키지들이 등장하고 있다. 전문가들

이 최근 가장 많이 사용하는 구글의 오픈소스 라이브러리인 텐서 플로우가 있으며, 그 외에 MxNet, 마이크로소프트의 Cognitive Toolkit 2.0, 사이킷 런 등이 있다[7].

한편 서비스로서의 인사이트는 빅데이터 분석 생태계의 새로운 패러다임으로 설명된다. 전 세계 기업의 66%는 비즈니스 인텔리전스(Business Intelligence, BI) 솔루션의 상당 부분(11%~75%)을 아웃소싱하고 있으며, 2018년에는 80% 이상의 기업이 통찰력을 얻기 위해 서비스로서의 인사이트 관련 업체에 의존할 것으로 전망하고 있다[6]. 서비스로서의 인사이트란 기업 또는 기관에 통찰력(insight)을 제공하고, 비즈니스 목표를 달성하기 위해 필요한 구체적인 활용 계획(action plan)까지 제공하는 서비스가 다[9]. 서비스로서의 인사이트에서 현업 담당자와 데이터 과학자 사이에서 분석 결과에 따른 통찰력 및 활용 계획을 적절히 생성 및 제공할 수 있는 전문 인력의 역할이 매우 중요 하다[10].

(그림 1)에서는 이러한 전문 인력을 커넥터(connector)라고 부른다. 커넥터는 현업 담당자를 통해 비즈니스 문제점을 이해해야 한다. 또한, 분석에 사용할 데이터를 이해하기 위해 노력해야 한다. 비즈니스 문제점과 데이터에 대한 이해



[출처: courtesy of Evangelos Simoudis]

(그림 1) Insight as a Service 프로세스

를 통해 문제를 정의하고 데이터 과학자가 이 문제를 해결한다.

데이터 과학자는 데이터에 대한 분석을 수행하고 다양한 패턴과 연관관계를 추출한다. 커넥터는 추출된 패턴과 연관관계를 분석하여 유의미한 항목들을 식별한다. 이러한 유의미한 항목을 기반으로 문제를 개선하고 비즈니스에 적용할 수 있는 활용 계획까지 수립하여 현업 담당자에게 제공한다. 현업 담당자는 활용 계획에 따른 수행 결과와 효과를 커넥터에게 피드백한다.

이처럼 커넥터는 비즈니스 문제점과 데이터를 결합하여 종합적인 문제 상황을 인식할 수 있어야 한다. 또한 비즈니스 문제점에 대해 데이터와 분석 관점에서 데이터 과학자와 적절한 의사소통이 가능해야 한다. 그리고 분석 결과를 토대로 업무적 수준에서의 적절한 해결책을 제시해야 하는 막중한 역할을 수행하는 만큼 다양한 분야에 걸쳐 전문적인 지식과 경험이 뒷받침되어야 한다.

3. 인공지능 기술동향

최근 인공지능 기술은 컴퓨팅 파워의 급격한 발전과 데이터의 폭증으로 그동안의 기술적 한

계를 극복하며 빠르게 발전하고 있다. 인공지능 기술은 전자·IT 산업뿐만 아니라 제조, 금융, 의료, 자동차, 헬스케어 등 거의 모든 산업에 그 영향력을 미치고 있으며, 다른 최신 ICT 기술과 복합적인 융합을 통해 계속해서 발전하고 있다.

글로벌 기업들은 자사의 인공지능 플랫폼을 활용한 다양한 지능형 서비스를 개발하여 시장을 점유하기 위해 노력하고 있다. 공공분야에서도 인공지능 기반의 공공행정 개선이나 공공서비스 지능화를 위해 다양한 노력을 기울이고 있다.

인공지능 시장은 앞으로 연평균 50% 이상 높은 성장세를 보일 전망이다. 특히 2018~2020년에 빠른 속도로 성장할 것으로 기대되고 있다. IDC[11], Tractica[12], Market and Markets[13] 등 시장조사 기관이 전망한 인공지능 시장 규모를 보면, 2020년 기준 IDC가 463억 달러로 가장 컸으며, Tractica는 105억 달러, Market and Markets는 56억 달러로 큰 차이를 보이고 있다. 이는 인공지능의 정의와 범위가 다르고, 아직 초기 단계로서 시장의 성장 잠재력에 대한 예측에 편차가 존재하기 때문이다[14]. 절대적인 시장 규모는 차이를 보이고 있지 만, 2020년까지 56~66% 수준의 높은 성장세를 보일 것이라는 점에서 각 기관들의 견해가 일치하고 있다는 점은

〈표 3〉 국내외 인공지능 기업들의 사업추진 동향

최근 사업 추진 동향	개요
소비자용 시장에서 기업용 시장으로 관심 이동	<ul style="list-style-type: none"> - 스마트폰, AI스피커, 스마트TV, 스마트 냉장고 등 소비자용 시장에 비해 기업용 시장의 비중이 더욱 높아질 전망 - 특히 의료, 금융, 자동차 분야의 시장이 크게 확대될 것으로 예상 (사례) 가천대길병원 포함 7개 거점병원에서 IBM 왓슨 온톨로지 도입, 은행들의 인공지능 챗봇 도입 추진 등
특정 분야에 특화된 인공지능의 확대	<ul style="list-style-type: none"> - 범용 인공지능으로 활용할 수 있는 기능이 제한적임 - 활용 가치가 높은 특정 분야에 전문화된 인공지능 서비스 확대 (사례) 마이크로소프트의 시각장애인용 인공지능 'Seeing AI', 아마존의 패션 코디네이터 인공지능 '에코 룩' 등
보는 인공지능으로의 진화	<ul style="list-style-type: none"> - '듣는 인공지능'에서 '보는 인공지능'으로 진화 - 카메라가 촬영하는 대상을 이해하는 인공지능 기술의 활용 증가 - 시각 인공지능은 컴퓨터비전 기술로 산업 현장에서 제한적으로 사용 되었으나, 최근 범용적인 시각 인공지능 서비스를 출시함 (사례) 구글 렌즈는 구글 어시스턴트와 통합되어 시각 검색의 대중화 예상, 아마존의 '에코 룩'은 카메라를 탑재한 AI 스피커로써 음성 명령으로 전신을 촬영하여 현재 입은 옷이 잘 어울리는지 스타일 수치를 제시, 마이크로소프트의 'Seeing AI'는 화폐와 필기체 글씨 등 인식 가능, 삼성전자의 '빅스비 비전'은 문자, 제품, 랜드마크 등을 인식하고 검색 및 번역 가능

주목해야 할 부분이다.

인공지능 기업의 동향을 살펴보면, 현재 미국과 중국 기업들이 인공지능 기술의 혁신을 선도하고 있다. 미국의 경우, 2016년까지는 주로 연구 개발에 힘을 쏟다가 2017년부터 본격적으로 사업화를 추진 중이다. 주로 활발한 스타트업 인수합병, R&D 투자를 기반으로 한다. 미국 인공지능 기업의 사업화 특징을 요약하면 다음과 같다. 먼저 아마존의 딥렌즈나 구글의 구글렌즈와 같이 음성 인공지능에서 시각 인공지능으로 진화하고 있다. 그리고 소비자용 시장에서 기업용 시장으로 관심이 옮겨가고 있다. 마지막으로 자사 제품이나 서비스에 본격적으로 인공지능을 탑재하고 있다.

중국은 국가적 역량을 인공지능 분야에 결집하여 2030년까지 인공지능 글로벌 리더로 부상한다는 계획을 세우고 있다. 투자 은행인 골드만삭스는 적극적인 투자를 기반으로 한 중국의 인공지능 기술력이 머지않아 미국을 따라잡을 것이라고 전망하였다[15].

현재 중국의 인공지능 연구 개발은 이른바 BAT(바이두(Baidu), 알리바바(Alibaba), 텐센트(Tencent)) 3인방이 이끌고 있는데, 특히 바이두는 가장 앞선 기술 경쟁력으로 검색, 모바일, 자율주행차 등 사업화를 적극적으로 추진 중이다.

최근 국내 인공지능 시장에서도 대기업을 중심으로 일반 소비자를 대상으로 한 음성 인식 플랫폼 기반의 다양한 제품과 서비스가 출시되면서 인공지능 플랫폼 시장 경쟁이 본격화되고 있다. 스마트폰 음성분야에서는 삼성이 인공지능 플랫폼 '빅스비'를 개발하였다. 홈 가상비서 분야에서는 통신 3사(SKT, KT, LG U+)가 인공지능 플랫폼 전담 조직을 꾸려 자체 서비스를 개발하거나 고도화를 추진 중에 있다.

2017년 국내 IT 서비스 업체들이 일반 소비자용 인공지능뿐만 아니라 기업용 인공지능 플랫폼 서비스를 출시하며, 국내 기업용 인공지능 시장 공략을 시작하였다.

삼성SDS는 지난해 6월 인공지능 기술을 활용해 대용량 데이터를 쉽고 빠르게 분석하는 기업

통합 분석 플랫폼 ‘브라이트 AI(Bright AI)’를 공개하였다[16]. 그리고 9월에는 자연어로 대화하며 고객이 요청하는 업무를 지원하고 수행하는 대화형 인공지능 플랫폼 ‘브라이티(Brity)’를 출시했다[17].

LG CNS는 데이터 수집부터 분석, 시각화까지 일련의 빅데이터를 처리하고 분석할 수 있는 인공지능 빅데이터 플랫폼 ‘DAP’을 출시하였다[18]. SK C&C도 지난해 9월 IBM 왓슨의 한국어판 인공지능 플랫폼 서비스 ‘에이브릴(AIBRIL)’을 발표하고, 자연어 이해, 자연어 분류, 검색 및 평가, 문서 변환, 언어 번역, 이미지 인식, 성향 분석 등 제공하는 8종의 API를 공개하였다[19].

이와 같이 최근 국내외 인공지능 관련 기업들의 연구개발 및 사업추진 동향을 정리해보면 다음과 같이 세 가지 특징으로 요약할 수 있다[14].

한편 데이터센터 운영 관점에서도 인공지능 기술의 활용이 더욱 중요해질 것으로 전망되고 있다. 많은 전문가들이 인공지능 기반의 자율 운영 데이터센터가 새로운 트렌드로 자리 잡을 것으로 예측하였다.

자율 운영 데이터센터는 글자 그대로 데이터센터가 스스로 운영할 수 있는 역량을 갖는다는 것이다. 예를 들면 인공지능이 스스로 어떤 패치를 적용해야 하는지, 언제 적용 해야 하는지를 알고 실행에 옮길 수 있다. 또한 장비의 문제 지점을 예측하고 그것을 해결하는 것도 인공지능이 사람을 대체할 날이 멀지 않았다고 전문가들은 예상한다[20]. 하지만 인공지능과 머신러닝을 활용한 데이터센터를 구현하는 데 많은 이슈가 있다.

장애는 기술적인 문제뿐만 아니라 가용 데이터의 양, 품질 그리고 속성과 관련된 것이 많다. 그래서 일부 전문가는 인공지능 기반 애플리케이션을 구동하려면 아주 많은 데이터가 필요하

다고 한다. 데이터를 수집하고 의미를 얻어내는 것이 가장 중요하다는 것이다.

즉, 모든 것이 계측되는 것은 아니고, 모든 것이 로그 작성 기능을 가지고 있지 않으며, 공통적인 표준도 존재하지 않기 때문에 이러한 데이터를 한 곳에 모으고 정리해 머신러닝 알고리즘이 학습할 수 있는 포맷으로 바꾸는 것이 과제라고 말한다.

과정이야 어떻든 먼 미래 혹은 가까운 미래에 인공지능 기반의 자율 운영 데이터센터가 보편화될 것이라는 전망은 공통적이다. 인공지능 기술의 확보를 위해서는 학습을 위한 고품질의 데이터, 인공지능 전문가, 프로세스의 표준화 등 다양한 과제들이 남아 있다. 그러나 보다 효율적인 데이터센터 운영을 위해 적극적인 대응이 필요하다.

4. 클라우드 기술동향

클라우드 컴퓨팅은 스토리지·플랫폼·소프트웨어와 같은 ICT 자원을 데이터센터에 설치 한 후, 개별 이용자가 요구하는 만큼 가상으로 분리하여 정보통신망을 통해 서비스를 제공하고 사용량에 비례하는 비용을 청구하는 방식을 의미한다[21]. 사용자 입장에서는 필요한 ICT 자원을 직접 구입하여 설치하는 전통적인 방식에 비해 클라우드 서비스를 이용하는 것이 비용이나 운영 측면에서 매우 효율적이다.

클라우드 서비스 모델은 SaaS(Software as a Service), PaaS(Platform as a Service), IaaS(Infrastructure as a Service)의 세 가지 유형으로 구분할 수 있는데, 물리적 인프라에서부터 소프트웨어까지 제공하는 영역의 범위에 따라 나눌 수 있다.

IaaS 서비스 제공 업체 가운데 글로벌 선두 업체는 최초의 IaaS 서비스를 시작한 아마존 웹 서비스(AWS)이다. 그 뒤를 마이크로소프트, IBM 등이 뒤따르고 있다. PaaS 서비스의 경우, 세일즈포스(SalesForce)의 Force.com과 사이보우즈(Cybozu)의 킨톤(kintone), 마이크로소프트 애저(Azure), 레드햇 오픈시프트(OpenShift), 피포탈 클라우드 파운드리(Cloud Foundry) 등이 있다. SaaS 시장의 리더는 단연 마이크로소프트이다. 이와 함께 세일즈포스, 어도비, 오라클, SAP 등이 두각을 나타내고 있다.

클라우드 컴퓨팅 기반의 서비스 활용 범위와 규모는 갈수록 확대되는 가운데, 최근 클라우드와 관련하여 새롭게 대두되는 기술적 동향에 주목할 필요가 있다. 언론 매체와 전문가 의견에 따르면, 서버리스 컴퓨팅(Serverless Computing)의 채택 확산, 하이브리드 클라우드(Hybrid Cloud) 활용 확대, 인공지능 채택(시스템 운영 등), 쿠버네티스(Kubernetes) 도입 확대 등으로 요약할 수 있다[22].

서버리스 컴퓨팅은 서버가 필요 없다는 것이 아니라, 개발자가 운영과 관련된 복잡한 문제를 걱정하지 않고 창의적인 코드 개발에만 집중할 수 있도록 한다는 개념이다. 서버리스 컴퓨팅의 이점은 크게 확장성과 비용 효율성이라고 할 수 있다. 확장성 차원에서 보면, 서버리스 환경에서는 사용자 요구를 충족하기 위해 애플리케이션을 자동으로 확장할 수 있다. 이는 사전 프로비저닝이나 오버 프로비저닝 문제를 줄여준다. 비용 효율성 차원에서 보면, 기존의 런타임 모델에는 지속적으로 실행되는 프로세스가 있고, 사용자는 이를 사용하지 않는 경우에도 요금을 내야 한다. 서버리스 환경은 배포된 인스턴스 당 고정된 요금을 지불하지 않고, 인스턴스의 작동 시간에 대해 요금을 내기 때문에 더 경제적이다.

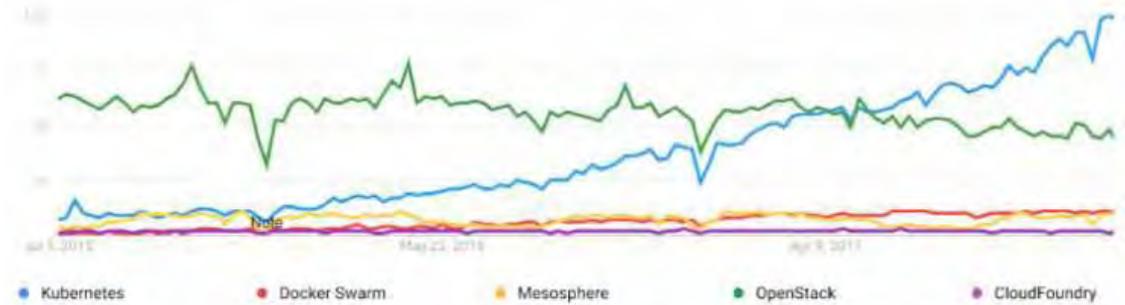
현재 글로벌 기업들은 이미 서버리스 컴퓨팅 관련된 제품을 출시한 바 있다. AWS의 람다(Lambda), 마이크로소프트의 애저 펑션(Azure Function), 구글의 클라우드 펑션(Cloud Function)이 대표적인 사례이다. 국내의 경우 네이버비즈니스플랫폼에서 자사 클라우드 서비스에 서버리스 컴퓨팅 상품 등 신규 상품 2종을 출시했다고 밝혔다[23].

한편 IaaS 퍼블릭 클라우드 업체들은 2017년에 하이브리드 클라우드의 전략을 규정하는 데 노력했다면, 2018년은 본격적인 도입의 해가 되었다고 설명한다.

가장 큰 공을 들인 기업은 마이크로소프트사이다. 마이크로소프트는 2년 이상의 마케팅 끝에 애저 퍼블릭 클라우드를 그대로 옮긴 프라이빗 클라우드 플랫폼인 애저 스택을 출시했다. 시장에서는 이미 애저 스택 초기 배포 사례가 등장하고 있다.

AWS는 실리콘밸리의 가상화 부문 선두 업체와 제휴를 체결하고 VMWare 온 AWS(VMware on AWS) 하이브리드 클라우드 상품을 판매한다. 구글 클라우드 플랫폼은 VMWare, 뉴타닉스(Nutanix) 두 업체와 협력관계를 맺었다. 오라클과 IBM은 단독으로 자체 하이브리드 클라우드 상품을 보유하고 있다[24].

클라우드 환경에서의 인공지능 채택은 보다 효율적이고 지능적인 자원 할당 및 관리에 활용될 수 있다. 예측 분석을 통해 IT 인프라 및 응용 프로그램 소유자는 실행 가능한 정보와 권장 사항을 받게 된다. 그에 대한 응답을 자동화하는 기능을 추가한다면 소유자 또는 사용자의 부담은 더 줄어들 것이다. 인공지능 기반의 분석 시스템은 인프라 및 응용 프로그램에 대한 통찰력을 갖고, 비정상적인 성능이나 보안 이슈가 발생했을 때 이에 대한 지능화된 대응을 할 수 있다.



(그림 2) 컨테이너 관리환경의 트렌트 그래프

즉 IT 인프라가 관리자의 명령 없이 스스로 문제에 대한 대응을 할 수 있다는 의미이다[22].

컨테이너 관리 플랫폼 분야에서는 쿠버네티스의 강점이 부각되고 있다. 쿠버네티스는 이제 기업 내 컨테이너 관리 경쟁(Docker Swarm, Kubernetes 및 Mesos 간 3자 경쟁)에서 완전히 승리한 것으로 보인다. AWS가 쿠버네티스 오픈소스 기술을 지원하기로 2017년 11월 리인벤트(re:Invent) 회의에서 발표했기 때문이다. 이제 퍼블릭 클라우드 업체(마이크로소프트, 구글, IBM, AWS 등)들은 모두 쿠버네티스를 완전히 지원한다[25]

쿠버네티스라는 컨테이너 관리 환경이 주목을 받게 된 배경은 다음과 같다. 컨테이너 기반의 환경은 기본적으로 배포에 장점이 있고 마이크로 서비스 아키텍처 구조에 적합 하다. 특히 최근 마이크로 서비스 아키텍처가 점점 더 발전하기 시작하면서 디자인 패턴과 이를 구현하기 위한 다양한 인프라 플랫폼이 소개되기 시작하였다. 결과적으로 서비스가 점점 작아지면서 1~2 코어로도 운영할 수 있는 작은 서비스들이 다수 등장하게 되었다. 이런 작은 서비스들은 VM 환경으로 운영하기에는 낭비가 너무 심하다(VM 이미지 크기가 너무 크며, 다양한 이미지를 VM으로 관리 배포하기에는 배포 속도 등 다양한 문

제 발생). 이러한 배경에서 점차 컨테이너 기반의 환경이 실용적이라는 판단 아래 주목을 받게 되었다.

컨테이너 기반 환경을 지원하는 다양한 오픈소스들이 전까지는 표준 없이 혼용되었다가 2017년 말을 기점으로 쿠버네티스가 사실상 표준으로 되어가는 형국이다. 쿠버네티스는 고(GO) 언어로 구현되었으며, 이는 벤더나 플랫폼에 종속되지 않아 대부분의 퍼블릭 클라우드 또는 오픈스택과 같은 프라이빗 클라우드, 베어메탈 방식의 클라우드 구축에도 배포가 가능하다는 장점이 있다.

5. 결 론

4차 산업혁명의 시대를 이끌어가는 빅데이터, 인공지능, 클라우드 등 최신 ICT 기술들의 활용 범위와 영향력이 점차 확대되어가고 있는 가운데, 각 기술들이 향후 어떠한 형태로 발전하고 서로 융합되어 성공적인 응용 사례들을 만들어 낼 것인지에 대해 면밀히 검토하고 준비해야 한다. 빅데이터는 데이터 경제의 시대에서 이제 모든 기술 분야의 근간이자 성공을 위한 동력으로 자리하고 있다. 빅데이터의 상징이라고 볼 수도 있는 하둡 생태계의 발전 동향과 함께 클라우드

기술과의 융합 동향에 대해 인지하고, 보다 효율적인 빅데이터 활용을 위한 연구에 힘을 쏟아야 할 것이다. 뿐만 아니라 복잡 다양한 빅데이터에 대해 효율적인 분석을 위해 사용되는 인공지능 기술의 발전 동향도 함께 연구해야 할 부분이다.

인공지능 기술은 최근 딥러닝 기술의 발전과 활용사례 확산에 따라 향후 점차 많은 산업분야에서 핵심 기술로 활용될 것으로 보인다. 보다 효율적인 딥러닝 활용을 위해 다양한 프레임워크에 대해 파악하고 적용 분야와 알고리즘적 특성을 함께 파악할 필요가 있다. 중요한 것은 인공지능 기술의 성공적 활용을 위해서는 학습을 위한 데이터의 확보가 반드시 필요하다는 것이다. 즉 빅데이터와 인공지능 기술은 상호 필수 불가결한 존재이며 향후 함께 발전하고 적용될 때 진정한 의미의 가치를 불러일으킬 수 있다는 것을 인지해야 할 것이다.

클라우드 기술 역시 빅데이터나 인공지능 기술과 융합되어 더욱 큰 가치를 만들어내고 있다. 빅데이터 활용을 위한 인프라 등 자원의 효율적 활용을 위해 점차 클라우드 기술의 활용이 확산되고 있으며, 인공지능 기반의 응용 서비스 개발도 클라우드 환경을 중심으로 점차 확대되어가고 있다. 반대로 클라우드 환경은 인공지능 기술의 도입으로 점차 지능적으로 관리될 수도 있다.

앞서 살펴본 것처럼 빅데이터, 인공지능, 클라우드 등 최신 ICT 기술들은 서로 독립적인 발전 양상을 보이는 것이 아니라 더욱 긴밀하게 융합되고 다양한 응용 사례들을 만들어내고 있다. 뿐만 아니라 블록체인, IoT 등 기술과도 여러 분야에서 융합되어 우리 삶의 모습을 혁신적으로 변화시키는데 작용하고 있다. 디지털 혁신의 시대에 살아가고 있는 우리 모두와 국가는 이들 기술들의 변화 양상을 정확하게 인지하고 다양한 산업분야에서 어떻게 적용되어져 갈 것인지에 대

해 파악하여 선제적인 대응을 해 나가야 할 것이며, 국가 경쟁력 확보 차원에서 새로운 모습의 혁신적 생태계 구축에도 많은 고민을 해야 할 것이다.

참 고 문 헌

- [1] Apache Hadoop, hadoop.apache.org
- [2] 한국정보화진흥원, 빅데이터 동향과 이슈, Vol.33, 2017.
- [3] Forrester, Move Big Data To The Public Cloud With An Insight PaaS
- [4] Towards Data Science, 10 hot trends of big data analytics for 2017
- [5] 투이컨설팅, 국내외 빅데이터 분석 프로젝트 추진 동향과 향후 발전방향
- [6] Forbes, 10 Predictions For AI, Big Data, and Analytics in 2018
- [7] CIO.com, 10 hot data analytics trends - and 5 going cold
- [8] NG Data, What is Big Data Analytics?
- [9] SimpliLearn, Insight-as-a-Service: The Next Big Thing in Analytics
- [10] InformationWeek, 8 Reasons To Consider Insight-As-A-Service
- [11] IDC, Worldwide semiannual Cognitive Artificial Intelligence Systems Spending by Industry Market 2016-2020 Forecast
- [12] Tractica, Artificial Intelligence Market Forecast
- [13] Market and Markets, Artificial Intelligence Market-Global Forecast to 2022
- [14] 정보통신기술진흥센터, AI First, AI Everywhere 로 전개되는 인공지능
- [15] Goldman Sachs, China's rise in Artificial Intelligence: The New New China
- [16] 삼성SDS, 삼성SDS, AI 기반의 분석 플랫폼 Brightics AI 공개, 2017.6.
- [17] 삼성SDS, 삼성SDS, 기업용 대화형 AI(Brity) 출시, 2017.9.

- [18] LG CNS, AI 빅데이터 플랫폼 DAP 출시, 혁신의 답을 찾다, 2017.9.
- [19] SK C&C, 인공지능 대중화... 데이터 산업화 큰 길 열렸다, 2017.9.
- [20] IDG, "자율 운영 데이터센터가 온다" AI와 머신러닝 기반의 데이터센터 운영 효율화
- [21] 국회입법조사처, 클라우드 컴퓨팅의 현황과 과제
- [22] steemit.com, 2018년 5대 Cloud Computing 예측
- [23] CIO Korea, '서버리스 컴퓨팅'이 뜬다... 개념에서 사용례까지, 2017.
- [24] IT World, 2018년 하이브리드 클라우드 시장에 대한 5가지 예측
- [25] CIO Korea, 서버리스의 부상, 쿠버네티스의 확산, AWS의 약화 외... 2018년 클라우드 트렌드 진단



김 영 배

이메일 : chensan@penta.co.kr

- 1994년 충남대학교 인문학부 (학사)
- 1995년~현재 펜타시스템테크놀러지 빅데이터팀 기술 총괄이사
- 관심분야: 빅데이터, 인공지능, 클라우드, 블록체인, IoT 등

저 자 약 력



김 대 엽

이메일 : dykim@penta.co.kr

- 2005년 충남대학교 정보통신공학부 (학사)
- 2015년 충남대학교 컴퓨터공학과 (박사(석박사 통합과정))
- 2016년~현재 펜타시스템테크놀러지 빅데이터팀 책임 컨설턴트
- 관심분야: 소프트웨어공학, 빅데이터, 인공지능, 클라우드, 블록체인, IoT 등