



2021. 08

유망시장 Issue Report

빅데이터

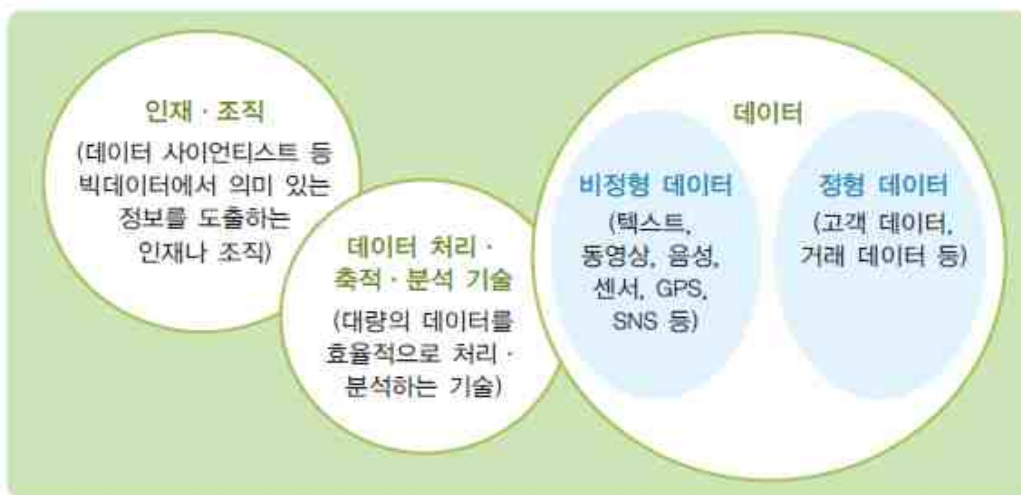
목 차

I. 개요	01
II. 정책동향	07
III. 기술동향	14
IV. 시장동향	19
V. 산업동향	26

I. 개요

1. 아이템 개요

- 빅데이터란 형식과 규모가 다양하고 데이터 축적 속도가 매우 빨라 기존의 데이터베이스로 처리할 수 있는 역량을 넘어서는 초대용량(테라바이트 또는 페타바이트급)의 정형, 비정형 데이터를 모두 포함하는 것을 말함
- 빅데이터는 다양한 종류의 대규모 데이터로부터 효율적으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처
- 거대한 데이터 집합 자체만을 의미했던 이전과 다르게 도구, 플랫폼, 분석 기법 등을 포함
- 빅데이터 분석 솔루션은 대용량의 정형 혹은 비정형 데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내어 비즈니스 의사결정을 지원할 수 있는 인사이트를 발굴·예측하는 소프트웨어 또는 하드웨어를 의미



[그림] 빅데이터 정의

※ 출처 : 빅데이터 컴퓨팅 기술, 한빛아카데미, 2016

- 빅데이터의 정의는 문헌마다 차이가 있으며, 바라보는 관점(데이터 규모와 기술, 업무 수행, 데이터 활용 등)에 따라서도 다름
- 빅데이터의 전통적인 개념은 구글, 마이크로소프트와 같은 대기업이나 NASA의 연구 프로젝트에서 분석하는 방대한 양의 데이터를 말함

[표] 빅데이터의 정의

기관	정의
Mckinsey (2011)	일반적인 데이터베이스 SW가 수집·저장·관리·분석할 수 있는 처리 능력을 초과하는 규모의 데이터 집합
IDC (2011)	다양한 형태의 방대한 데이터로부터 고속 캡처, 데이터 탐색 및 분석을 통해 경제적으로 필요한 가치를 추출하기 위해 설계된 차세대 기술 및 아키텍처
Gartner (2012)	더 나은 의사결정, 시사점 발견 및 프로세스 최적화를 위해 사용되는 새로운 형태의 정보처리가 필요한 대용량, 초고속 및 다양성의 특성을 가진 정보자산
교육과학기술부 외 (2012)	데이터 형식이 다양하고 생성 속도가 매우 빨라, 새로운 관리 및 분석 방법이 필요한 대용량 데이터
정보통신기술협회 (2015)	새로운 가치 추출을 위해 기존의 기술 또는 기법으로 처리하기 어려운 특장(규모, 신속성, 가변성, 다양성, 진정성)을 갖는 데이터 모음

※ 출처 : 서울시 공공빅데이터 활성화방안, 서울연구원, 2018

○ (빅데이터 등장 배경) 1990년 이후 인터넷이 확산되면서 정형화 및 비정형화된 형태의 데이터가 무수히 발생하여 정보 홍수 개념이 등장하였고, 이것이 오늘날 빅데이터 개념으로 이어짐

- 개인화 서비스와 SNS의 확산으로 기본 인터넷 서비스 환경이 재구성되어 검색과 포털 위주였던 인터넷 서비스가 통신, 게임, 음악 등의 영역에서 개인화 서비스와 소셜 네트워크 서비스를 제공하는 환경으로 변화함
- 데이터양이 많아짐에 따라 기존의 데이터 저장, 관리, 분석 기법으로 데이터를 처리하는데 한계가 발생하면서 정보기술의 패러다임이 변화하게 되었고 이를 통해 빅데이터 용어가 등장하게 됨

[표] 정보기술의 패러다임 변화

	PC 시대	인터넷 시대	모바일 시대	스마트 시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물 정보화
정보기술 이슈	PC, PC통신, 데이터베이스	초고속 인터넷, www, 웹 서버	모바일 인터넷, 스마트폰	빅데이터, 차세대 PC, 사물 네트워크
핵심 분야 (서비스)	PC, OS	포털, 검색 엔진, Web 2.0	스마트폰, SNS, 웹 서비스	미래 전망, 상황 인식, 개인화 서비스
대표기업	MS, IBM	구글, 네이버, 유튜브	애플, 페이스북, 트위터	구글, 삼성, 애플, 페이스북, 트위터
정보기술 비전	1인 1PC	클릭 e-Korea	손 안의 PC, 소통	IT everywhere, 신가치창출

※ 출처 : 빅데이터 컴퓨팅 기술, 한빛아카데미, 2016

○ **(빅데이터의 분류)** 빅데이터는 데이터의 정형화 정도에 따라 정형, 반정형, 비정형으로 분류

- **(정형 데이터)** 일정한 규칙에 따라 체계적으로 정리한 데이터로 표와 같이 고정된 열에 저장되는 데이터와 파일을 의미
 - 관계형 데이터베이스, 스프레드시트가 있음
- **(반정형 데이터)** 고정된 필드에 저장되어 있지는 않지만, 메타데이터나 스키마 등을 포함하는 데이터를 의미
 - XML, HTML 텍스트와 페이스북, 카카오톡 등 소셜 네트워크 서비스 사용자가 생성하는 데이터들이 있음
- **(비정형 데이터)** 고정된 필드에 저장되어 있지 않은 데이터로 텍스트, 이미지, 동영상과 같이 사전에 정의된 구조가 없는 형태의 데이터를 의미
 - 비정형 데이터의 증가 속도는 누구도 예측할 수 없을 정도로 빠름

○ **(빅데이터의 속성)** 가트너의 애널리스트 더그 레이니는 현재 가장 널리 사용하는 빅데이터의 속성을 규모, 다양성, 속도인 3V로 정의했고 IBM은 여기에 정확성 요소를 더해 4V로, 최근에는 가치를 포함하여 5V로 정의함

- **(규모)** 미디어나 위치 정보, 동영상 등과 같이 다루어야 할 데이터의 크기로, 기술의 발달에 따라 테라바이트를 넘어 요타바이트까지를 빅데이터로 통칭
- **(다양성)** 다양한 종류의 데이터를 수용하는 속성을 말하며, 형식이 정해져 있는 정형 데이터뿐만 아니라 동영상, 사진, 메시지 등 다양한 비정형 데이터도 생성
- **(속도)** 대용량의 데이터를 빠르게 처리하고 분석할 수 있는 속성으로 데이터를 자동으로 생성하는 센서, 스마트폰 등 데이터 생성 및 유통 채널의 다변화로 데이터 생성 속도가 빨라지면서 처리 속도의 가속화를 요구함
- **(정확성)** 데이터에 부여할 수 있는 신뢰 수준을 말하며, 높은 데이터 품질을 유지하는 것은 중요한 요구사항이자 어려운 과제
- **(가치)** 빅데이터를 저장하려고 IT 인프라 구조 시스템을 구현하는 비용을 말하며, 빅데이터의 규모는 엄청나며 대부분은 비정형적인 텍스트와 이미지 등을 구성



[그림] 빅데이터의 속성

※ 출처 : 빅데이터 컴퓨팅 기술, 한빛아카데미, 2016

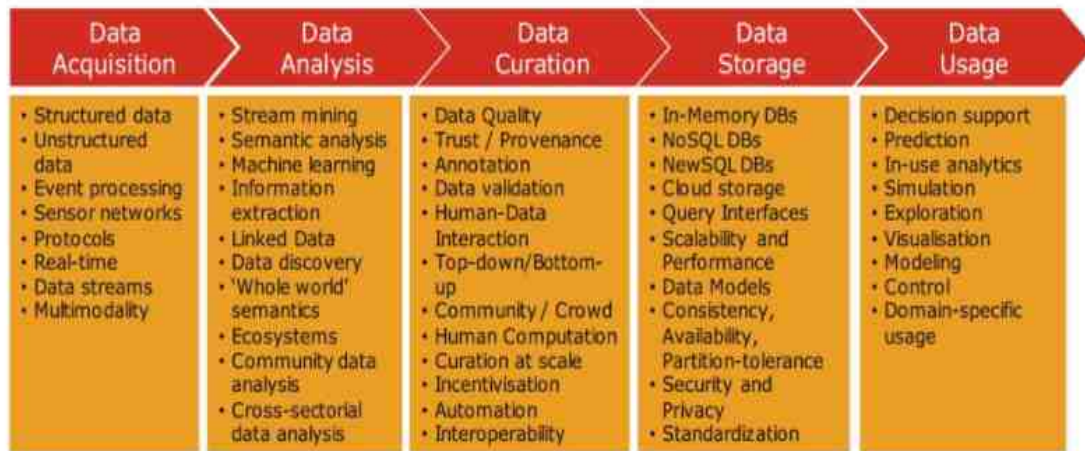
2. Value Chain

○ 빅데이터 산업은 데이터 생산, 수집, 처리, 분석, 유통, 활용 등을 통해 가치를 창출하는 상품과 서비스를 생산·제공하는 산업

- 후방산업에는 데이터 생산 및 수집에 제반이 되는 ICT 인프라 기술, IoT 분야 등이 존재하며, 전방산업에는 생성된 데이터를 다루는 데이터솔루션, 데이터 구축, 빅데이터 분석 컨설팅 및 서비스와 기술 발전의 중심인 인공지능 시스템이 존재

○ 빅데이터 밸류체인은 다음과 같이 5가지 과정으로 구성

- **(데이터 수집)** 데이터 웨어하우스 또는 데이터 분석을 수행할 수 있는 다른 스토리지 솔루션에 배치하기 전에 데이터를 수집 및 정리하는 프로세스
- **(데이터 분석)** 획득한 원시 데이터를 도메인 별 용도뿐만 아니라 의사결정에 사용할 수 있도록 만드는 작업을 말하며, 관련 데이터를 강조 표시하고 비즈니스 관점에서 잠재력이 높은 유용한 정보를 합성 및 추출하는 데이터 탐색, 변환 및 모델링 과정이 포함
- **(데이터 큐레이션)** 효과적인 사용에 필요한 데이터 품질 요구사항을 충족하도록 수명주기 동안 데이터를 적극적으로 관리하는 단계로써, 콘텐츠 생성, 선택, 분류, 변환, 검증 및 보존과 같은 활동으로 분류 가능
- **(데이터 저장)** 데이터에 대한 빠른 액세스가 필요한 애플리케이션의 요구를 충족하는 확장 가능한 방식으로 데이터를 지속하고 관리
- **(데이터 사용)** 데이터에 대한 액세스가 필요한 데이터 기반 비즈니스 활동, 분석 및 비즈니스 활동 내에서 데이터 분석을 통합하는데 필요한 도구를 포함하며 비즈니스 의사결정에서 데이터 사용은 비용 절감, 부가가치 증가 등 경쟁력을 향상시킴



[그림] The Big Data Value Chain

※ 출처 : Definitions, Concepts, and Theoretical Approaches, Edward Curry, 2015

II. 정책동향

1. 해외 정책동향

○ (미국) 2012년 과학기술정책국(Office of Science and Technology)에서 빅데이터 기술 개발 및 활용, 빅데이터 전문인력 양성 등을 주요 목적으로 ‘빅데이터 R&D 이니셔티브’를 발표

- 2016년 NITRD(Networking and Information Technology Research and Development)에서 ‘빅데이터 R&D 전략 계획’을 발표

• 범부처 차원에서 빅데이터 7대 R&D 전략과 18개 세부과제를 제시

- 대통령 직속 기구인 국가과학기술위원회(National Science and Technology Council, NSTC)에서 연방정부 부처 기관의 ICT 연구개발 프레임워크인 NITRD 프로그램 운영

• 2018년 빅데이터, 로봇, 인공지능, 고성능 컴퓨팅 시스템 인프라 운용으로 영역을 확대하였으며, 2019년 사이버보안을 비롯한 개인 정보 등으로 확대 변경하여 시행 중

• 빅데이터와 인공지능 분야를 강화하기 위해 투자 확대 및 프로그램을 신설하였고 실제로 예산은 인프라 및 소프트웨어보다 증가

- 2019년 1월, 미국 연방정부는 정책 평가에 필요한 증거를 통계로 규정하는 증거 기반정책법(Foundations for Evidence-Based Policymaking Act of 2018) 제정

• 증거기반정책법은 연방정부의 증거 구축 활동을 규정하고 연방정부가 생산하는 공공데이터의 활용에 초점을 맞춘 OPEN(Open, Public, Electronic, and Necessary) 정부 데이터 법과 비밀정보보호와 통계적 효율성으로 구성하여 데이터와 통계의 유기적인 연계를 강조

[표] 미국 빅데이터 R&D의 7대 전략과 18개 세부과제

	전략	세부과제
1	미래 빅데이터 특성을 반영한 기술 개발로 차세대 능력 함양	• 데이터의 크기, 전달/처리 속도, 복잡성에 보조를 맞춘 기술 개발 • 미래에 요구되는 새로운 빅데이터 기술의 방법론 개발
2	데이터의 신뢰성 및 더 나은 빅데이터 기반 의사결정을 위한 R&D 지원	• 데이터의 신뢰성과 타당성을 제고시켜 더 나은 결과도출 • 데이터 기반 의사결정을 지원하는 도구 개발
3	빅데이터 혁신을 가능하게 하는 사이버 인프라 구축 및 강화	• 국가 데이터 인프라 강화 • 빅데이터에 대한 응용과학 사이버 인프라 역량 강화 • 유연하고 다양한 인프라 자원
4	데이터 공유 및 관리를 촉진하는 정책을 통한 데이터 가치 향상	• 데이터 투명성, 효율성을 증가시키는 메타데이터의 모범 사례 개발 • 데이터 자산에 효율적이고 지속적이며 안전한 접근을 제공
5	개인 정보 보호, 보안 및 빅데이터의 수집·공유·활용의 윤리적 측면 이해	• 올바른 개인 정보 보호 • 안전한 빅데이터 사이버공간 구축 • 데이터 거버넌스를 위한 정보윤리 이해
6	국가의 빅데이터 교육 및 훈련 환경 개선, 폭넓은 인력 확충	• 데이터 과학자의 양성 • 데이터 영역 전문가 커뮤니티 확장 • 데이터 사용이 가능한 인력 확충 • 공공의 데이터 활용 역량 개선
7	정부 기관, 대학, 기업, 비영리단체와의 협력에 의한 빅데이터 혁신 생태계 지원	• 기관 간 빅데이터 협력 장려 • 빠른 대응과 영향력 측정이 가능한 정책 및 정책추진 프레임워크 구축

※ 출처 : 데이터산업 백서, 한국데이터진흥원, 2017

○ (유럽) 유럽연합은 2015년부터 빅데이터에 대해 향후 5년간 산업 컨소시엄의 20억 유로와 공적 자금 5억 유로를 합친 25억 유로(약 3조 3,870원)를 투자

- 아토스(Atos), IBM, 노키아, 솔루션엔테크웍스(Nokia Solution and Networks, Microsoft(N)), 오렌지(Orange Telecom), SAP, Siemens 등의 기업들과 연구 기관들 주도로 민관 협력을 추진
- 2017년 1월, 유럽 내 47개 주요 교통, 물류, IT 관련 기업 및 기관으로 구성된 컨소시엄이 30개월간 1,870억 유로의 예산으로 빅데이터 기반의 Transforming Transport 프로젝트를 추진
- 2017년 유럽 내 통합 디지털 플랫폼을 기반으로 데이터 접근·분석·활용 강화를 통한 새로운 데이터 비즈니스 창출을 목표로 데이터 경제 육성전략을 제시
- 데이터 접근권 강화, 기술 표준 제정, 법적 책임 명확화와 개인 정보 보호 규정

(GDPR) 제정으로 데이터 삭제권, 정보 이동권, 프로파일링에 대한 권리 등 개인 정보 보호 강화와 합법적 데이터 유통을 동시에 추구

○ (일본) 2017년 일본 정부는 ‘미래투자전략’을 수립하여 빅데이터를 ICT 융합 활성화 촉매로 활용하기 위해 이노베이션 벤처를 탄생시키는 선순환 시스템을 구축하여 빅데이터 산업에 집중 투자 지원 추진

- 2020년까지 ‘관민 데이터 활용추진 전략회의’를 중심으로 공공데이터 개방 집중 실시, 데이터 활용 권한을 명확히 하는 기업 간 계약 가이드라인 전개, 데이터 유통을 위한 법 제정비, 정보은행·데이터 거래시장의 제도 정비 등을 실시
- 2018년 정부 주도로 정부, 지자체, 민간 보유의 모든 데이터를 정리하여 개방할 계획을 발표
 - 2014년부터 정부의 투명성, 경제 활성화, 공공서비스 향상이 목적인 ‘빅데이터 활용을 통한 Open Data 정책’의 연장선으로 공공기관의 데이터 공개를 추진
 - 국가가 보유한 각종 통계 데이터, 지도데이터, 기상 데이터와 민간에서 보유한 도로 상황 및 교통량과 관련된 데이터, 작물의 재배 상황 정보, 위치 정보 등 광범위한 분야의 정보를 취합하는 것이 목표
 - 인공지능 분석이 용이하도록 행정용어부터 산업용어까지 빅데이터 내 용어를 통일·정리하여 취합
 - G20 정상회의에서 데이터베이스 구축에 미국과 유럽의 참가를 요청, 10억 명 이상의 정보가 집적된 초국가적인 데이터베이스 Pool 구축을 최종 목표로 함

○ (중국) 2015년 8월, 리커창 총리 주재로 열린 국무원 상무 회의에서 빅데이터 개발 및 응용을 통한 대중창업, 만민혁신 추진 및 경제성장을 위한 신동력 개발, 정부의 관리 감독 서비스 개선을 추진

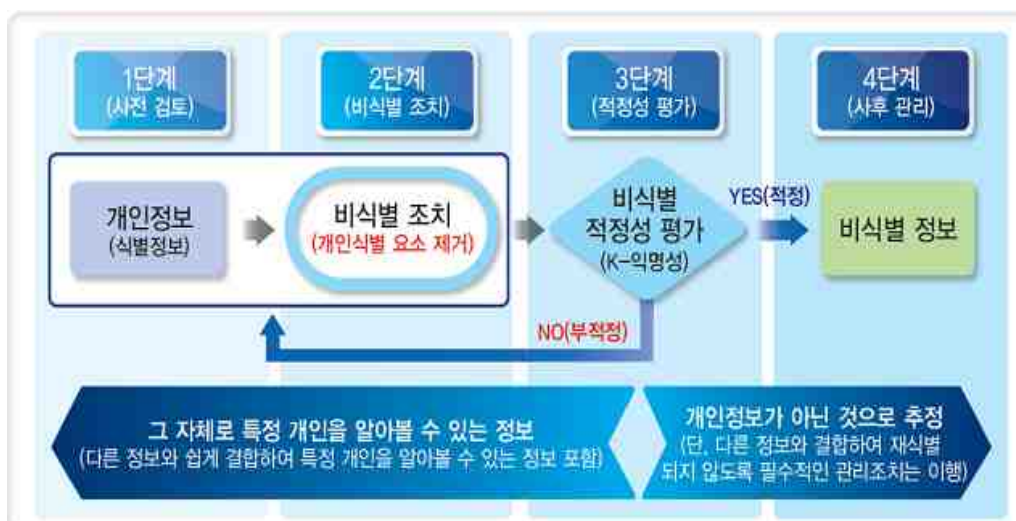
- 2016년 베이징시 빅데이터 및 클라우드 컴퓨팅 발전 행동계획 발표
 - 향후 5년간 정부 데이터의 민간 개방 확대, 핵심기술 개발 및 응용 확대 등을 통한 빅데이터 산업 육성 계획과 경제·사업의 광범위한 영역에서의 빅데이터 응용 계획을 제시
- 2017년 빅데이터 산업 발전계획에서 빅데이터를 핵심 산업으로 추진하기

위해 10개 이상 글로벌 빅데이터 선도기업 육성 정책을 제시

- 데이터 개방 확대, 플랫폼·오픈소스 기술 지원, 빅데이터 전문 SW 수준 향상, 전문 인재의 공급, 데이터거래소 등 생태계 조성
- 정확한 고객 수요 파악과 분석을 통해 R&D의 효율성을 높이는 한편, 조립·생산 라인의 경쟁력을 높임으로써 생산예측과 관리를 강화하고 경영 효율화를 재고할 수 있을 것으로 기대
- 통계국은 국가통계국과 텐센트 등 6개 기업 공동으로 연구를 진행하여 정부 통계 데이터를 이용한 빅데이터 기반 공동 수집·처리·분석·탐색·발표 기술 개발 표준과 통계 표준을 제정

2. 국내 정책동향

- 2016년 빅데이터 분야 개인정보 규제혁신 방안 마련 및 빅데이터 분석 활용을 위한 ‘개인정보 비식별 조치 가이드라인’ 발표
 - 관계부처가 합동으로 현행 개인 정보 보호 법령의 틀 내에서 빅데이터가 안전하게 활용될 수 있도록 하는데 필요한 개인정보의 비식별 조치 기준과 비식별 정보의 활용 범위 등을 명확히 제시하여 기업투자와 산업 발전을 도모함과 동시에 국민의 개인정보 인권 보호를 목적으로 함
 - 정보 주체를 알아볼 수 없도록 비식별 조치를 적정하게 한 비식별 정보는 개인정보가 아닌 것으로 추정하여 빅데이터 분석에 폭넓게 활용하도록 허용



[그림] 개인정보 비식별 조치 가이드라인

※ 출처 : 개인정보 비식별 조치 가이드라인, 국무조정실, 2016

- 2017년 11월 발표된 ‘4차 산업혁명 대응계획’은 지능화 기반 산업 및 사회·공공 분야에 빅데이터 기술을 활용하여 지능화 혁신을 이루고자 함
 - 의료기관 간 진료 정보 전자 교류 전국 확대 및 보건의료 데이터 구축·개방을 추진하여 데이터 인프라 확보, 신약 개발 주기 및 비용 절감을 위한 빅데이터 활용 등을 추진
 - 제조업, 에너지, 농수산업, 스마트 시티 등의 분야에 빅데이터 기술을 활용하거나, 빅데이터를 구축하여 지능화 혁신을 추진

○ 2017년 12월, 정부는 빅데이터를 혁신성장동력으로 지정하고 2018년 5월, 2022년까지 해당 분야 조기 상용화를 위해 제도개선, 연구개발, 인프라 구축 등의 지원 방안을 발표함

- 개인정보의 안전한 활용을 위해 개인정보 범위 명확화, 비식별 조치 법제화 등의 제도개선 논의를 추진하여 합의안을 도출하고, 정보 주체 중심의 데이터 활용 촉진을 지원
- 인공지능 및 초연결 사회를 위한 빅데이터 연구개발을 지원하고 빅데이터 분석플랫폼과 유통 플랫폼의 구축 및 고도화 지원

○ 2018년 대통령 직속 4차산업혁명위원회에서는 데이터 산업 활성화 전략을 발표하여 개인 정보 보호와 활용의 균형 및 데이터 경제 활성화를 추진

- 정보 주체인 개인이 데이터 통제·활용권을 가지도록 하고 개인정보의 안전한 활용을 위한 법제화·안심 존 구축·신기술 적용 등을 지원
- 데이터 구축·개방, 저장·유통, 분석·활용의 지원체계를 마련하고, 핵심 기술 확보 및 전문인력 확보 등을 통해 기반 조성을 추진
- 국회 4차산업혁명특별위원회는 개인정보 보호법과 정보통신망법, 신용정보법 등 중복 조항 정비 및 강력한 사후규제를 도입하되 가공정보에 대한 안전한 활용이 가능하도록 입법권고안을 발표

○ 2018년 6월, 데이터 산업 활성화 전략으로 I-Korea 4.0 데이터 분야 계획인 I-Data를 발표

- 4차 산업혁명 시대 데이터가 모든 산업의 발전과 새로운 가치 창출의 촉매 역할을 하는 ‘데이터 경제’로의 패러다임 전환의 입각하여 데이터 이용제도 패러다임 전환, 데이터 가치 사슬 전방위 확산, 글로벌 데이터 산업 육성기반 조성 등을 추진과제로 발표
- 공공데이터의 원칙적 개방(국가안보·개인 정보 제외) 등을 통해 공공부문 핵심 데이터 구축 개방 전면화
- 데이터의 효율적인 저장관리를 위한 클라우드 본격 확산 및 데이터 유통 촉진을 위해 민간 공공을 연계한 개방형 데이터 기반구축, 2022년까지 중소 벤처기업 1만 개 이상에 클라우드 도입을 지원
- 빅데이터 전문기업 성장 지원으로 데이터 강소기업 100개를 육성하고

빅데이터 개발에 필수적인 컴퓨팅 자원, 데이터셋 등이 집적된 판교 글로벌 ICT 혁신 클러스터 지속 고도화 및 활용을 지원

○ 2021년 4차산업혁명위원회는 ‘민관 협력 기반 데이터 플랫폼 발전전략’을 발표

- 데이터를 체계적으로 수집·개방하는 공공 민간의 다양한 데이터 플랫폼들의 시장 안착과 지속가능한 성장을 지원하는 데이터 플랫폼 활성화 방안을 심의·의결

- (전략1) 분야별 특성에 따른 대표 데이터 플랫폼 확충, 민·관 데이터 플랫폼 발전 협의체를 구성
- (전략2) 수요자가 데이터 구매 또는 가공을 지원받을 수 있는 데이터 바우처 신설, 데이터 수요를 미리 알려주는 수요 예보제 도입
- (전략3) 분석 서비스 바우처 지원, 혁신 서비스 개발 등을 위한 플랫폼 데이터 할인 및 무상으로 제공, 데이터 플랫폼의 개인정보 관리체계 점검·개선
- (전략4) 데이터 입력 항목의 용어를 표준화, 비정형 데이터 품질관리 방법을 개발하는 등 품질관리 가이드를 지속적으로 고도화



[그림] 민관협력기반 데이터 플랫폼 발전전략 비전 및 추진전략

※ 출처 : 민관협력기반 데이터 플랫폼 발전전략 발표 보도자료, 과학기술정보통신부, 2021

Ⅲ. 기술동향

1. 기술범의 및 특징

- 빅데이터는 데이터 수집·저장·처리 등 빅데이터 플랫폼 기술과 빅데이터 분석·예측 기술을 활용하여 새로운 비즈니스 가치를 창출하는 빅데이터 분석·활용기술을 포함

[표] 빅데이터 기술 분류

	소분류	세분류	요소기술
빅데이터	빅데이터 처리·유통	빅데이터 수집·유통기술	빅데이터 수집/정제/융합/가공, 데이터 품질관리, 실시간 ETL/ELT, 데이터 생성/증강, 비식별화 및 필터링, 데이터 마켓, 데이터 레이크, 데이터 프리퍼 레이션, 마스터 데이터 관리
		빅데이터 저장·처리·관리 기술	실시간 스트림 처리, 데이터 처리 및 관리, 데이터 라이프 사이클, 가상 데이터맵 관리, 데이터 처리 프레임 워크
	빅데이터 분석·활용	빅데이터 분석·예측 기술	심층 분석, 실시간 분석, 그래프 분석, 예지분석, 예측분석, 지시적 분석, 인지 분석, 시공간 복합분석, 분석 모델링, 시뮬레이션, 데이터 기계학습/딥러닝, 이상치 검출
		빅데이터 활용·시각화 기술	빅데이터 서비스, 빅데이터 응용, 사회변화 예측, 데이터 시각화, 분석 시각화, 대쉬 보드

※ 출처 : ICT R&D 기술로드맵 2023, 정보통신기획평가원, 2018

- 빅데이터는 데이터의 생성·저장·처리·분석·표현의 과정으로 나뉘며, 기존의 데이터와 속성이 달라 각 처리 과정별로 새로운 방법들이 요구되면서 다양한 기술들이 등장함

- **(소스 생성과 수집기술)** 데이터는 소스 위치에 따라 내부 데이터와 외부 데이터로 구분하며, 데이터 수집은 주로 톨, 프로그램으로 자동 진행됨
 - 로그 수집기, 크롤링, 센싱, RSS 리더/오픈 API, ETL 등 수집 방법을 사용
- **(저장 기술)** 데이터 저장 관리는 추후 사용할 수 있도록 데이터를 안전하고 효율적으로 저장하는 것으로, ‘대용량, 비정형, 실시간성’ 속성을 수용할 수 있는 저장 방식이 필요함
 - 대량의 데이터를 파일 형태로 저장할 수 있는 기술과 비정형 데이터를

정형화된 데이터 형태로 저장하는 기술이 중요

- 분산 파일 시스템, NoSQL, 병렬 DBMS, 네트워크 구성 저장 시스템 등이 대표적인 기술임
- **(처리 기술)** 방대한 양의 데이터와 데이터 생성 속도, 데이터 종류의 다양성을 모두 고려할 수 있는 기술이 필요함
 - 대표적인 빅데이터 처리 기술로 분산 병렬 데이터 처리 기술의 표준인 맵리듀스가 있으며, 정형·비정형 빅데이터 분석에 가장 선호되는 솔루션인 하둡, 비 관계형 데이터베이스인 NoSQL 등이 있음
- **(분석 기술)** 빅데이터 분석에 사용되는 기술은 대부분 통계학과 전산학, 기계 학습과 데이터 마이닝 분야에서 사용되는 기술들로 이 기술들의 알고리즘을 대규모 처리에 맞게 개선하여 빅데이터 처리에 적용
- **(표현 기술)** 어렵고 복잡한 정보를 한눈에 쉽게 이해할 수 있도록 간단한 도표나 3D 이미지 등으로 표현하는 정보 표현 기술이 발전
 - 최근 빅데이터 표현 기술 중 2009년 구글에서 개발한 Fusion Tables은 방대한 양의 데이터를 표현해 주는 온라인 서비스



[그림] 빅데이터 처리과정

※ 출처 : 빅데이터 컴퓨팅 기술, 한빛아카데미, 2016

2. 국내/외 Trend

- 빅데이터 기술은 전 산업에 영향을 미치는 기반 기술로 자리매김하고 있으며, 지능형 분석을 통한 ‘변화 예측형 빅데이터’ 기술로 발전
 - 데이터 간 연계활용, 실시간 분석 능력 향상, 지능·예측형 분석 기술 확보, 빅데이터 수집·처리·저장·분석에 최적의 성능을 얻을 수 있는 기술이 개발되고 있음
 - 정보 획득이라는 기존 빅데이터 기술의 목적이 변화 예측으로 바뀌는 추세
- 빅데이터는 2010년대 이후 인공지능 기술 발전을 촉진하면서 그 중요성이 확대되었으며, 인공지능, IoT, 클라우드와 더불어 지속적으로 발전
 - 빅데이터는 지금까지 양적인 측면이 강조되어 플랫폼과 분석 기술이 발전되었으나, 데이터 개방·유통·공유할 수 있는 플랫폼의 발전과 인공지능 분석 방법론을 적용한 데이터 분석 지능의 다양화로 발전이 전망됨
 - 점차 초연결 기술의 발전과 병행하여 실시간성과 초연결 지능화를 달성하는 방향으로 발전할 것으로 전망

□ 이종 데이터를 통합하여 처리할 수 있는 플랫폼의 개발

- (다양한 데이터 통합) 이종의 빅데이터를 단일시스템으로 처리하는 것은 한계가 있어 이종 데이터를 통합하여 고속 처리할 수 있는 플랫폼이 확대됨
 - Oracle, SAP 등과 같은 글로벌 데이터 관리 기업들은 기존 제품을 확장하여 여러 데이터 모델을 지원하고 있으나 한계가 있어 이종 데이터 모델의 통합이 대두됨
 - 단일지능에서 다양한 데이터를 통합 분석할 수 있는 복합지능으로의 진화를 위하여 다양한 데이터를 통합관리 할 수 있는 기술이 발전하고 있음

□ 신속한 정보 확보를 위한 데이터 처리 가속화 기술의 개발

- (데이터 처리 가속화) 사회 현안 해결 등 다양한 니즈에 부합할 수 있는 실시간성 확보가 중요해짐에 따라 데이터 처리 가속화 기술의 중요성이 확대됨
 - 다양한 센서에서 발생하는 데이터로부터 유의미한 시간 안에 데이터를

처리할 수 있는 기술 개발 중

- 이종의 고성능 컴퓨팅 자원을 저비용으로 사용하는 빅데이터 처리 플랫폼, 데이터 저장관리 기술의 필요성 증대
- 분산원장 데이터의 활용성을 촉진하기 위하여 신뢰성을 가지고 데이터를 빠르게 분석할 수 있도록 하는 분산원장 데이터 관리 기술 출현이 기대됨

□ 데이터 활용을 촉진할 수 있는 데이터 유통 기술의 개발

○ (데이터 유통 활성화) 데이터 수집, 등록, 유통, 거래, 관리기술을 확보하여 데이터 활용의 활성화 촉진 가능성이 높음

- 메타데이터 프레임워크를 활용한 도메인별 빅데이터 확보를 통한 빅데이터 유통 선순환 체계 구축이 예상됨
- 데이터 품질이 인공지능 기술의 가치를 좌우하면서 대량 데이터의 품질을 강화함
 - 기업에서 보유한 데이터의 70~80%가 비정형 데이터로 추정되고 있으며, IBM은 Apache Kafka 오픈소스를 활용하여 IoT 데이터 수집 구축 중

□ 분석 정확도를 높일 수 있는 데이터 분석 알고리즘의 진화

○ (분석 알고리즘 진화) 빅데이터 분석 정확도 및 분석대상 특성에 따른 분석 알고리즘의 진화가 지속적으로 확대되고 있음

- 딥러닝 등 기계학습의 효율성을 높이기 위한 분산형 훈련 방법을 도입할 것으로 보임
 - 구글은 ‘분산형’ Federated Learning 기술을 도입하고 있으며, ‘AI로부터 배우는 AI 기술’의 시발점인 AutoML 연구 진행 중
- 사회 관계망 분석에 주로 사용되던 그래프 분석은 최근 스마트 시티 등 사회 인프라 분석용으로 확대됨
- 최근 처방 분석(Prescriptive Analytics)에 대한 연구가 활발하게 진행
 - IBM은 Decision Optimization이라는 수학적 모델기반 분석 솔루션을 개발하여 상업, 제조, 금융, 의료 등의 분야에 적용을 시도
 - 영국 프로우저 아이오는 복잡하고 역동적인 불확실한 환경에서 자율적이고 적응력이 뛰어난 ‘AI 기반 의사결정 플랫폼’ vuku를 개발 중

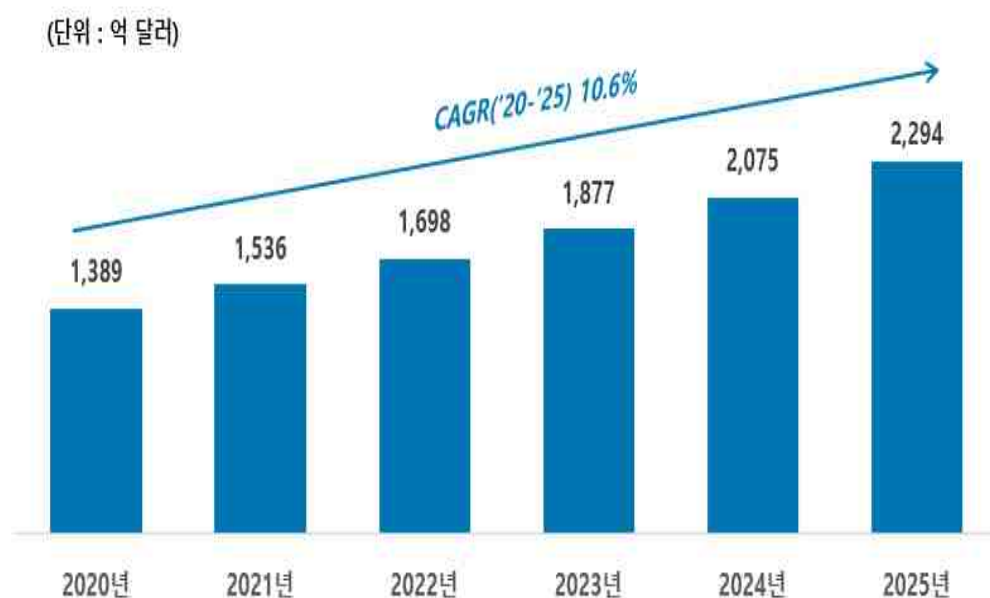
□ 데이터 분석 서비스 활성화를 위한 분석 기술 개발

- (분석 편이성 증대) 데이터 분석 활용 영역이 넓어지면서 빅데이터 분석 서비스가 활성화되고, 분석을 용이하게 할 수 있는 기술이 개발 중
 - 데이터를 분석하고자 하는 업계 수요가 상승하면서, 빅데이터 표준 아키텍처와 클라우드 기반 분석 서비스가 출현
 - 구글(데이터 플로우), 아마존(EMR: Elastic MapReduce) 등 클라우드 서비스 제공업체들은 빅데이터 데이터 저장소뿐만 아니라, 분석 및 관리 기능들까지 개발 제공 중
 - 스마트 시티, 스마트 국가 등 대규모 영역의 복잡한 사회 인프라를 최적화하기 위한 빅데이터를 처리하기 위한 기술들이 발전하고 있음
 - 누구나 쉽게 데이터 분석을 할 수 있도록, 데이터 준비에서부터 시각화 및 해석까지 처리할 수 있는 기술 개발이 필요함

IV. 시장동향

가. 해외 시장 동향

- 전 세계 빅데이터 시장은 2020년 1,388억 8,600만 달러에서 연평균 성장률 10.6%로 증가하여, 2025년에는 2,294억 2,300만 달러에 이를 것으로 전망

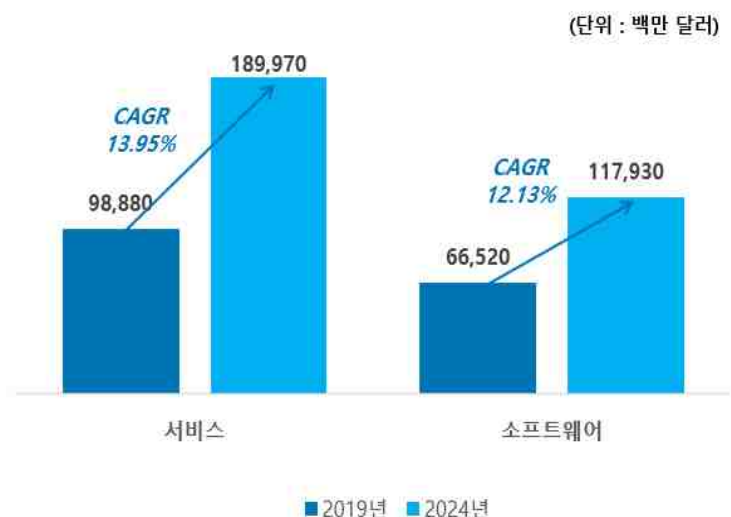


[그림] 글로벌 빅데이터 시장 규모 및 전망

※ 출처 : MarketsandMarkets, Big Data Market, 2020

- 전 세계 빅데이터 시장은 종류에 따라 서비스, 소프트웨어로 분류

- 서비스는 2019년 988억 8,000만 달러에서 연평균 성장률 13.95%로 증가하여, 2024년에는 1,899억 7,000만 달러에 이를 것으로 전망
- 소프트웨어는 2019년 665억 2,000만 달러에서 연평균 성장률 12.13%로 증가하여, 2024년에는 1,179억 3,000만 달러에 이를 것으로 전망



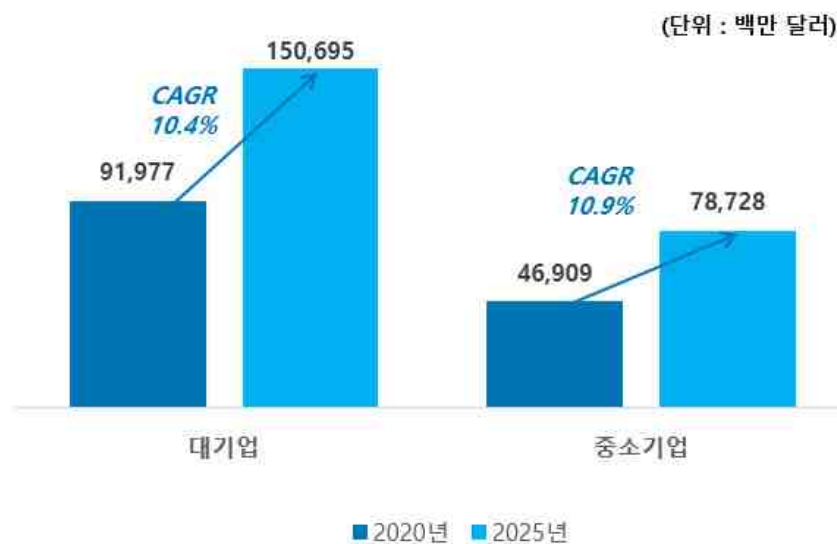
[그림] 글로벌 빅데이터 시장의 종류별 시장 규모 및 전망
 ※ 출처 : TechNavio, Global Big Data Market, 2020

- 전 세계 빅데이터 시장은 배포 방식에 따라 온프레미스(On-premise), 클라우드로 분류
- 온프레미스(On-premise)는 2020년 588억 4,400만 달러에서 연평균 성장률 9.7%로 증가하여, 2025년에는 935억 3,100만 달러에 이를 것으로 전망
 - 클라우드는 2020년 269억 8,100만 달러에서 연평균 성장률 10.9%로 증가하여, 2025년에는 453억 4,600만 달러에 이를 것으로 전망



[그림] 글로벌 빅데이터 시장의 배포 방식별 시장 규모 및 전망
 ※ 출처 : MarketsandMarkets, Big Data Market, 2020

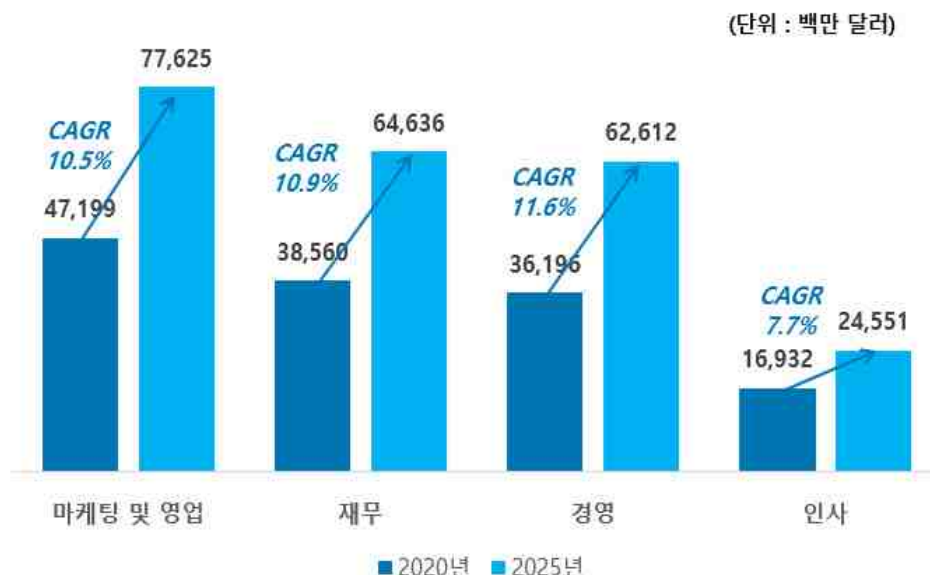
- 전 세계 빅데이터 시장은 조직 규모에 따라 대기업, 중소기업으로 분류
- 대기업은 2020년 919억 7,700만 달러에서 연평균 성장률 10.4%로 증가하여, 2025년에는 1,506억 9,500만 달러에 이를 것으로 전망
 - 중소기업은 2020년 469억 900만 달러에서 연평균 성장률 10.9%로 증가하여, 2025년에는 787억 2,800만 달러에 이를 것으로 전망



[그림] 글로벌 빅데이터 시장의 조직 규모별 시장 규모 및 전망

※ 출처 : MarketsandMarkets, Big Data Market, 2020

- 전 세계 빅데이터 시장은 비즈니스 기능에 따라 마케팅 및 영업, 재무, 경영, 인사로 분류
- 마케팅 및 영업은 2020년 471억 9,900만 달러에서 연평균 성장률 10.5%로 증가하여, 2025년에는 776억 2,500만 달러에 이를 것으로 전망
 - 재무는 2020년 385억 6,000만 달러에서 연평균 성장률 10.9%로 증가하여, 2025년에는 646억 3,600만 달러에 이를 것으로 전망
 - 경영은 2020년 361억 9,600만 달러에서 연평균 성장률 11.6%로 증가하여, 2025년에는 626억 1,200만 달러에 이를 것으로 전망
 - 인사는 2020년 169억 3,200만 달러에서 연평균 성장률 7.7%로 증가하여, 2025년에는 245억 5,100만 달러에 이를 것으로 전망



[그림] 글로벌 빅데이터 시장의 비즈니스 기능별 시장 규모 및 전망

※ 출처 : MarketsandMarkets, Big Data Market, 2020

- 전 세계 빅데이터 시장은 산업 분야에 따라 은행·금융서비스·보험(BFSI), 정부 및 방위, 헬스케어 및 생명과학, 제조, 소매 및 소비재, 미디어 및 엔터테인먼트, 통신 및 IT, 운송 및 물류, 기타로 분류
 - 은행·금융서비스·보험(BFSI)은 2020년 255억 2,000만 달러에서 연평균 성장률 6.6%로 증가하여, 2025년에는 350억 5,800만 달러에 이를 것으로 전망
 - 정부 및 방위는 2020년 149억 4,600만 달러에서 연평균 성장률 12.2%로 증가하여, 2025년에는 266억 3,200만 달러에 이를 것으로 전망
 - 헬스케어 및 생명과학은 2020년 97억 2,900만 달러에서 연평균 성장률 12.5%로 증가하여, 2025년에는 175억 2,800만 달러에 이를 것으로 전망
 - 제조는 2020년 133억 6,600만 달러에서 연평균 성장률 13.3%로 증가하여, 2025년에는 249억 900만 달러에 이를 것으로 전망
 - 소매 및 소비재는 2020년 202억 5,500만 달러에서 연평균 성장률 10.7%로 증가하여, 2025년에는 336억 9,600만 달러에 이를 것으로 전망
 - 미디어 및 엔터테인먼트는 2020년 140억 4,100만 달러에서 연평균 성장률 11.0%로 증가하여, 2025년에는 236억 1,200만 달러에 이를 것으로 전망

- 통신 및 IT는 2020년 159억 4,000만 달러에서 연평균 성장률 9.2%로 증가하여, 2025년에는 247억 9,200만 달러에 이를 것으로 전망
- 운송 및 물류는 2020년 113억 달러에서 연평균 성장률 11.9%로 증가하여, 2025년에는 197억 9,700만 달러에 이를 것으로 전망
- 기타는 2020년 137억 8,900만 달러에서 연평균 성장률 11.2%로 증가하여, 2025년에는 234억 달러에 이를 것으로 전망

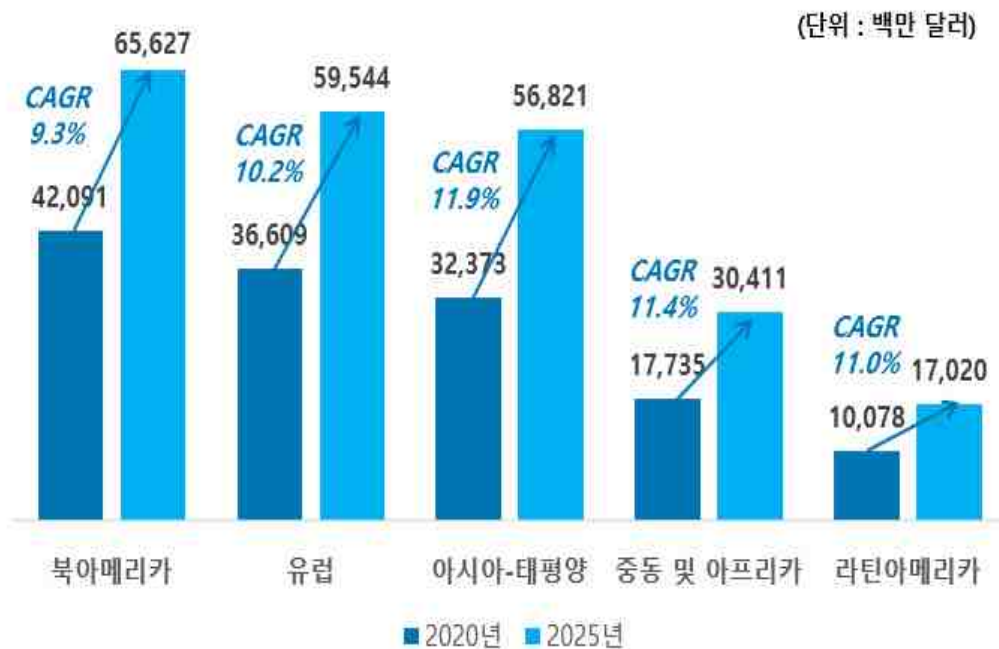


[그림] 글로벌 빅데이터 시장의 산업 분야별 시장 규모 및 전망

※ 출처 : MarketsandMarkets, Big Data Market, 2020

- 전 세계 빅데이터 시장을 지역별로 살펴보면, 2020년을 기준으로 북아메리카 지역이 30.3%로 가장 높은 점유율을 차지
- 북아메리카는 2020년 420억 9,100만 달러에서 연평균 성장률 9.3%로 증가하여, 2025년에는 656억 2,700만 달러에 이를 것으로 전망
- 유럽은 2020년 366억 900만 달러에서 연평균 성장률 10.2%로 증가하여, 2025년에는 595억 4,400만 달러에 이를 것으로 전망

- 아시아-태평양은 2020년 323억 7,300만 달러에서 연평균 성장률 11.9%로 증가하여, 2025년에는 568억 2,100만 달러에 이를 것으로 전망
- 중동 및 아프리카 지역은 2020년 177억 3,500만 달러에서 연평균 성장률 11.4%로 증가하여, 2025년에는 304억 1,100만 달러에 이를 것으로 전망
- 라틴아메리카 지역은 2020년 100억 7,800만 달러에서 연평균 성장률 11.0%로 증가하여, 2025년에는 170억 2,000만 달러에 이를 것으로 전망

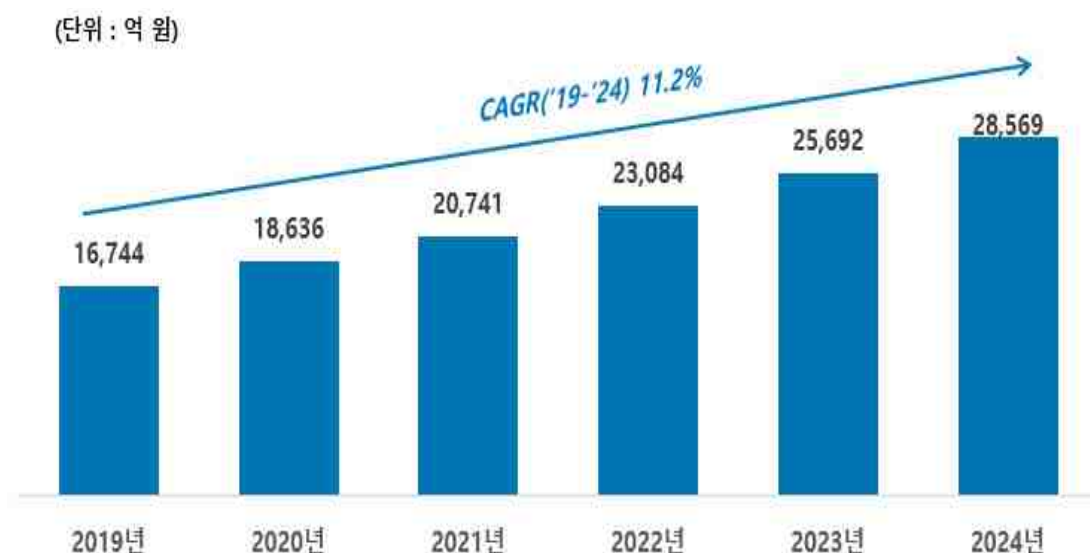


[그림] 글로벌 빅데이터 시장의 지역별 시장 규모 및 전망

※ 출처 : MarketsandMarkets, Big Data Market, 2020

나. 국내 시장 동향

- 국내 빅데이터 시장은 2019년 1조 6,744억원에서 연평균 성장률 11.2%로 증가하여, 2024년에는 2조 8,569억 원에 이를 것으로 전망



[그림] 빅데이터 국내 시장 규모 및 전망

※ 출처 : 국내 빅데이터 및 분석 2018-2024 전망, 한국IDC, 2019

- 국내 빅데이터 시장은 제품별 빅데이터 인프라(서버, 스토리지), 빅데이터 소프트웨어, 빅데이터 서비스로 나뉨
 - 빅데이터 인프라 시장이 55.9%로 가장 큰 부분을 차지하고 있으나, 소프트웨어(23.2%), 서비스(20.9%) 시장은 활성화되고 있지 못한 것으로 나타남
- 기업의 빅데이터에 대한 인식 호전과 중견·대기업의 투자 증가, 정부의 강력한 빅데이터 산업 육성 의지에 따라 성장세를 보이고 있으며, 결과적으로 시장이 확대됨
 - 국내 빅데이터 시장은 포털사, 통신사, IT 업체 등이 자사 보유 데이터를 바탕으로 빅데이터 서비스 제공을 시작하는 단계

V. 산업동향

1. 국내/외 산업 동향

□ 빅데이터 기반의 AI 기술과의 융복합을 통한 사업 영역 확대

- 아마존(Amazon)은 아마존 웹서비스(AWS)를 통해 클라우드와 연계한 빅데이터 제품의 시장 지배력을 강화하고, 구글(Google)은 폭넓은 빅데이터 제품 구축 및 기계학습 관련 커뮤니티인 캐글(Kaggle)을 인수하여 새로운 기계학습 API 구축을 진행
- 마이크로소프트(Microsoft), 오라클(Oracle) 등은 클라우드, AI와 연계된 빅데이터 서비스를 추진
- IBM과 Salesforce, SAP와 Google처럼 ‘데이터(리포지토리) 보유 기업’ 과 ‘인공지능기술 보유 기업’ 간의 협력관계 형성이 증가

□ 빅데이터 스타트업의 우수기술력을 바탕으로 글로벌기업과의 M&A 증가

- 2018년 상반기 기준으로 전 세계 빅데이터 관련 스타트업이 약 3,500여 개로 집계되었으며, 우수 기술을 보유한 스타트업은 꾸준히 M&A가 이루어짐
- 2018년 8월, 데이터 분야에 3,538개의 스타트업과 19,323명의 엔젤 투자자가 있으며, 기업당 평균 투자금액은 460만 달러에 이름
- 벤처캐피탈리스트의 빅데이터 기업투자는 꾸준히 이루어지고 있으며, 2017년 대규모 자금이 투자된 기업은 Looker(\$81M), InsideSales(\$50M), DataRobot(\$54M), Confluent(\$50M), DataBricks(\$60M) 등이 있음
- 구글(Google)은 빅데이터 분석업체인 루커(Looker)를 26억 달러에 인수하여 클라우드 컴퓨팅 사업을 강화하였으며, 마이크로소프트(Microsoft)는 소셜 서비스 플랫폼인 틱톡 인수를 추진 중

[표] 주요 스타트업 M&A 현황(2016년 및 2017년)

연도	스타트업		인수 기업	인수 금액
	기업명	주력 기술		
2017	Mobileye	비전기반 자율주행 기술	Intel	\$15.3B
	AppDynamics	애플리케이션 성능 관리 기술	Cisco	\$3.7B
	Nimble Storage	예측 플래시 저장 기술	HPE	\$1.1B
	Kaggle	빅데이터 예측분석 플랫폼	Google	-
2016	Qlik	비즈니스 인텔리전스	Thoma Bravo	\$3B
	Cruise Automation	비즈니스 인텔리전스	General Motors	\$1B
	Apigee	API 매니지먼트	Google	\$625M
	OPower	클라우드 기반 에너지 관리 플랫폼	Oracle	\$532M

※ 출처 : MarketsandMarkets, Big Data Market, 2020

○ 크런치베이스* 조사 결과 2019년 10월에는 전 세계 빅데이터 스타트업이 9,746개로 집계되었으며 이들 기업에 투자된 자금만 805억 7,953만 달러 (약 94조 원)에 달함

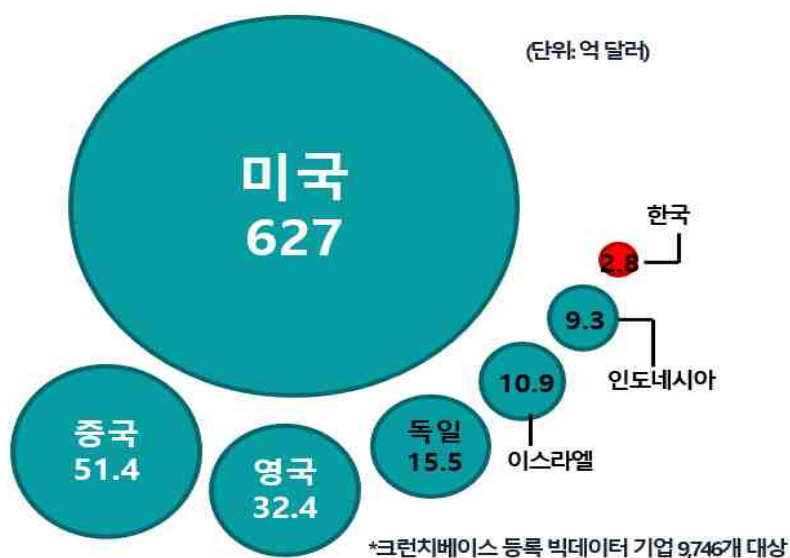
* 빅데이터 기반 스타트업 분석 플랫폼

- 이들 기업은 빅데이터 솔루션, 컨설팅, 서비스 등을 제공하는 스타트업으로 빅데이터를 간접적으로 활용하는 기업까지 합하면 더 많을 것으로 예상
- 빅데이터 스타트업 9,746개 중 미국에 본사를 둔 기업은 5,035개로 51.7%의 점유율에 달함

□ 미국과 중국을 중심으로 전 세계 빅데이터 기업에 대한 투자 증가

○ 미국 기업은 투자 유치 총액이 627억 달러 이상을 차지했으며, 투자 유치액에서 절대적 지위를 차지

- 중국은 막대한 인구를 바탕으로 자국 내 방대한 데이터를 활용해 국가 주도로 활용하고 있으며, 기업당 투자 유치액에서 1위를 차지
- 중국 기업의 기업당 투자 유치액은 2,663만 달러, 미국은 1,246만 달러로 2배 이상 차이를 보임



[그림] 국가별 빅데이터 스타트업 투자금액

※ 출처 : 빅데이터 혁명 마중 경쟁에 이미 뒤쳐진 韓 'AI 강국'도 요원., 매일경제 2019, 네이버 재가공

- 국내 기업의 빅데이터 관련 투자가 확대되고 매출 또한 증가하고 있으나, 관련 인프라 및 글로벌 선도 기업이 부족한 상황
- 국내 기업의 빅데이터 시스템 도입률은 2016년 약 50개 기업에서 2017년 약 65개 기업으로 증가하였고 평균 투자금액은 2016년 5억 2,100만 원에서 2017년 6억 100만 원으로 증가

2. 국내/외 기업 동향

가. 해외 기업 동향

□ 빅데이터 선도기업

○ Microsoft (미국)

- 지능형 클라우드 및 지능형 에지 시대에 경쟁할 소프트웨어, 서비스, 장치 및 솔루션을 개발하고 있음
- Microsoft의 제품에는 고객에게 소프트웨어, 플랫폼, 콘텐츠를 제공하고 솔루션 지원 및 고객을 위한 컨설팅 서비스를 제공하는 클라우드 기반 솔루션이 포함됨
- 2015년 11월, Cortana Analytics 제품군으로 Big Data Mainstream을 출시하였으며, 2016년 1월, 오픈소스 R 프로그래밍 언어를 기반으로 한 고급 분석플랫폼을 도입

○ AWS (미국)

- 아마존(Amazon)의 자회사이며 주로 웹 서비스 형태로 클라우드 컴퓨팅 서비스를 제공
- 아마존(Amazon)의 제품 포트폴리오는 컴퓨팅, 스토리지, 데이터베이스, 네트워크 및 콘텐츠 제공, 개발자 도구, 관리 도구, 미디어 서비스, 분석 등으로 구성

○ Oracle (미국)

- 2015년 8월, Oracle Data Integrator for Big Data를 출시했으며, 2015년 10월, 빅데이터를 위한 Oracle 클라우드 플랫폼을 출시했음
- Oracle은 플랫폼, 애플리케이션 및 인프라와 같은 기업 IT 환경의 요구 사항을 충족하도록 설계된 광범위한 제품, 솔루션 및 서비스를 제공
- Oracle의 제품 포트폴리오는 자율 데이터베이스, 개발자 도구, 엔지니어링 시스템, IT 인프라, Java 및 미들웨어로 구성

○ Google (미국)

- Google의 제품 혁신은 서비스 사용을 증가시켰으며 세계에서 가장 인정받는 브랜드 기업 중 하나로 자리 잡음
- 핵심 제품 및 플랫폼에는 Android, Chrome, Gmail, Google 드라이브, Google 지도, Google Play, 검색 및 YouTube가 포함되며, 각 제품에는 매달 10억 명 이상의 활성 사용자가 있음

○ IBM (미국)

- 2017년 기준 IBM 빅데이터 매출은 26억 5,700만 달러로 빅데이터 시장에서 1위를 기록
- 실험 기반에서 디지털 재창조에 초점을 맞춘 하이브리드 클라우드 기반 AI 기반 애플리케이션으로 개발을 전환
- 인공지능 기반의 빅데이터 분석 프로그램 ‘왓슨(Watson)’을 출시하였으며, 주로 활용되는 분야는 의료 분야로 방대한 양의 의료 데이터를 인공지능 기술을 활용하여 처리 및 분석하여 질병 연구에 적용
- 2017년 인도 수자원 문제의 해결을 위해 IBM 빅데이터 기술을 제공하였으며, 지능형 물 관리 소프트웨어를 사용하는 ‘수자원 관리 센터’를 설립하고 모든 물 사용 데이터를 모아 관리체계를 일원화하였음

○ Splunk (미국)

- 2003년에 설립된 빅데이터 소프트웨어 기업으로 빅데이터 검색과 모니터링, 분석 소프트웨어를 개발
- 데이터 패턴을 분석하고 문제를 진단하여 조직 전체가 데이터에 접근할 수 있는 시스템을 제공
- 2018년 IT 운영관리 시스템 제공업체 중 가장 큰 매출 증가율을 달성하였고 빅데이터 기반 보안관제 시스템(SIEM, Security Information and Event Management) 부문에서도 점유율 1위를 기록
- 다양한 기업들과의 인수합병을 통해 회사 서비스 분야를 넓혀가고 있는 스플링크는 2018년 보안 오케스트레이션 회사 팬텀사이버(Phantom Cyber)를 인수
- ‘SOAR’ 기술을 활용해 차세대 데이터 분석 기반의 보안 기술을 구현할 것으로 예상

○ Dell Technologies (미국)

- 2019년 6월 기준 글로벌 서버 시장 매출은 810억 달러이며 벤더 별 시장 점유율은 1위를 기록
- 2015년 컴퓨터 수요 감소로 인한 수익 악화 상황에서 데이터 스토리지 분야로 수익원을 확대하기 위해 스토리지 전문업체 EMC를 인수
- 2019년 호주 퀸즐랜드 대학교의 알츠하이머 연구에 데이터 시각화와 기계학습 등의 작업을 포함하여 대량의 계산 작업을 병렬로 처리하는데 적합한 델테크놀로지의 고성능 컴퓨터 시스템이 도입

○ SAP SE (독일)

- 2014년 11월, 엔터프라이즈 분석 데이터 관리 분야의 선두업체인 클라우데라와 파트너십을 맺었으며, 2015년 1월, 새로운 SAP HANA 도구인 SAP HANA Big Data Intelligence를 출시했음
- 응용 프로그램, 분석 소프트웨어 및 소프트웨어 관련 서비스를 제공하는 주요 공급업체 중 하나임
- SAP ERP, SAP Business Warehouse, SAP Business Object Software, Sybase 모바일 제품 및 인메모리 컴퓨팅 어플라이언스 SAP HANA 등을 제공하고 있음

□ 빅데이터 유망기업

○ Palantir Technologies (미국)

- 빅데이터 분석을 전문으로 하는 미국 소프트웨어 회사로, ‘페이팔 마피아 (PayPal mafia)’ 중 한 명으로 알려진 피터 티(Peter Thiel) 회장이 설립
- 정부 기관과 기업을 대상으로 다양한 빅데이터 분석 서비스를 제공하고 있으며, 팰런티어고담(Palantir Gotham)과 팰런티어파운드리(Palantir Foundry) 두 종류의 서비스를 제공
- 미 중앙정보국(CIA, CentralIntelligence Agency)에 범죄 예측 분석 소프트웨어를 제공하고 있으며 미 연방 마약국(DEA, Drug Enforcement Administration)이 보유한 방대한 자료를 분석하여 대규모 마약 조직 소탕 작전에도 기여

[표] Palantir Technologies 서비스 종류

서비스명	주요 내용
팔런티어고담 (Palantir Gotham)	• 정부기관이 주 고객 • 복잡한 네트워크상에 숨겨진 부정행위 적발에 활용 • 테러조직 검거, 자금세탁 방지, 밀수추적 등에 사용
팔런티어파운드리 (Palantir Foundry)	• 상업 부문에 활용하기 위해 개발한 서비스 • 정부기관 대비보다 복잡하고 비체계적인 일반 기업 데이터에 대응하기 위하여 분석력 강화 • 매일 생성되는 데이터를 기반으로 분석이 이루어지기 때문에 실질적인 의사결정에 도움을 줌 • 제약회사의 임상실험 계획 수립, 제조기업의 효율적인 공정 수립 및 품질관리, 광고 타겟 결정, 결제 데이터 분석 등에 사용

※ 출처 : 글로벌 빅데이터 시장 보고서, 정보통신산업진흥원, 2019

○ Uptake Technologies (미국)

- 사물 인터넷을 활용해 다양한 산업 분야에서 이용되는 기계와 장비의 센서 데이터를 분석한 뒤 소프트웨어를 통해 오류 발생 전 예측할 수 있도록 도와주는 시스템을 제공
- 농업과 빌딩, 건축 및 건설, 에너지, 정부기관, 군대, 제조, 광산 등이 주요 활용 분야임
- 2018년 미 육군에 100만 달러 규모의 기계 오류 예측 소프트웨어 납품 계획을 발표하였으며 2019년 미 해병대에 자사 Asset IO 애플리케이션을 M88 장갑차의 고장 예측에 활용할 것이라고 발표
- 자사 인공지능과 기계학습 알고리즘을 브래들리 M2A3에 적용하여 장갑차 고장 유무를 사전에 파악할 수 있도록 함

[표] Uptake Technologies 서비스 종류

서비스명	주요 내용
Asset IO	기계 및 장비를 비롯한 개인 자산의 데이터 관리 시스템
Dealer Suite	산업용 플랫폼으로, 고객의 자산 관리에 필요한 플랫폼 서비스

※ 출처 : 글로벌 빅데이터 시장 보고서, 정보통신산업진흥원, 2019

나. 국내 기업 동향

○ 바이브컴퍼니

- 2000년 7월 다음커뮤니케이션에서 분사한 인공지능 기업으로 빅데이터 자료를 수집하고 분석해 인사이트를 도출해내는 국내 최초이자 최대 빅데이터 전문기업
- NLP 연구소를 설립해 정형 데이터뿐만 아니라 소셜 빅데이터(인스타그램, 트위터, 블로그, 뉴스 등) 비정형 데이터까지 수집·분석함으로써 기업, 기관, 개인 등 고객사에 최적의 맞춤형 솔루션을 제안하는 서비스를 제공
- 인공지능·빅데이터 플랫폼 ‘소피아(SOFIA)’를 구축하고 빅데이터 분석, 문제 해결 솔루션, AI 리포트, 비대면 전문 상담 서비스 등을 개발



[그림] 소피아(SOFIA) 플랫폼

※ 출처 : 바이브컴퍼니 홈페이지

○ 알체라

- 2016년 설립된 AI 영상인식 전문기업으로, 대용량 영상데이터 수집·편집부터 딥러닝 학습, 인공지능 엔진 최적화, 운용 앱 제작까지 전 과정의 토털 솔루션을 제공
- 얼굴인식, 이상 상황 감지 AI 기술을 보유하고 있으며, 이 두 기술은 인천공항 자동 출입국 시스템, 외교부 여권 정보 통합관리 시스템, 금융권 여권 위조 여부 검출 시스템, 신한카드 Face Pay, 한국전력 화재 감시

시스템 등 다양한 영역에서 활용되고 있음

- AIIR Pass Pro는 AI 기반 출입관리 솔루션으로 다수 인원이 동시에 출입하는 상황에서 발생할 수 있는 대기 시간과 정체 현상을 관리하고, AIIR ID는 클라우드 기반 본인 인증 솔루션으로 신분증 사진과 사용자 얼굴을 비교하여 사용자의 신분 확인이 가능함

○ 씨이랩

- 2010년에 설립된 기업으로, 글로벌 AI 선두 기업인 엔비디아(NVIDIA)의 프리퍼드 파트너(Preferred Partner)로 영상 AI 기술에 필요한 엔비디아 그래픽처리장치(GPU) 시스템 공급, 소프트웨어 개발, 데이터 유통 사업 등을 수행
- 아큐튜닝 for 씨이랩(XIII Lab)을 출시해 AI 통합기술 수준을 한 단계 더 높였으며, 이 기술은 별도의 학습 없이 빅데이터 전처리와 영상 속 자연어 자동분류, 자동 데이터 탐색·분석(EDA), 최적의 분류 예측 모델 자동 생성 등을 제공
- 대용량 데이터 처리 기술을 접목한 영상데이터 중심 AI 분석 서비스 ‘X-VIVA(AI Video Analytics)’를 출시



[그림] X-VIVA 개요

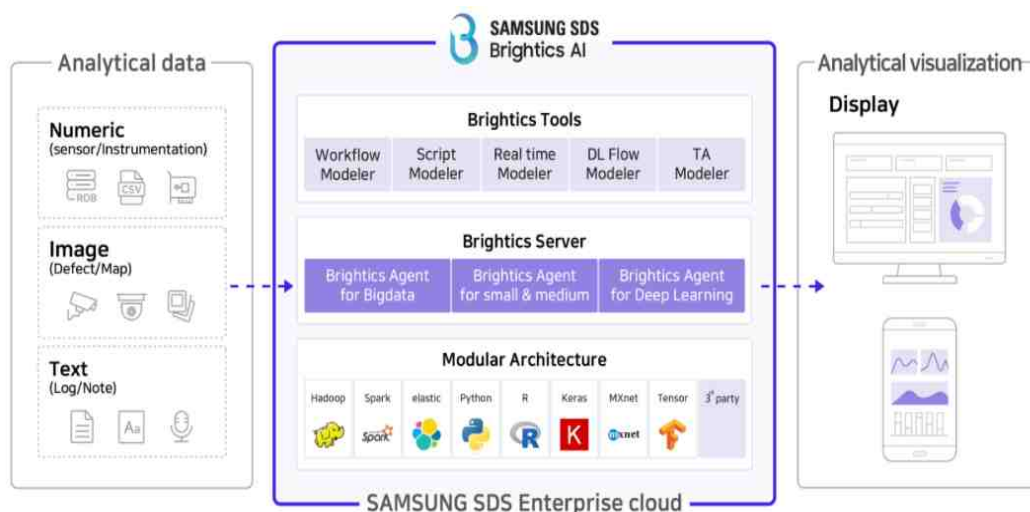
※ 출처 : 씨이랩 홈페이지

○ LG CNS

- 인공지능(AI) 기술을 활용해 데이터 수집에서 분석까지 한 눈에 보여주는 빅데이터 플랫폼 ‘DAP’ 를 출시하였고, DAP는 빅데이터 분석 환경 구축에 걸리는 시간을 1시간 내, 3~4개월가량 걸리던 분석작업 시간을 1~2주로 단축
- 아마존의 클라우드 시스템인 ‘아마존웹서비스(AWS)’ 와 연계해 더 다양한 기능을 제공하고 국내 유망 스타트업과 협업을 통해 추가적인 기능도 추가할 계획을 발표

○ 삼성 SDS

- 2016년 6월, AI 기반 플랫폼인 ‘브라이틱스’ 를 선보였으며, 이 플랫폼은 3시간 이상 걸리는 데이터 수집·분석과 시각화 과정을 10분 내로 단축
- 브라이틱스는 제조업과 마케팅·물류·보안·사물 인터넷(IoT) 등 다양한 분야에서 여러 기업에 적용되었으며, 향후 서비스·금융 등을 범위를 넓혀갈 예정
- 브라이틱스 AI 플랫폼을 활용한 미래 AI 전문가 양성을 위해 노력하고 있으며, 2020년 글로벌 컨설팅 기관 포레스터에 이어 2021년 가트너의 머신 러닝 플랫폼 부문에서도 한국 기업 최초로 등재됨



[그림] 브라이틱스 AI 개요

※ 출처 : 삼성SDS 홈페이지

○ SK C&C

- 제조업 분야에 특화된 AI 기반 빅데이터 분석플랫폼 ‘스키테일’을 구축해 기업에 제공하였고 2016년 4월, AI 솔루션 전문 개발업체 ‘베가스’와 양해각서를 맺고 최신 알고리즘을 적용
- 가명 정보에 기반한 원스톱 고객 맞춤형 ‘토탈 데이터 결합 서비스’를 공개하였으며, 이 서비스는 데이터 결합 활용 컨설팅, 데이터 수집·연계, 가명 처리 결합·분석 등을 제공
- 데이터 시각화 전문업체인 태블로와 산업별 시스템 기반의 빅데이터 분석·활용 시장 공동 발굴을 위한 전략적 업무 제휴를 맺었으며, SK C&C의 산업별 시스템 구축 및 운용 역량과 아큐인사이트 플러스 등 빅데이터 플랫폼에 태블로의 데이터 시각화 전문 기술을 결합하여 산업 현장의 빅데이터 분석 및 활용 수준을 높이기 위해 노력

[참고문헌]

- 2021년 빅데이터 활용 사례 10가지, 업계별 추천한다, FineReport, 2021
- 민관협력기반 데이터 플랫폼 발전전략 발표 보도자료, 과학기술정보통신부, 2021
- 삼성SDS, 빅데이터 분석 전문가 키운다, 머니투데이, 2021
- 인공지능과 빅데이터 기술 동향, TTA Journal, 2020
- MarketsandMarkets, Big Data Market, 2020
- TechNavio, Global Big Data Market, 2020
- “디지털뉴딜, 물 들어올 때 노 젓자”...빅데이터 기업들 줄줄이 IPO 도전, 한국경제, 2020
- 중소기업 전략기술로드맵 2021-2023 빅데이터, 중소기업기술정보진흥원, 2020
- 2018 빅데이터 시장현황 보고서, 한국정보화진흥원, 2019
- 인공지능(빅데이터) 시장 및 기술 동향, 과학기술일자리진흥원, 2019
- 국내 빅데이터 및 분석 2018-2024 전망, 한국IDC, 2019
- 빅데이터 시장, 연구개발특구진흥재단, 2019
- 2019 데이터 산업 백서, 한국데이터산업진흥원, 2019
- 글로벌 빅데이터 시장 보고서, 정보통신산업진흥원, 2019
- 빅데이터 특허 동향, 한국특허전략개발원 심사지원센터, 2019
- 빅데이터 혁명 미·중 경쟁에 이미 뒤쳐진 韓 ‘AI 강국’도 요원... 데이터 거래 활성화해야, 美 민간 시장 200조원, 매일경제, 2019
- 빅데이터, 한국과학기술기획평가원, 2018
- 빅데이터 정책 추진 현황과 활용도 제고방안, 국회입법조사처, 2018
- ICT R&D 기술로드맵 2023, 정보통신기획평가원, 2018
- “AI 빅데이터 플랫폼 선점 “ 막 오른 SI기업 ‘삼각 경쟁’, 서울경제, 2017
- Technology Advancements Shaping Big Data Progress, Frost&sullivan, 2017
- 데이터산업 백서, 한국데이터진흥원, 2017
- Technology Advancements Shaping Big Data Progress, Frost&sullivan, 2017
- 빅데이터 컴퓨팅 기술, 한빛아카데미, 2016
- 개인정보 비식별 조치 가이드라인, 국무조정실, 2016
- Definitions, Concepts, and Theoretical Approaches, Edward Curry, 2015

- 빅데이터 기술 및 시장 동향, 정보통신기술진흥센터, 2014