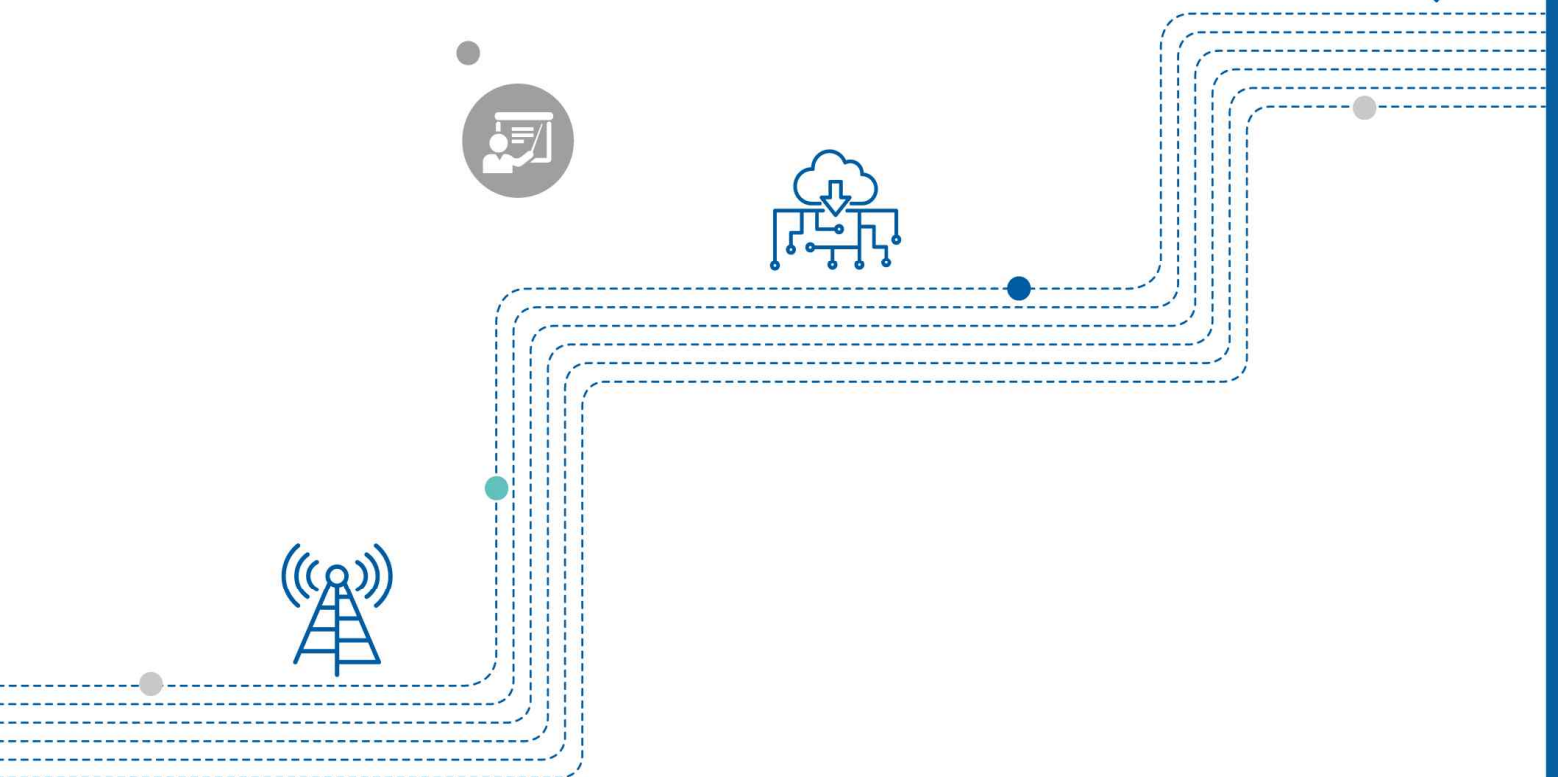


빅데이터 플랫폼의 산업생태계 현황과 주요 이슈

김문구 · 박종현



본 보고서는 ETRI 기술경제연구본부 주요사업인 “ICT R&D 경쟁력 제고를 위한 기술경제 연구”의 결과물입니다.



* 본 보고서의 내용은 연구자의 견해이며 ETRI의 공식 의견이 아님을 알려드립니다.



Contents



요약	1
I. 개요	7
1. 정의 및 범위	7
2. 기술 분류 및 대상 시장	10
II. 시장 및 산업생태계 현황	13
1. 시장 동향 및 전망	13
2. 빅데이터 플랫폼 생태계 현황	18
III. 정책 및 기술 동향	25
1. 국내외 정책 동향	25
2. 기술 동향 및 전망	29
IV. 주요 이슈 및 핵심가치	35
1. 주요 이슈	35
2. 핵심가치 및 R&D 시사점	43
참고문헌	47

표목차



[표 1] 빅데이터 플랫폼 분야 기술 범위	10
[표 2] 국내 빅데이터 플랫폼 시장 규모 전망	17
[표 3] 세계 빅데이터 플랫폼 시장 규모 전망	17
[표 4] 빅데이터 플랫폼 국내외 기업 현황	19
[표 5] 국외 빅데이터 플랫폼 기업 현황	19
[표 6] 국내 빅데이터 플랫폼 기업 현황	21
[표 7] 국내외 주요국의 빅데이터 R&D 투자 현황	31
[표 8] 국내 빅데이터 투자 현황(2015~2017년)	32
[표 9] 빅데이터 플랫폼 주요 이슈	35
[표 10] 국내 직무별 빅데이터 전문인력 현황 및 전망	38
[표 11] 국내 공공데이터 미활용 이유	39

그림목차



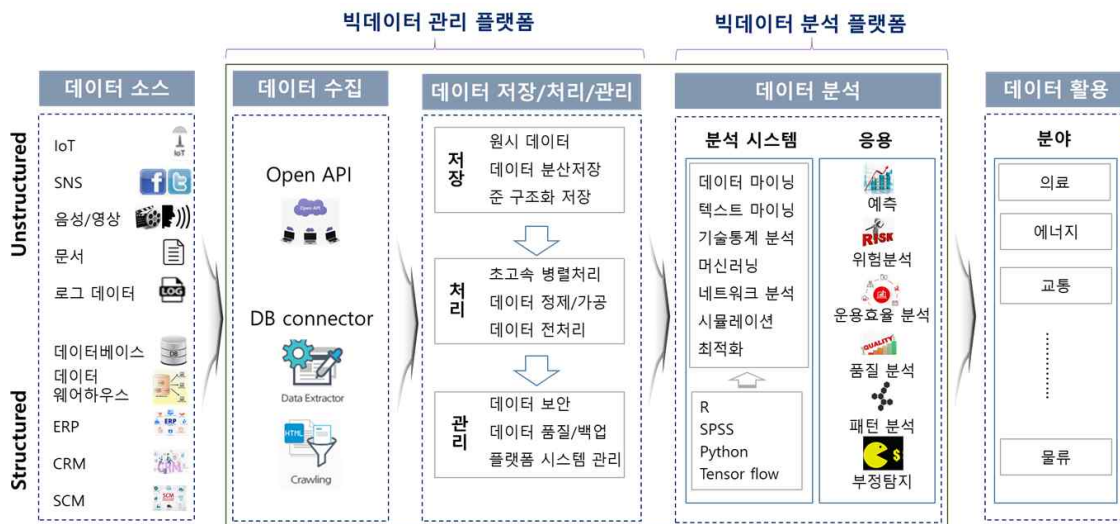
[그림 1] 빅데이터 플랫폼	7
[그림 2] 빅데이터 플랫폼 대상 시장	11
[그림 3] 빅데이터 기술수준 및 기술격차 현황	33
[그림 4] 주요 국가별 빅데이터 기술수준 격차 변동 추이	33
[그림 5] 국내 기업의 빅데이터 활용 현황	36
[그림 6] 국내 빅데이터 전문인력 채용 요건	38
[그림 7] 빅데이터 활용과 미활용 이유 간 상관관계	40
[그림 8] 빅데이터 및 타 기술과의 융합	42
[그림 9] 빅데이터 플랫폼 활용 사회문제 해결	43

요약

개요

[정의 및 범위]

- 빅데이터 플랫폼은 데이터의 수집·저장·처리·관리 및 분석 등의 역할 수행을 지원함으로써 새로운 인사이트(Insight)와 비즈니스 가치 창출이 가능한 데이터 프로세스 환경을 의미
 - 빅데이터 플랫폼은 데이터의 수집·저장·처리·관리를 담당하는 빅데이터 관리 플랫폼과 데이터 분석을 지원하는 빅데이터 분석 플랫폼으로 구분
 - 빅데이터 플랫폼의 핵심 역할은 원시데이터로부터 새로운 통찰력과 가치의 창출



[대상 시장]

- 빅데이터 플랫폼 시장은 크게 관리 영역과 분석 영역으로 구분 가능
 - 빅데이터 관리 영역: 데이터의 수집에서부터 처리·저장·품질관리를 위한 소프트웨어 영역
 - 빅데이터 분석 영역: 데이터 분석을 위한 소프트웨어 영역



※ 교육, 의료, 컨설팅, 물류 등 빅데이터를 활용한 서비스 제공 분야는 본 빅데이터 플랫폼 대상 시장에 미포함

시장 동향 및 전망

[니즈 변화]

- (데이터 증가) 사물인터넷, 스마트 디바이스, 소셜 미디어 서비스 등의 활성화로 비정형 데이터 규모가 빠르게 증가
- (데이터 경제) 데이터 중심의 새로운 경제 패러다임 도래로 데이터가 경제적 가치 창출의 핵심 원유로 부상
- (수요 증대) 새로운 가치 창출의 추동으로서 빅데이터 플랫폼에 대한 수요 증대

[성장 동인]

- (데이터 관리 기술 역량 확보) 실시간 데이터의 전송·처리를 위한 기술역량 강화 및 안정적인 네트워크 인프라 고도화
- (데이터 활용 생태계 성장) 사물인터넷의 빠른 확산으로 인한 빅데이터 이용 환경 확대
- (인공지능 연계 분석 기술 고도화) 인공지능 연계의 빅데이터 분석 기술의 정확도 향상 및 분석 대상 특성에 따른 분석 알고리즘의 지속적인 진화
- (정부 지원) 빅데이터 분석 결과의 활용성 제고를 위한 정부의 정책적 지원 확대

[국내외 시장 전망]

- 국내 빅데이터 플랫폼 총 시장 규모는 '15년 1.39조 원에서 연평균 8.1%의 성장을 통해 '25년 약 2.2배 증가한 3.05조 원의 시장이 형성될 것으로 전망
- 세계 빅데이터 플랫폼 총 시장 규모는 '15년 69.6조 원에서 '25년 152.5조 원의 시장으로 확대될 전망

[주요 기업 현황 및 전략]

- 빅데이터 플랫폼 기업들은 대용량 데이터의 실시간 처리, 다양한 파일 시스템에 대한 접근성 강화, 클라우드와 오픈 소스 기반 빅데이터 솔루션 연계 등을 통해 빅데이터 기반 기업 생태계 형성 및 산업 성장에 자원과 역량을 집중
- 빅데이터 플랫폼 기업의 핵심 기능에 따라 빅데이터 관리와 분석 기업으로 구분
 - (빅데이터 관리 분야 플랫폼 기업) 하둡을 기반으로 데이터 수집, 분산 컴퓨팅 및 병렬 처리에 집중
 - ※ (국외) Google, IBM, MS, Amazon, Oracle, SAP, Intel, Cisco, Splunk, Teradata, (국내) 넥스알, 데이터스트림즈, 선재소프트 등
 - (빅데이터 분석 분야 플랫폼 기업) 데이터 마이닝, 텍스트 마이닝, 인공지능 알고리즘 등의 분석 툴을 활용하여 가치 있는 솔루션 제공
 - ※ (국외) Google, IBM, MS, Amazon, Oracle, SAP, TIBCO Spotfire, Tableau Software, (국내) 다음소프트, 그루터, 사이람, 솔트룩스, 마인즈랩 등

정책 및 기술 동향

[정책 동향]

- (빅데이터 관리) 미국을 비롯한 주요국은 빅데이터 수집·저장·처리·관리 및 공공 데이터의 공유·개방 정책을 통해 빅데이터로부터 새로운 가치창출을 위한 기반 환경·인프라 구축에 역량을 집중하고 있음
- (빅데이터 분석) 수집된 데이터의 가치 실현을 위해 분석기술 강화, 산업에의 활용·응용 확대 및 빅데이터 전문가 양성 등에 자원과 역량을 집중함으로써 새로운 혁신적인 비즈니스 창출을 촉진하는 정책 추진

[기술진화의 특성]

- (복합 데이터 처리) 단일 데이터 분석 중심에서 다양한 데이터를 통합 분석 가능한 다중 지능 발전 요구 및 이를 위한 빅데이터 플랫폼 기술로 발전 전망



- (분석의 편리성 증대) 빅데이터 분석 전문가 외에 일반인도 데이터 분석을 쉽게 할 수 환경 조성
- (데이터 분산 관리·분석 확대) 초연결 시대로 발전함에 따라 중앙 서버에서 모든 데이터를 관리하는 것에서 분산된 형태의 저장 관리 기술로 발전 전망
- (빅데이터 분석 알고리즘의 지속적인 진화) 기계학습, 딥러닝 등 인공지능 알고리즘의 진화로 보다 효율적이고 정확한 빅데이터 분석 기술 발전 전망
- (분석 기법의 다양화) 최근 유튜브, 소셜네트워크서비스 등 비정형데이터의 빠른 증가로 인해 텍스트·영상 기반 데이터 분석기법 확산

[기술역량 수준]

- 우리나라의 빅데이터 분야 기술수준은 최고기술 보유국(미국) 대비 79.0%, 1.6년의 기술격차 발생
 - 우리나라는 일본(82.8), 유럽(87.0)등 선진국뿐만 아니라 중국(82.5)에도 기술이 열위한 상황에 직면
 - 빅데이터 분석·예측 기술은 타국 대비 상대적으로 기술수준이 더 낮게 나타남



주요 이슈 및 핵심가치

[빅데이터 플랫폼 주요 이슈]

- 빅데이터 플랫폼을 둘러싼 다양한 이슈들이 제기됨에 따라 이를 산업생태계, 정책·제도, 기술·R&D 등의 3개 분야로 구분, 각 분야별 핵심 이슈 요약

분야	주요 이슈
산업생태계	<ul style="list-style-type: none"> ■ 협업의 생태계 환경 미흡 ■ 글로벌 빅데이터 플랫폼 기업 부재 ■ 활용 가능한 빅데이터 부재 ■ 데이터 남용·오용 ■ 기업 보안 및 영업 비밀 유출

정책·제도	<ul style="list-style-type: none"> ■ 빅데이터 관리·분석에 필요한 전문인력 부족 ■ 개인정보 침해 ■ 데이터 공유·활용 미흡
기술·R&D	<ul style="list-style-type: none"> ■ 데이터 분석 알고리즘 오류 ■ 선진국 대비 낮은 기술 역량 ■ M&A를 통한 빅데이터 관리·분석 기술 확보 미흡

[새로운 시도와 확장(융합) 가능성]

- (빅데이터와 타 기술과의 융합 확대) 빅데이터와 인공지능, 사물인터넷, 클라우드, 블록체인 등 타 기술과의 융합을 통한 빅데이터 플랫폼 기술 발전 및 혁신적인 새로운 시장 기회 창출 기대
- (사회문제 해결) 다양한 사회 현안 이슈 해결의 핵심 기술로 빅데이터 플랫폼 기술을 활용하며 정부 주도의 빅데이터 플래그십 투자 확대 추진

[기술개발의 핵심가치]

- (산업생태계) 타 기술과의 융합을 통한 혁신 창출 및 활용성 높은 빅데이터 확보
- (정책·제도) 이용자 보호와 빅데이터 산업 진흥의 유연한 제도 정착 및 빅데이터 전문가 육성
- (기술·R&D) 빅데이터 분석 알고리즘의 고도화 및 빅데이터 접근 및 분석 방법의 수월성 향상

[R&D에의 시사점]

- 빅데이터 관리 플랫폼
 - (초연결 지능형 기반 원천기술 개발) 초연결 지능화 추세에 대응한 빅데이터 관리, 인공지능 및 초연결 통신 결합의 핵심 기술력 강화
 - (데이터의 신뢰성 향상 기술개발) 빅데이터 분석시 활용되는 데이터의 신뢰성 제고를 위한 데이터 전처리, 정제, 학습용 데이터 셋 등 데이터 품질 향상위한 핵심기술 개발 강화



- (사물인터넷 데이터의 실시간 처리 기술 역량 강화) 사물인터넷 환경에서 대규모의 방대한 데이터가 생성됨에 따라 실시간 빠른 데이터 처리 및 분석이 중요함에 따라 이에 최적화된 기술개발 투자 강화
- (빅데이터 보안 기술개발 강화) 빅데이터 플랫폼 기술개발 단계에서 개인정보 침해 및 데이터 위변조 관련 사전 대응 가능한 빅데이터 보안 기술개발

■ 빅데이터 분석 플랫폼

- (빅데이터 분석 원천기술 확보) 신뢰성 높은 빅데이터 분석 결과 도출을 위한 빅데이터 분석 원천기술에 대한 R&D 지원 강화
- (산업별 특화된 빅데이터 분석 플랫폼 기술개발) 다양한 산업의 특성을 반영한 빅데이터 분석 플랫폼 기술개발
- (지능화된 빅데이터 분석 툴(tool) 개발) 인공지능 분석 방법론을 적용한 다양한 빅데이터 분석 방법론 기술개발 필요

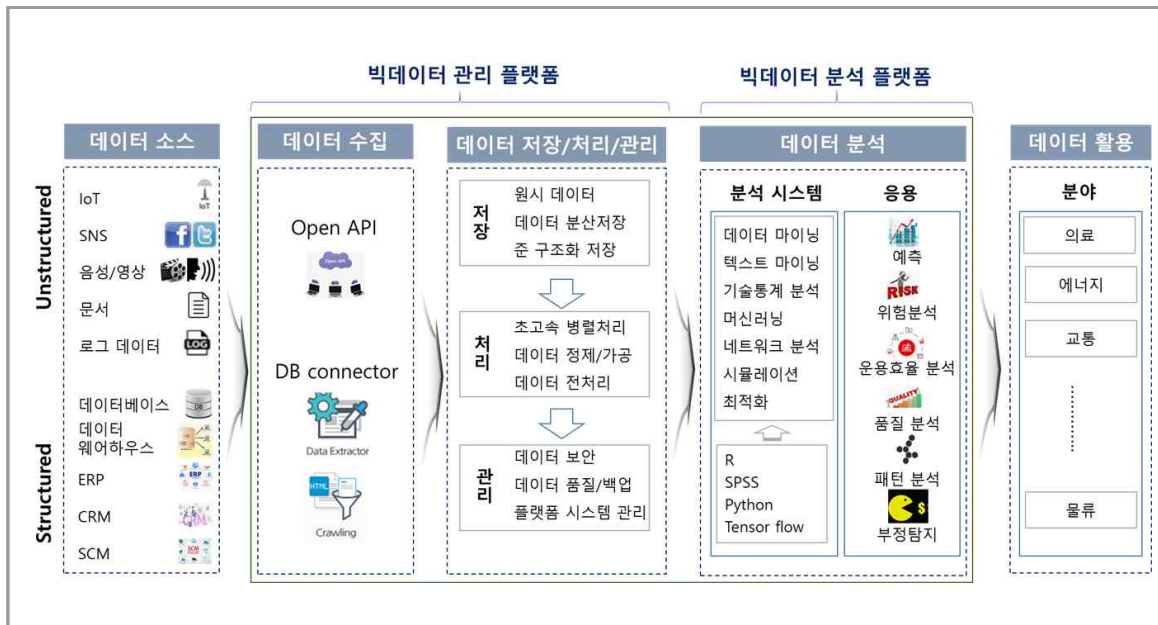
I 개요

1 정의 및 범위

가. 정의

- 빅데이터 플랫폼은 데이터의 수집·저장·처리·관리 및 분석 등의 역할 수행을 지원함으로써 새로운 인사이트(Insight)와 비즈니스 가치 창출이 가능한 데이터 프로세스 환경을 의미
 - 빅데이터 플랫폼은 데이터의 수집·저장·처리·관리를 담당하는 빅데이터 관리 플랫폼과 데이터 분석을 지원하는 빅데이터 분석 플랫폼으로 구분
 - 빅데이터 플랫폼의 핵심 역할은 원시데이터로부터 새로운 통찰력과 가치의 창출

(그림-1) 빅데이터 플랫폼





나. 범위

1 빅데이터 관리 플랫폼

- (빅데이터 수집) 다양한 데이터 소스를 Open API, DB Connector 등을 통해 데이터를 수집하는 단계로서 수집된 데이터는 크게 정형 데이터(Structured Data)와 비정형 데이터(Unstructured Data)로 구분
 - 정형 데이터: 고정된 필드에 저장된 데이터를 말하며 관계형 데이터베이스와 스프레드시트 등을 예로 들 수 있으며, 기업의 ERP(Environment Resources Management)·CRM(Customer Relationship Management)·SCM(Supply Chain Management) 데이터, 인구 통계 데이터 등이 해당
 - 비정형 데이터: 고정된 필드 형태로 항상 저장이 가능하지 않은 데이터로 페이스북, 유튜브 영상, 이미지 파일, 파일 음원 등이 해당되며 다양한 SNS 등에서 생성되는 데이터가 다수를 차지
 - 데이터 수집 기술: 크롤링(Crawling), Open API, Streaming, RDB Aggregator, ETL(Extraction, Transformation, Loading), IoT 센싱 등
 - ※ 크롤링: 소셜네트워크서비스(SNS), 뉴스, 웹 정보 등의 외부 데이터 수집 방법으로 인터넷상에서 제공되는 웹문서 정보 수집
 - ※ Open API: 인터넷 사용자가 집적 응용 프로그램과 서비스를 개발할 수 있도록 웹 운영 주체가 정보 및 데이터를 개발자와 사용자에게 공개하는 수집 기술
 - ※ Streaming: 온도, 습도 등 다양한 IoT 센서, CCTV, 음성·오디오·비디오 데이터 등 연속적으로 발생하는 데이터를 실시간으로 수집하는 기술
 - ※ Log Aggregator: 웹 로그, 트랜잭션 로그, 클릭 로그 등 다양한 로그 데이터를 수집하는 오픈 소스 기술로 Apache Flume, Apache Chukwa 등이 있음
 - ※ RDB Aggregator: 관계형 데이터베이스에서 정형 데이터를 수집하여 하둡 분산파일 시스템(HDFS)와 같은 NoSQL에 저장하는 오픈 소스 기술
- (빅데이터 저장) 다양한 형태의 데이터로부터 의미 있는 정보를 추출하기 위해 데이터의 단순 저장이 아닌 향후 사용·분석 할 수 있는 형태로 안전하고 효율적으로 저장하는 단계를 의미

- 웹·소셜 미디어 데이터, 기업 비즈니스 데이터, 사물인터넷 센싱 데이터 등 다양한 형태의 대용량비정형 데이터를 실시간으로 저장·관리할 수 있는 분산 컴퓨팅 기술 필요
- 대용량의 데이터를 파일 형태로 저장할 수 있는 기술과 비정형 데이터를 정형화된 데이터로 저장하는 기술 중요
- 데이터 저장 기술: 분산 파일 시스템(Distributed File System), 병렬 DBMS(Data Base Managed System), 네트워크 구성 저장 시스템 등
 - ※ 분산파일 시스템: 컴퓨터 네트워크로 공유하는 여러 호스트 컴퓨터 파일에 접근할 수 있는 파일 시스템으로 Hadoop 시스템, 구글 파일 시스템, 아마존 S3 파일 시스템 등이 대표적임
 - ※ 병렬 DBMS: 데이터베이스 처리를 동시에 수행하는 기술로 SAP, HANA, 오라클 Smart Flash Cache 등이 대표적임
 - ※ 네트워크 구성 저장 시스템: 서로 다른 종류의 데이터 저장장치를 하나의 데이터 서버에 연결하여 총괄적으로 데이터를 저장 및 관리하는 기술로 SAN, NAS 등이 대표적임
- (빅데이터 처리·관리) 저장된 빅데이터에서 유용한 정보 및 의미 있는 가치를 찾아내기 위한 데이터 가공 및 분석을 지원하며 데이터의 보안, 품질 관리 등을 수행
 - 수요대응 적시 데이터 처리 기술 중요: 다량의 데이터에 대한 빠른 처리 과정에서 사용자가 요구하는 정보를 적시에 처리하는 것이 중요해짐
 - 데이터 처리 및 관리 방법: 실시간 처리, 분산병렬 처리, 인-메모리 처리, 인-데이터 베이스 처리, 데이터 보안, 통합형 데이터 관리, 데이터 품질 관리 등이 있음
 - 빅데이터 전처리: 수집·저장된 방대한 양의 데이터중 분석에 필요한 데이터를 여과(Filtering), 사용하기에 적절한 형태로 변환(Transformation), 정제(Cleansing) 과정을 거쳐 통합(Integration)하여 저장
 - 주요 기술: 빅데이터 분산 병렬 및 실시간 데이터 처리를 위해 구글의 맵 리듀스(Map Reduce), 하둡(Hadoop)의 맵 리듀스, 마이크로 소프트의 Dryad 등이 있음

2 빅데이터 분석 플랫폼

- (빅데이터 분석) 텍스트 마이닝, 기계학습, 소셜 네트워크 분석 등 다양한 분석 기술로



데이터를 보다 효율적이고 정확하게 분석하여 데이터로부터 잠재된 가치 추출과 비즈니스 영역에 적용하기 위한 단계

- 방대한 양의 비정형 데이터에서 정확한 정보 추출을 위해서는 빅데이터 분석 기술역량의 확보가 매우 중요함
- 대용량 빅데이터 처리에 적합하도록 기존 분석 기술의 알고리즘을 개선하여 적용하고 있으며, 최근에는 인공지능 기술을 이용한 분석 및 예측 기술이 크게 발전하고 있음
- 분석 기술: 데이터·텍스트 마이닝, 머신러닝, 네트워크 분석, 군집화, 최적화, 감성 분석 등이 있으며 분석 툴로는 R, SPSS, Python, Tensor Flow 등이 있음

2 기술 분류 및 대상 시장

가. 기술 분류

- 의미 있는 정보를 추출하고 새로운 비즈니스 가치 창출의 기반 인프라 환경을 지원하는 빅데이터 플랫폼은 아래의 <표-1>과 같이 데이터의 수집·저장·처리·관리 및 분석 기술로 분류가 가능하며, 세부 핵심기술로 구성됨

< 표-1 > 빅데이터 플랫폼 분야 기술 범위

분야	기술분류	핵심기술
빅데이터 관리	데이터 수집 기술	빅데이터 수집·정제·융합·가공 기술 센서·머신 데이터 수집 및 정제 기술 웹·소셜 데이터 수집 및 정제 기술 실시간 데이터 수집 기술
	데이터 처리·저장·관리 기술	대규모 데이터 분산 저장 기술 병렬 DBMS 네트워크 구성 저장 시스템 기술 대규모 데이터 분산 처리 기술 대용량 데이터의 실시간 처리 기술 운영·분석 통합형 데이터 및 품질 관리 기술 고속 빅데이터 처리 및 관리 기술
빅데이터	빅데이터 분석 기술	데이터·텍스트 마이닝

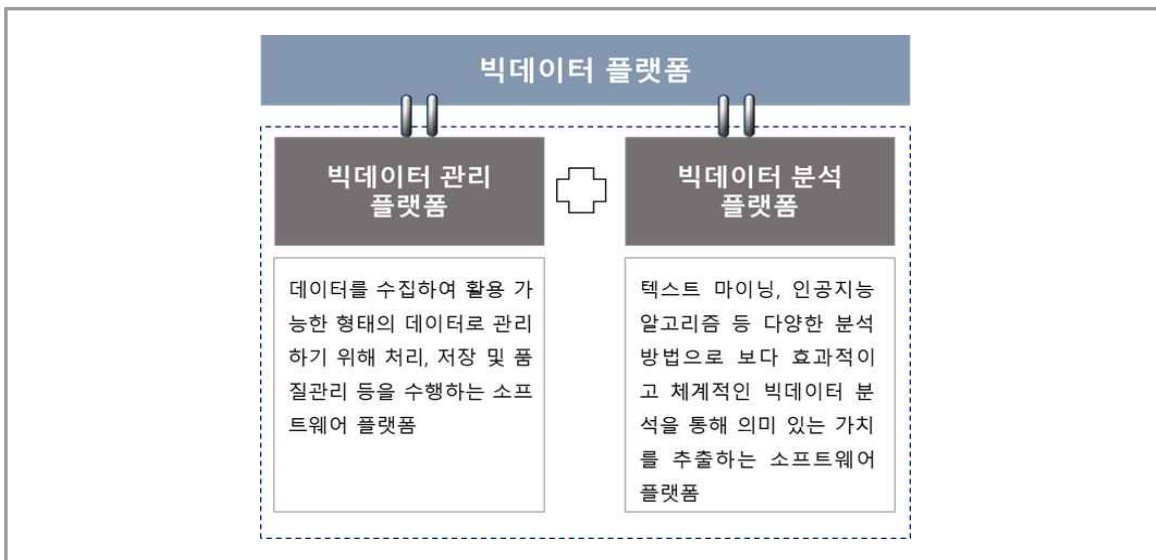
분석	대규모 기계학습 네트워크 분석 모델링 및 시뮬레이션
----	------------------------------------

* 출처 : KISTEP(2018), 중소기업기술정보진흥원(2018, 2019) 참조하여 구성

나. 대상 시장

- 빅데이터 플랫폼 시장은 크게 관리 영역과 분석 영역으로 구분 가능
 - 빅데이터 관리 영역: 데이터의 수집에서부터 처리·저장·품질관리를 위한 소프트웨어 영역
 - 빅데이터 분석 영역: 데이터 분석을 위한 소프트웨어 영역
- ※ 교육, 의료, 컨설팅, 물류 등 빅데이터를 활용한 서비스 제공 분야는 본 빅데이터 플랫폼 대상 시장에 미포함

(그림-2) 빅데이터 플랫폼 대상 시장



II 시장 및 산업생태계 현황

1 시장 동향 및 전망

가. 니즈 변화

- (데이터 증가) 사물인터넷, 스마트 디바이스, 소셜 미디어 서비스 등의 활성화로 비정형 데이터 규모가 빠르게 증가
 - 스마트폰, 태블릿 PC 등 스마트 기기의 빠른 보급, 소셜 미디어 이용 증대, 사물인터넷의 확산 등은 다양한 대용량 콘텐츠의 증가와 함께 데이터 사용 및 생산을 촉진
 - 글로벌 시장조사업체 IDC에 의하면¹⁾, 전 세계에서 생성되는 데이터양이 '18년 33ZB에서 연평균 61% 성장을 통해 '25년 175ZB로 급격히 증가할 것으로 전망
 - ※ 1ZB: 3MB 용량의 MP3를 281조 5,000억 개 저장 가능
 - 특히, 페이스북, 유튜브 등 소셜 네트워크·미디어 서비스 등의 이용 확산으로 비정형 데이터(사진, 그림, 동영상 등)의 기하급수적 증가 전망
 - ※ 페이스북은 일일 업로드 이미지가 3억 5,000만 장, 유튜브는 1분당 400시간 이상의 대용량 영상 데이터가 업로드 되고 있으며 월 기준 글로벌 유튜브 이용자 수 19억 명 및 매일 10억 시간의 유튜브 시청으로 향후 '20년에는 전체 데이터에서 비정형데이터 비중이 80% 이상을 차지할 것으로 예측²⁾
- (데이터 경제) 데이터 중심의 새로운 경제 패러다임 도래
 - 제4차 산업혁명시대로의 진전으로 IoT 센서, 소셜 네트워크·미디어 서비스 등에서 발생하는 대량의 데이터가 데이터 기반 산업을 견인하는 원동력으로 작용할 전망
 - ※ 전 세계 데이터 시장규모 전망: 1,508억 달러('17)에서 2,100억 달러('20)로 증가 전망³⁾
 - 데이터가 경제적 가치 창출의 역동적인 핵심 원유로 부상

1) IDC(2018)

2) 중앙일보(2019)

3) 관계부처합동(2018) 재인용



- ※ 기업이 데이터 기반 의사결정시 5~6%의 생산성이 향상되며, 1%의 효율성 향상이 전 세계 GDP에 약 15조 달러를 증대시킬 것으로 전망⁴⁾
- 전통적인 경제활동의 차별화 포인트로 토지, 노동력, 자본 등의 사용가능성 외에 데이터가 기업의 새로운 경쟁원천 및 차별화 요소로 인식⁵⁾
- (수요 증대) 새로운 가치 창출의 추동력으로서 빅데이터 플랫폼에 대한 수요 증대
 - 데이터 분석을 통해 파생되는 비즈니스 주도권 확보가 기업의 미래 성장 전략의 핵심 동력으로 작용할 전망
 - 글로벌 시장의 경쟁 속에서 타 기업대비 자사의 차별적 역량 확보를 통한 글로벌 시장 경쟁력 확보가 점점 어려워짐에 따라 고객 데이터로부터 잠재된 수요를 활용한 차별화된 기업 전략 추구를 위해 빅데이터에 대한 기업 수요 증대
 - 글로벌 ICT 기업들은 데이터를 활용한 비즈니스 전략 추진
 - ※ 애플은 모바일을 활용한 고객의 소비행태, 아마존은 상품 판매 및 클라우드를 활용한 고객 데이터 활용 비즈니스 전략 추진
 - 대량의 디지털 데이터 생성과 이를 통한 빅데이터 분석 영역의 확대는 인공지능과 빅데이터 분석의 결합을 통해 빅데이터 플랫폼 역량 강화로 연결되어 혁신적이고 새로운 비즈니스 창출 기회를 증대시킬 것으로 기대

나. 성장 동인

- (데이터 관리 기술 역량 확보) 실시간 데이터 전송·처리를 위한 기술역량 강화 및 안정적인 네트워크 인프라 고도화
 - 효율적인 데이터 처리 및 실시간 데이터 전송을 위한 클라우드 서비스와 네트워크 인프라 구축은 빅데이터 플랫폼 생태계 성장의 필수 요인
 - 클라우드 서비스는 데이터의 수집·저장·관리 등에 소요되는 비용 절감을 통해 빅데이터 플랫폼 활성화에 기여
 - 양질의 대용량 데이터의 끊임 없는 실시간 전송을 위해서는 지속적인 네트워크 인프라의 고도화가 필수

4) Economist Intelligence Unit(2012), BSA(2015)

5) Economist Intelligence Unit(2012)

- 특히, 음악, 영화 등 대용량 데이터의 수집·저장·처리에 이르는 일련의 과정을 실시간 처리가 가능한 컴퓨팅 기능의 강화 중요

- (데이터 활용 생태계 성장) 사물인터넷의 빠른 확산으로 인한 빅데이터 이용 환경 확대
 - 네트워크 엣지에 연결되는 디바이스의 확대는 다양한 데이터의 빠른 수집과 분석 시장의 촉매제로 작용
 - 사물인터넷에 대한 투자 확대 및 이로 인한 시장규모 증가는 궁극적으로 빅데이터 플랫폼 시장의 증대로 이어질 것으로 전망
 - ※ 글로벌 사물인터넷 시장규모 전망⁶⁾: 약 5,550억 달러('17) → 10, 529억 달러('22) (CAGR 13.7%)

- (인공지능 연계 분석 기술 고도화) 인공지능 연계의 빅데이터 분석 기술 정확도 향상 및 분석 대상 특성에 따른 분석 알고리즘의 지속적인 진화
 - 사진·동영상·텍스트 등 빠르게 대량 생성되는 비정형 데이터를 효과적으로 관리하고 유용한 의미 있는 정보를 추출해 낼 수 있는 분석 기술 확보가 중요
 - 인공지능이 결합된 빅데이터 분석 플랫폼 개발로 정확도가 높은 분석과 사용자 편의성 향상된 서비스 제공 가능
 - ※ 다양한 통계기법 및 머신러닝 등의 빅데이터 분석을 통해 분석 결과의 정확도 향상에 기여
 - ※ 구글의 통계적 기계번역을 통한 정확도 높은 번역 서비스 제공 및 검색 데이터와 머신러닝을 결합한 인공지능 비서 서비스 등의 제공 가능
 - 딥러닝 등 기계학습의 효율을 높이기 위한 인공지능 알고리즘의 성능 향상 및 제조·금융·의료 등 다양한 분야에 특성화된 알고리즘의 진화는 소비자 기반 확대와 새로운 서비스 시장 창출을 촉진

- (정부 지원) 빅데이터 분석 결과의 활용성 제고를 위한 정부의 정책적 지원 역할 확대
 - (인력양성) 빅데이터 분석 및 활용 촉진을 위해서는 숙련된 경험과 역량을 갖춘 데이터 과학자 양성 프로그램 운영 및 지원

6) IDC(2019)



- ※ 산업에 대한 풍부한 경험과 통계학·경제학·인지과학·심리학 등 다양한 분야의 지식을 갖춘 빅데이터 전문인력 양성 프로그램 마련
- (규제완화·제도 개선) 산업적 활용을 극대화 할 수 있도록 개인정보 제도 개선 및 사회적 합의 도출 필요
 - ※ 개인정보보호법상 개인정보 범위의 합리화, 개인정보 비식별 조치의 법적 근거 명확화 등 제도 개선 논의 필요
- (개방형 데이터 환경 구축) 공공·민간 데이터의 획기적인 개방 확대를 통해 민간기업의 새로운 혁신 사업 기회 발굴 촉진
 - ※ 공공데이터의 원칙적 개방(국가안보·개인정보 제외) 등을 통해 공공부문 핵심데이터(통계, 연구, 참조표준 등) 구축·개방 확대
- (기업생태계 성장 지원) 빅데이터 전문 스타트업·강소기업의 중점 발굴 및 육성 지원

다. 시장 규모 전망

■ 국내 시장

- 국내 빅데이터 플랫폼 총 시장 규모는 '15년 1.39조 원에서 연평균 8.1%의 성장을 통해 '25년 약 2.2배 증가한 3.05조 원의 시장으로 성장할 전망
- 빅데이터 관리 플랫폼은 '15년 1.13조 원에서 연평균 7.6%의 성장을 통해 '25년 2.35조 원의 시장 규모가 될 것으로 전망되며 전체 빅데이터 플랫폼 시장의 대부분('25년 기준 약 77%)을 차지하는 매우 중요한 시장
 - ※ 빅데이터 관리 플랫폼 시장 전망은 데이터 수집, 저장, 품질, 관리 등을 대상으로 한국데이터진흥원(2018) 보고서의 2015년부터 2017년까지의 시장규모에 대한 CAGR 7.6%를 2018년 이후 시장에 적용하여 추정
- 빅데이터 분석 플랫폼은 '15년 0.26조 원에서 연평균 10.4%의 빠른 성장을 통해 '25년 0.7조 원의 시장 규모가 될 것으로 전망되며 관리 플랫폼 대비 상대적으로 높은 CAGR을 통한 시장 성장 기대
 - ※ 빅데이터 분석 플랫폼 시장 전망은 데이터 분석·플랫폼 대상으로 한국데이터진흥원(2018) 보고서의 2015년부터 2017년까지의 시장규모에 대한 CAGR 10.4%를 2018년 이후 시장에 적용하여 추정

< 표-2 > 국내 빅데이터 플랫폼 시장 규모 전망 (단위: 조 원, %)

구분	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	CAGR
빅데이터 관리 플랫폼	1.13	1.26	1.31	1.41	1.52	1.63	1.76	1.89	2.03	2.19	2.35	7.6%
빅데이터 분석 플랫폼	0.26	0.29	0.31	0.35	0.38	0.42	0.47	0.52	0.57	0.63	0.70	10.4%
합계	1.39	1.55	1.63	1.76	1.90	2.06	2.23	2.41	2.60	2.82	3.05	8.1%

* 출처 : 한국데이터진흥원(2018)

주: 2015년~2017년 데이터를 기반으로 2018년 이후는 3개년(2015년~2017년)의 CAGR을 적용하여 추정

■ 세계 시장

- 세계 빅데이터 플랫폼 총 시장 규모는 '15년 69.6조 원에서 '25년 152.5조 원의 시장
으로 확대될 전망
 - ※ 세계 시장 규모는 국내 소프트웨어가 세계 시장에서 차지하는 비중인 약 2%를 반영
하여 추정하였음
- 빅데이터 관리 플랫폼 시장은 '15년 56.7조 원에서 '25년 117.7조 원의 시장 규모가
형성될 것으로 전망
- 빅데이터 분석 플랫폼 시장은 '15년 12.9조 원에서 '25년 34.8조 원의 시장 규모가 형
성될 것으로 전망

< 표-3 > 세계 빅데이터 플랫폼 시장 규모 전망 (단위: 조 원, %)

구분	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	CAGR
빅데이터 관리 플랫폼	56.67	63.18	65.59	70.56	75.91	81.67	87.86	94.53	101.69	109.40	117.70	7.6%
빅데이터 분석 플랫폼	12.92	14.30	15.75	17.38	19.20	21.19	23.40	25.84	28.53	31.50	34.78	10.4%
합계	69.59	77.48	81.34	87.95	95.11	102.86	111.26	120.36	130.22	140.90	152.48	8.1%

* 출처 : 한국데이터진흥원(2018)

주: 세계에서 국내 소프트웨어 시장이 차지하는 비중(2%)을 반영하여 세계 시장 규모 추정
2015년~2017년 데이터를 기반으로 2018년 이후는 3개년(2015년~2017년)의 CAGR을 적용하여 추정



2 빅데이터 플랫폼 생태계 현황

가. 주요 기업 현황 및 전략

- 빅데이터 플랫폼 기업들은 대용량 데이터의 실시간 처리, 다양한 파일 시스템에 대한 접근성 강화, 클라우드와 오픈 소스 기반 빅데이터 솔루션 연계 등을 통해 기업 생태계 형성 및 산업 성장에 자원과 역량을 집중
 - 주요 기업들은 분산형 병렬처리 및 실시간 데이터 처리를 위해 하둡 등의 솔루션을 활용하고 있음
 - 아마존은 자사의 AWS(Amazon Web Service) 클라우드 서비스와 오픈 소스 빅데이터 솔루션의 연계를 통해 차별화된 플랫폼 환경 구축 추진
 - 구글과 애플은 모바일 기반 고객 데이터와 자사의 분석 플랫폼을 기반으로 빅데이터 분석 환경 마련

- 빅데이터 플랫폼 기업의 핵심 기능을 크게 빅데이터 관리와 분석으로 구분 가능
 - (빅데이터 관리 분야 플랫폼 기업) 하둡을 기반으로 데이터 수집, 분산 컴퓨팅 및 병렬 처리에 집중
 - ※ 주요 기업: (국외) Google, IBM, MS, Amazon, Oracle, SAP, Intel, Cisco, Splunk, Teradata, (국내) 넥스알, 데이터스트림즈, 선재소프트 등
 - (빅데이터 분석 분야 플랫폼 기업) 데이터 마이닝, 텍스트 마이닝, 인공지능 알고리즘 등의 분석 툴을 활용하여 가치 있는 솔루션 제공
 - ※ 주요 기업: (국외) Google, IBM, MS, Amazon, Oracle, SAP, TIBCO Spotfire, Tableau Software, (국내) 다음소프트, 그루터, 사이람, 솔트룩스, 마인즈랩 등

- Google, Amazon, MS 등 글로벌 기업들은 빅데이터 관리 및 분석 관련 자체 플랫폼을 바탕으로 자사의 기업 고객에게 토탈 솔루션을 제공

< 표-4 > 빅데이터 플랫폼 국내외 기업 현황

구분	빅데이터 관리	빅데이터 분석
주요 제품·기술	<ul style="list-style-type: none"> 하둡, HDFS, 맵리듀스, Spark, 데이터 수집, 분산 컴퓨팅, 분산 파일 시스템, NoSQL, 병렬 DBMS 	<ul style="list-style-type: none"> Mahout, R, Spotfire, Tableau 통계분석, 데이터·텍스트 마이닝, 자연어 처리, 기계학습
국외 기업	<ul style="list-style-type: none"> Google, IBM, MS, Amazon, Oracle, SAP, Intel, Cisco, Splunk, Teradata 등 	<ul style="list-style-type: none"> Google, IBM, MS, Amazon, Oracle, SAP, TIBCO Spotfire, Tableau Software 등
국내 기업	<ul style="list-style-type: none"> 넥스알, 데이터스트림즈, 선재소프트 등 	<ul style="list-style-type: none"> 다음소프트, 그루터, 사이람, 솔트룩스, 마인즈랩 등

* 출처 : 안춘모(2017), 중소기업기술정보진흥원(2018, 2019) 참조하여 구성

< 표-5 > 국외 빅데이터 플랫폼 기업 현황

기업	내용
Google	<ul style="list-style-type: none"> 플랫폼: 구글 클라우드 빅데이터 저장 및 처리를 위한 맵리듀스와 Big Table 원천기술 보유
Amazon	<ul style="list-style-type: none"> 플랫폼: AWS 고객의 디지털 경험 개선을 위해 AWS를 통해 클릭스트림을 수집 및 처리 새로운 데이터에 대한 변환 작업(필터링, 정렬, 조인, 집계 등) 수행
MS	<ul style="list-style-type: none"> 플랫폼: Azure 모바일, 스토리지, 데이터 관리, 메시징 등 다양한 서비스 제공
IBM	<ul style="list-style-type: none"> 플랫폼: InfoSphere Information Server IBM BigSheets를 포함한 성능, 신뢰성, 보안 및 관리 기능으로 하둡 강화 InfoSphere Information Server: 포괄적 데이터 통합 및 품질 기능으로 신뢰성 있는 정보 제공
Oracle	<ul style="list-style-type: none"> 플랫폼: 오라클 클라우드 데이터 수집·저장·처리 과정의 간소화 대용량 빅데이터 병목 현상 제거 하둡, 스파트, NoSQL과 같은 데이터 서비스들과의 통합 절차 개선
SAP	<ul style="list-style-type: none"> 플랫폼: HANA 클라우드 플랫폼, SAT 클라우드 플랫폼 혁신적이고 다양한 사물인터넷 환경 적용 인 메모리 컴퓨팅 플랫폼 제공



빅데이터 플랫폼 분석 기업	Intel	<ul style="list-style-type: none"> ■ 플랫폼: Xeon 스케일러블 프로세서 ■ 실시간 애널리틱스, 가상화된 인프라, 고성능 컴퓨팅 연산 기능 ■ 클라우드-빅데이터-5G 무선통신기술의 결합
	Cisco	<ul style="list-style-type: none"> ■ 플랫폼: CA(Connected analytics), CPA(Cisco Prime Analytics), CPA(Common Platform Architecture) ■ Connected analytics: 하드웨어, 소프트웨어 및 센서 데이터 기반 서비스들을 결합한 사물인터넷용 데이터 분석 플랫폼 ■ Cisco Prime Analytics: 실시간 네트워크 중심의 분석 서비스를 제공하는 확장 가능한 플랫폼으로 다양한 소스로부터 네트워크 데이터를 쉽게 수집 가능 ■ Common Platform Architecture: 하둡 기반 개방형 플랫폼으로 다양한 어플라이언스를 지원
	Splunk	<ul style="list-style-type: none"> ■ 플랫폼: 통합로그관리시스템 ■ 소규모 IT 환경 비정형 데이터 분석을 위한 전용 솔루션 기업 ■ IT에서 발생하는 로그 데이터, 실시간 이벤트 데이터 및 다양한 장비 데이터를 수집, 모니터링, 검색, 분류, 분석할 수 있는 엔진 제공
	Teradata	<ul style="list-style-type: none"> ■ 플랫폼: TAP (Teradata Analytics Platform) ■ 스파크(Spark), 텐서플로우(TensorFlow), 글루온(Gluon), 티아노(Theano) 등의 엔진이 탑재되며, 인공지능 및 딥러닝 등의 알고리즘에 빠르게 접근 가능 ■ 상용 및 오픈소스 분석 기술 및 프로그래밍 언어에 대한 용이한 액세스 제동 ■ 파이썬(Python), R, SAS, SQL 등을 지원
	Google	<ul style="list-style-type: none"> ■ 플랫폼: 구글 클라우드 ■ 클라우드 기반의 빅데이터 분석서비스와 딥러닝 플랫폼 제공
	Amazon	<ul style="list-style-type: none"> ■ 플랫폼: AWS ■ 클라우드 컴퓨팅 서비스 제공 ■ 클라우드 서비스 AWS와 인공지능 비서 알렉스의 연계
MS	<ul style="list-style-type: none"> ■ 플랫폼: Azure ■ 클라우드 컴퓨팅 및 인프라스트럭처 서비스 	
IBM	<ul style="list-style-type: none"> ■ 플랫폼: InfoSphere BigInsights, InfoSphere Streams ■ InfoSphere BigInsights: 고급 텍스트 분석 모듈 및 데이터 탐색용 ■ InfoSphere Streams: 짧은 시간에 다양한 유형의 데이터 분석 지원 	
Oracle	<ul style="list-style-type: none"> ■ 플랫폼: 오라클 클라우드 ■ 하둡, 스파크, NoSQL과 같은 데이터 서비스들과의 통합 절차 개선 	
SAP	<ul style="list-style-type: none"> ■ 플랫폼: HANA 클라우드 플랫폼, SAT 클라우드 플랫폼 ■ 혁신적이고 다양한 사물인터넷 환경 적용 ■ 인 메모리 컴퓨팅 플랫폼 제공 	

	TIBCO Spotfire	<ul style="list-style-type: none"> ■ 플랫폼: Spotfire X, Spotfire Data System ■ Spotfire X, 자연어 검색 및 인공지능 기반 인사이트를 제공 ■ Spotfire Data System: 다양한 데이터 소스에 대한 스트리밍 분석 제공
	Tableau Software	<ul style="list-style-type: none"> ■ 플랫폼: Tableau Desktop ■ 각종 데이터 소스에 연결·분석으로 대시보드, 리포트를 생성하는 분석 플랫폼

* 출처 : 안춘모(2017), 중소기업기술정보진흥원(2018, 2019), IBM(2012), Amazon 홈페이지 등

< 표-6 > 국내 빅데이터 플랫폼 기업 현황

기업		내용
빅데이터 플랫폼 관리 기업	KT 넥스알	<ul style="list-style-type: none"> ■ 플랫폼: NDAP(NexR Data Analytic Platform) ■ 하둡 기반 기업 환경에 최적화된 성능, 신뢰성, 사용편의성 제공 ■ 실시간 발생 데이터를 수집하여 비즈니스 의사결정을 지원하는 Lean Stram 솔루션 제공
	데이터스트림즈 (DataStreams)	<ul style="list-style-type: none"> ■ 플랫폼: TeraONE ■ 하둡을 기반으로 정형·비정형 데이터, 빅데이터의 통합과 데이터 거버넌스를 종합적으로 구성한 One-Stop 데이터 처리 플랫폼 ■ 수집된 데이터의 품질 개선으로 데이터 분석의 신뢰성 향상
	선재소프트 (Sunjesoft)	<ul style="list-style-type: none"> ■ 플랫폼: Goldilocks ■ 인메모리 DB 기술을 채택하여 초고속으로 대용량 데이터 처리 서비스 제공
빅데이터 플랫폼 분석 기업	그루터 (Gruter)	<ul style="list-style-type: none"> ■ 플랫폼: BAAS(Bigdata Analytics & Application System) ■ 하둡 관련 기술을 보유한 빅데이터 솔루션 전문기업 ■ 다양한 오픈소스와 소프트웨어 스택 솔루션 제공
	사이람	<ul style="list-style-type: none"> ■ 플랫폼: 넷마이너(NetMiner) ■ 소셜 네트워크 분석 응용솔루션 및 컨설팅 제공 ■ 대규모 소셜 네트워크 및 데이터 간의 관계를 계량적으로 분석 및 패턴 파악
	솔트룩스	<ul style="list-style-type: none"> ■ 플랫폼: Truestory, STORM, O2 ■ 비정형 빅데이터 분석 전문기업 ■ Truestory는 클라우드 컴퓨팅과 인공지능 기술이 결합된 빅데이터 분석 수행
	마인즈랩 (MINDsLab)	<ul style="list-style-type: none"> ■ 플랫폼: maum.ai ■ 음성인식, 텍스트 마이닝 관련 종합 인공지능 서비스 기업 ■ 인공지능 알고리즘을 토대로 소셜 빅데이터 분석정보 및 기업 내부의 정형 데이터와 교차 분석하여 비즈니스 활동에



		필요한 정보 제공
	다음소프트	<ul style="list-style-type: none"> ■ 플랫폼: Contextual Finder ■ 소셜 네트워크 마이닝 분야의 전문 분석 수행 ■ 소셜 미디어 상의 데이터에 대한 의미 있는 정보 서칭, 정보간의 관계 및 패턴 파악, 트렌드 분석 등 서비스 제공

* 출처 : 안춘모(2017), 중소기업기술정보진흥원(2018, 2019) 등 참조하여 구성

나. 스타트업 투자 동향

- 2019년 4월 기준 전 세계 빅데이터 애널리틱스 분야에 약 3,800개의 스타트업이 존재하고, 우수한 기술력을 지닌 스마트업 대상 인수합병이 활발하게 전개되고 있음
 - 2019년 4월 1일 기준 3,821개의 빅데이터 애널리틱스 스타트업과 19,613명의 엔젤 투자자가 있으며 한 기업당 평균 투자금액은 460만 달러로 나타남⁷⁾
 - 2017년 기준 인수합병 주요 기업으로는 구글에 인수된 빅데이터 예측 분석 플랫폼 기업 Kaggle, 2016년에는 오라클에 인수된 클라우드 기반 에너지 관리 플랫폼 기업 OPwer 등이 대표적임⁸⁾

- 국내 빅데이터 생태계는 빅데이터 전문기술을 보유한 스타트업이 부족하며, 글로벌 수준의 빅데이터 기술혁신 기업도 전무한 상황
 - 미국의 CRN사(2017)에 의하면, 글로벌 100대 빅데이터 기술혁신 기업 중 국내 기업은 없는 것으로 나타남⁹⁾

- 우리나라는 빅데이터 스타트업 지원을 목적으로 인프라 지원 강화 및 맞춤형 성장 지원 등을 추진
 - 빅데이터 인프라 지원을 위한 판교 글로벌 ICT 클러스터를 추진 예정
 - 정부는 K-Global DB-Stars 사업을 통해 빅데이터 활용 스타트업을 선발, 컨설팅·투자

7) <https://angel.co/big-data-analytics>

8) Matt Turck(2017)

9) 관계부처합동(2018)



유치 등의 지원 확대

- ※ '14년부터 DB-Stars사업을 통해 성장한 67개 팀은 민간투자유치 518.6억 원, 전략적 M&A 5건, 데이터 제휴·판매 29건, 해외진출 및 국제수상 23건 등 매년 괄목할 만한 성과 달성

Ⅲ 정책 및 기술 동향

1 국내외 정책 동향

가. 빅데이터 관리

■ 미국

- '12년 “빅데이터 R&D 이니셔티브”를 발표, 빠르게 성장하는 디지털 데이터의 수집·저장·처리·관리 등의 추진계획 포함
 - ※ 국립과학재단(NSF): 대용량 데이터 저장 연구, 국립보건원(NIH): 신경 과학 관련 데이터 수집 및 접근성 개선 연구와 게놈 프로젝트를 통한 인체 유전자 데이터 공개, 에너지부(DoE): 대용량 데이터의 관리, 접근, 보존 관련 기술개발
- '16년 “빅데이터 R&D 전략계획”를 발표, 방대한 양의 데이터 처리, 공유 및 관리를 위한 전략과제 제시¹⁰⁾
 - ※ 국가 데이터 인프라 구축 강화 및 데이터 표준 공유와 데이터간 상호운영성 촉진 정책 추진

■ 유럽

- 빅데이터의 공유 및 개방을 통한 새로운 가치창출에 자원과 역량 집중
 - ※ 영국은 데이터 수집·활용을 위한 파트너십 구축에 정부지원 확대¹¹⁾
- 개인정보보호규정(GDPR, General Data Protection Regulation) 시행: 투명한 데이터 처리 방식, 데이터 삭제권 등 개인정보 보호 강화와 합법적인 데이터 유통을 동시에 추구함

■ 일본

- 빅데이터 활용의 기반 인프라 구축을 위한 ‘미래투자전략 2017’을 발표, 빅데이터에 대한 접근성 강화

10) IITP(2016)

11) 정도영, 김민창, 김재화(2018)



- ‘관민데이터 활용추진 전략회’를 중심으로 2020년까지 공공데이터 개방, 데이터 유통 기반의 정비 및 데이터 활용을 위한 종합적인 민관데이터 활용에 관한 법제 정비 추진
- 2017년 ‘개인정보의 보호에 관한 법률’ 개정으로 익명가공정보의 활용을 촉진시킴으로써 빅데이터 정보의 수집 등 데이터 접근·활용성 강화

■ 중국

- ‘빅데이터 산업발전 계획(2016~2020년)’을 발표¹²⁾, 빅데이터 기술개발 및 제조업 생산 과정에 빅데이터의 전면 도입을 강화
- 특히 빅데이터 수집·저장·처리·관리 등 핵심기술에 대한 연구개발 중점 프로젝트 추진

■ 우리나라

- 빅데이터가 4차 산업혁명 시대 산업 발전과 가치 창출의 촉매 역할을 함에 따라 ‘데이터 산업 활성화 전략’¹³⁾을 통해 데이터 구축·개방, 저장·유통 등의 세부 과제 추진
 - ※ 데이터 구축·개방: 산업적 활용도가 높은 영역별 기기·센서 등을 통해 수집·생성되는 산업별 실제 빅데이터 구축, 공공데이터의 개방으로 공공부분 핵심 데이터 오픈
 - ※ 데이터 저장·유통: 데이터의 효율적 저장 관리를 위한 클라우드 기반 데이터 관리 확대 및 민간-공공 연계의 개방형 데이터 기반 구축
- ‘4차 산업혁명 대응계획’¹⁴⁾을 통해 빅데이터의 산업·공공분야에서의 산업적 활용을 높이기 위해 빅데이터의 개방형 혁신과 인프라 구축 추진
 - ※ (데이터 인프라 구축) 인공지능 학습용·연구개발용 의료영상 데이터셋 등 보건의료 데이터 구축 및 개방 촉진
 - ※ (빅데이터 개방형 혁신) 연구데이터 공유·활용 제도화, 국가 연구데이터 플랫폼 구축, 공공·민간 분야 양질의 데이터 구축·분석을 지원하는 빅데이터 전문센터 구축·육성

12) 인천발전연구원(2018), 중소기업정보기술진흥원(2018, 2019)

13) 관계부처합동(2018)

14) 관계부처합동(2017)

나. 빅데이터 분석

■ 미국

- ‘빅데이터 R&D 이니셔티브’를 통해 빅데이터 분석기술 확보, 활용성 강화, 전문인력 양성을 중점 추진
 - ※ (기술) 다량의 데이터로부터 가치 있는 정보 창출을 위해 빅데이터 분석 기술의 최첨단화 추진
 - ※ (응용·활용) 산업의 새로운 부가가치 창출 견인 및 공공분야에서의 활용을 통한 사회문제 해결의 핵심 추동으로 빅데이터 분석, 응용, 활용 강화 및 확대
 - ※ (전문가 양성) 빅데이터 기획에서부터 분석 및 인사이트 추출에 이르는 전 과정의 숙련된 경험·노하우 및 관련 도메인 특성에 대한 지식을 보유한 빅데이터 전문인력 양성
- ‘빅데이터 R&D 전략계획’을 통해 빅데이터 분석관련 기술 방법론 개발, 인력양성 및 활용 역량 개선 강조
 - ※ 빠르게 변화하는 미래 수요에 대응한 빅데이터 분석 방법론 개발, 데이터 과학자·빅데이터 전문가 양성 및 공공분야에서의 데이터 분석 결과의 활용 역량 개선

■ 유럽

- 유럽내 통합 디지털 플랫폼인 ‘Digital European’을 통해 데이터 경제 육성 전략 방향을 제시하였으며¹⁵⁾, 특히 데이터에 대한 분석·활용을 강화함으로써 새로운 데이터 비즈니스 창출을 촉진
- 영국은 장기적인 측면에서 데이터 분석 및 활용을 활성화하기 위해 빅데이터 전문가 양성 정책 강화

■ 일본

- ‘미래전략투자 2017’의 빅데이터 활용을 위한 기반 인프라 구축 지원 정책을 통해 빅데이터 분석 및 응용을 위한 토대 마련
 - ※ 건강, 의료, 간호 데이터를 개개인이 파악할 수 있는 시스템 구축에 따른 빅데이터 분석

15) 중소기업정보기술진흥원(2018, 2019), 관계부처합동(2018)



및 활용의 기반 구축으로 개개인에 보다 의미있는 빅데이터 분석결과 도출 환경 구축
촉진

■ 중국

- ‘빅데이터 산업발전 계획(2016~2020년)’을 통해 빅데이터의 제조업 분야에서의 응용 확대, 글로벌 빅데이터 리딩 기업 육성 정책 방향 제시함으로써 빅데이터가 중국 전 산업발전의 추동으로 인식
- 빅데이터 분석·응용 관련 핵심 기술에 대한 연구개발 강화와 빅데이터 분석 상품개발·솔루션 개발 촉진 정책을 통해 2020년까지 10개 이상의 글로벌 빅데이터 선도기업과 500개의 빅데이터 응용서비스 기업 육성 추진

■ 우리나라

- 빅데이터의 산업적 가치 창출 활성화를 위한 정책 지원으로 ‘데이터 산업 활성화 전략’에 빅데이터 분석·활용 관련 세부 과제 추진¹⁶⁾
 - ※ (데이터 분석·활용) 데이터 기반 산업경쟁력 제고 및 빅데이터를 활용한 사회문제 해결 강화
 - ※ 데이터 분석을 위해 빅데이터 기반 실시간 모니터링(데이터 수집), 사전예측, 조기 대응 등을 활용
- 빅데이터 기술을 활용하여 기반 산업과 공공분야에서의 지능형 혁신 추진을 목표로 ‘4차 산업혁명 대응계획’에서 빅데이터의 응용 환경 조성 마련¹⁷⁾
 - ※ (빅데이터 분석·응용) 빅데이터 기반 스마트 시티 운영 체계 구축, 인공지능과 빅데이터 기반의 위험감지와 빅데이터 분석을 통한 지능형 의사결정 지원 체계 마련

16) 관계부처합동(2018)

17) 관계부처합동(2017)

2 기술 동향 및 전망

가. 기술진화의 특성

- (복합 데이터 처리) 단일 데이터 분석 중심에서 다양한 데이터를 통합 분석 가능한 다중 지능 발전 요구 및 이를 위한 빅데이터 플랫폼 기술로 발전 전망
 - 데이터 종류에 따른 상이한 데이터 관리방법 사용 및 각 종류별로 다른 방식으로 데이터에 접근하는 방식에서 데이터 사용자가 일관된 방법으로 여러 종류의 데이터를 동일한 방식으로 관리하고 처리할 수 있는 방식으로 진화
- (분석의 편리성 증대) 빅데이터 분석 전문가 외에 일반인도 데이터 분석을 쉽게 할 수 환경 조성 전망
 - 데이터 분석이 점차 일반화되면서 다양한 분야에서의 데이터 활용 니즈가 증가하고 있으며 비전문가 누구라도 편리하게 분석 가능한 방법론 활용 기대
- (데이터 분산 관리·분석 확대) 초연결 시대로 발전함에 따라 중앙 서버에서 모든 데이터를 관리하는 방식에서 분산된 형태의 저장 관리 기술로 발전 전망
 - 초연결 환경에서 빠르게 발생하는 다양한 형태의 실시간 데이터를 효율적으로 관리 및 분석을 위해 탈 중앙의 데이터 분산 분석 및 처리 기술 활용 기대
 - 기업이 IoT 빅데이터를 처리하는 신기술로 중앙 서버가 아닌 엣지(edge), 개별기기 등에 초점을 두는 엣지 컴퓨팅 기술의 발전
- (빅데이터 분석 알고리즘의 지속적인 진화) 기계학습, 딥러닝 등 인공지능 알고리즘의 진화로 보다 효율적이고 정확한 빅데이터 분석 기술 발전 전망
 - 여러 분야에서 데이터 특성을 반영한 다양한 빅데이터 분석을 위해 빅데이터 분석 알고리즘의 지속적인 진화·발전 전망



- (분석 기법의 다양화) 최근 유튜브, 소셜네트워크서비스 등 비정형데이터의 빠른 증가로 인해 텍스트·영상 기반 데이터 분석기법 확산
 - 전통적인 통계기법에 의한 비정형데이터 분석의 한계로 인해 비정형데이터 분석에 보다 효율적인 텍스트 마이닝, 형태소분석, 자연어 처리, 파이썬 등 분석기법 활용성 증대

나. 기술동향

- 미국이 빅데이터 분야 원천기술의 대부분을 보유하고 있으며, 빅데이터 기반 비즈니스 및 기업 분야의 다양한 활용에 집중함과 동시에 지속적인 R&D 역량 강화 추진
 - 빅데이터를 보유하고 있는 SNS 기업(구글, 페이스북, 트위터)이 대부분 미국 기업으로, 이들 글로벌 기업들이 빅데이터 수집, 저장관리, 분석, 서비스 등에 대한 기술의 주도권을 가지고 있음
- 우리나라는 정부 주도의 데이터 산업 전략 활성화에 자원과 역량을 집중하고 있으며, 빅데이터 플랫폼 고도화를 위한 기술개발 추진
 - 인공지능, 사물인터넷 등 빅데이터 활용 기술을 위한 양질의 데이터 구축개발 및 확산과 클라우드를 이용한 데이터 유통 및 거래 촉진 추구
 - 인공지능 기술을 활용할 수 있는 빅데이터 기반 산업 특화 활용을 위한 플랫폼 기술개발 추진
 - 실시간 빅데이터 분석이 가능한 트랜잭션·복합분석 일체형 데이터 엔지니어링 플랫폼 기술개발 추진
 - 분석·예측 기술에서는 사업화가 부진하며, 대기업에서도 쿼리분석 기술은 우수하나 알고리즘 분석 기술이 미흡함에 따라 이에 대한 전문 인력이 부족한 실정임
- 데이터 관리 기술
 - 데이터의 수집·저장과 밀접히 관련된 오픈소스 소프트웨어 솔루션 개발이 활발하게 진행되고 있으며, 하둡 기반의 오픈소스 플랫폼인 스파크가 주로 활용
 - 스파크는 디스크 액세스가 필요하지 않는 병렬 인-메모리 프로세싱 기능을 기반으로

결과를 빨리 제시할 수 있는 장점 지님

- 사물인터넷 기술의 발전과 더불어 클라우드 기반 빅데이터 플랫폼, 실시간 처리 기술 등의 개발과 활용이 확대
- 기존 기업들이 자체 구축한 빅데이터 처리 기술(수집, 저장, 분석, 유통, 시각화 등)이 다양한 클라우드 기반 빅데이터 서비스로 대체될 전망이며, 블록체인, 사물인터넷, 스트리밍 처리 등 차세대 기술을 융합한 새로운 형태의 빅데이터 플랫폼 서비스 등장 전망
- 실시간 분석을 위한 방안으로 실시간 컴퓨팅 시스템, 분산 스트리밍 플랫폼 등의 기술 개발 진행 중

■ 데이터 분석 기술

- 글로벌 IT 기업 중심으로 인공지능 연계 빅데이터 분석 플랫폼 개발과 이종 소스 심층 융합 분석 기술개발 진행
- 구글, MS 등은 통계기법, 머신러닝 등 기술 활용으로 빅데이터 분석의 정확도를 높이고 사용자 편의를 높이는 방향으로 기술 개발 진행
- 영상인식 기술과 스트림 분석 플랫폼 분야에 아이옴니사이언트, 보쉬, 오라클, MS, IBM 등이 핵심 역량을 보유

다. R&D 투자 동향 및 기술역량 수준

- 국내외 주요국은 4차 산업혁명의 핵심 기술인 빅데이터에 대한 투자를 확대하고 있으며, 특히 인공지능 연계 빅데이터 투자를 통해 결합 시너지 극대화 추구

< 표-7 > 국내외 주요국의 빅데이터 R&D 투자 현황

국가	주요 내용
미국	<ul style="list-style-type: none"> ■ 인공지능과 고성능 컴퓨팅 투자와 함께 빅데이터 분야에 대한 투자 증가 전망
유럽	<ul style="list-style-type: none"> ■ 빅데이터의 자원 최적화 및 정확성 높은 기술 개발 관련 투자 확대 ■ EC는 방대한 양의 빅데이터에 대한 정확한 분석과 최적화된 전산 자원 사용으로 에너지 효율성을 높일 수 있는 방향으로 기술개발 추진
일본	<ul style="list-style-type: none"> ■ 'Society 5.0 실현 플랫폼'을 위한 빅데이터 투자에 집중



	- 2018년 1,295억 엔 투자하였으며 이중 문부과학성은 인공지능·빅데이터·사물인터넷 등 통합프로젝트에 86.5억 엔 투자
우리나라	<ul style="list-style-type: none"> ■ 전체 정부 R&D 대비 빅데이터 분야 R&D 비중은 과제수 기준 1.4%~2.8%, 예산 기준 1.6%~2.3% 수준이며, 연구단계 기준으로 연구개발단계 중 응용 및 개발연구단계에 상대적으로 많은 예산 투입 - 2015년~2017년 기간 동안 연구개발단계별 과제당 평균 예산의 경우, 기초(220백만원, 198백만원, 166백만원), 응용(666백만원, 375백만원, 410백만원), 개발(531백만원, 315백만원, 383백만원)으로 나타남

* 출처 : STEPI(2017), KISTEP(2018)

< 표-8 > 국내 빅데이터 투자 현황(2015년~2017년)

구분	과제 수			예산(백만원)		
	2015년	2016년	2017년	2015년	2016년	2017년
기초연구	320	491	711	64,891	97,334	118,683
응용연구	123	196	216	81,943	73,601	88,551
개발연구	259	760	490	137,599	239,873	187,848
기타	36	104	174	17,296	32,949	31,769
소계	738	1,551	1,591	301,729	443,757	426,851

* 출처: KISTEP(2018)

- 우리나라의 빅데이터 분야 기술수준은 최고기술 보유국(미국) 대비 79.0%로 1.6년의 기술격차 발생
 - 우리나라는 일본(82.8%), 유럽(87.0%)등 선진국뿐만 아니라 중국(82.5%)에도 기술이 열위한 상황에 직면
 - 빅데이터 분석·예측 기술은 타국 대비 상대적으로 기술수준이 낮게 나타남

(그림-3) 빅데이터 기술주준 및 기술격차 현황

빅데이터	상대수준 (100%)															기술격차 (0년)					
	한국			미국			일본			중국			유럽			한국	미국	일본	중국	유럽	
	기초	응용	사업화	기초	응용	사업화	기초	응용	사업화	기초	응용	사업화	기초	응용	사업화						
소분류	빅데이터 수집기술	79.0	78.0	82.0	100	100	100	82.0	83.0	83.0	82.0	83.0	83.0	89.0	87.0	86.0	1.5	0.0	1.2	1.2	0.9
	빅데이터 저장-처리-관리기술	80.0	82.0	82.0	100	100	100	82.0	83.0	82.0	76.0	80.0	88.0	87.0	87.0	86.0	1.2	0.0	1.2	1.0	0.8
	빅데이터 분석-예측기술	78.0	75.0	75.0	100	100	100	83.0	82.0	83.0	83.0	82.0	83.0	86.0	87.0	87.0	2.0	0.0	1.3	1.3	1.0
	서비스 및 응용	78.0	77.0	79.0	100	100	100	83.0	82.0	82.0	83.0	82.0	82.0	87.0	86.0	87.0	2.0	0.0	1.4	1.4	1.0
	기타 빅데이터 기술	82.0	82.0	78.0	100	100	100	85.0	85.0	85.0	85.0	85.0	85.0	89.0	88.0	88.0	1.3	0.0	1.1	1.0	1.0
상대수준 및 기술격차	평균	79.2	78.6	79.2	100	100	100	82.7	81.9	82.8	80.9	81.9	84.6	87.3	87.0	86.6					
	중분류	79.0			100			82.8			82.5			87.0			1.6	0.0	1.2	1.2	0.9

* 출처 : IITP(2018)

- 중국의 빅데이터 기술 역량 상승 추세가 빠르게 전개
 - 우리나라 기술수준의 변화는 거의 없으나, 타국 대비 중국은 논문, 특허 등의 활발한 활동을 통해 기술수준이 2016년 대비 2017년에 8.1%포인트 상승하여 빠르게 선진국을 추격하고 있음

(그림-4) 주요 국가별 빅데이터 기술수준 격차 변동 추이



* 출처 : IITP(2018)

IV 주요 이슈 및 핵심가치

1 주요 이슈

가. 분야별 주요 이슈

- 빅데이터 플랫폼을 둘러싼 다양한 이슈들이 제기됨에 따라 이를 산업생태계, 정책·제도, 기술·R&D 등의 3개 분야로 구분, 각 분야별 핵심 이슈를 요약하면 아래 표 9와 같음

< 표-9 > 빅데이터 플랫폼 주요 이슈

분야	주요 이슈
산업생태계	<ul style="list-style-type: none"> ■ 협업의 생태계 환경 미흡 ■ 글로벌 빅데이터 플랫폼 기업 부재 ■ 활용 가능한 빅데이터 부재 ■ 데이터 남용·오용 ■ 기업 보안 및 영업 비밀 유출
정책·제도	<ul style="list-style-type: none"> ■ 빅데이터 관리·분석에 필요한 전문인력 부족 ■ 개인정보 침해 ■ 데이터 공유·활용 미흡
기술·R&D	<ul style="list-style-type: none"> ■ 데이터 분석 알고리즘 오류 ■ 선진국 대비 낮은 기술 역량 ■ M&A를 통한 빅데이터 관리·분석 기술 확보 미흡

* 빅데이터 플랫폼 관련 국내외 문헌과 관련 분야 전문가와의 브레인 스토밍을 통해 분야별 주요 이슈 도출

[산업생태계]

- (협업의 생태계 환경 미흡) 빅데이터는 IoT, AI, 클라우드 등과의 협업을 통한 통합적 접근이 필수적이나 각 영역별 개별접근에 치중
 - 빅데이터는 별개의 영역이 아닌 IoT, AI, 클라우드간 긴밀한 협력 관계: IoT가 빅데이터를 생산하고 AI는 빅데이터를 기반으로 작동하며 이런 일련의 과정이 클라우드



시스템을 통해 이뤄지고 있음

- 그러나, 학문간 높은 장벽에 의한 개별 접근은 빅데이터를 둘러싼 협업 환경 구축을 지연시켜 동반성장의 시너지 창출의 한계 직면

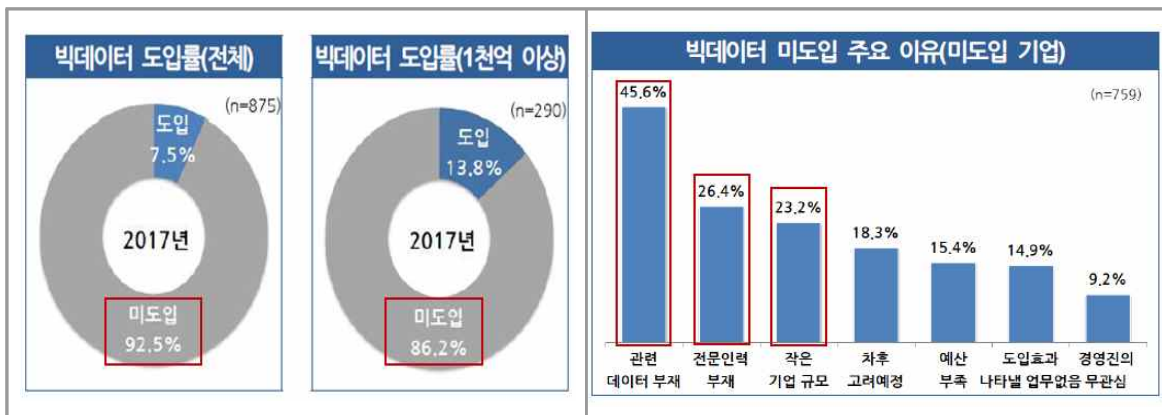
■ (글로벌 빅데이터 플랫폼 국내 기업 부족) 구글, 아마존 등 해외 주요 기업을 중심으로 글로벌 빅데이터 플랫폼 시장을 주도

- 독자적 글로벌 경쟁력을 보유한 빅데이터 플랫폼을 가진 구글, 아마존, IBM 등 글로벌 ICT 기업이 세계 시장을 리딩
- 그러나, 국내 기업은 글로벌 ICT 기업 대비 열위한 플랫폼 역량으로 국내 시장에 한정되어 사업 전개

■ (활용 가능한 빅데이터 부재) 기업의 전략적 의사결정에 활용할 만한 충분한 데이터가 없음

- 국내 기업의 빅데이터 도입률은 7.5%(전체)이며 매출액 1천억 이상 기업의 도입률은 13.8%로 나타나 전반적으로 국내 기업의 빅데이터 도입률이 저조
- 국내 기업이 빅데이터를 도입하지 않는 주된 이유로는 '관련 데이터 부재'가 45.6% 가장 높게 났으며 그 다음으로 '전문 인력 부재', '작은 기업 규모' 등으로 조사됨

(그림-5) 국내 기업의 빅데이터 활용 현황



* 출처 : 과학기술정보통신부(2018)

- (데이터 남용·오용) 빅데이터 분석의 확률적 결과에 대한 과해석 및 의미 없는 데이터간 연관성 부여로 데이터 남용·오용 발생
 - 빅데이터 분석 결과에 대해 다분히 의도성이 가미된 해석을 통해 범죄 예방 명목으로 개인의 권리 침해 가능
 - ※ 예) 영화 '마이내리티 리포트'에서 빅데이터 분석으로 미래 범죄를 사전에 감지하여 향후 죄를 지을 확률이 높은 사람을 체포
 - ※ 예) 기업이 고객 데이터 정보를 바탕으로 과도한 제품 판촉을 함으로써 고객들의 불편 가중시킴

- (기업 보안 및 영업비밀 유출 가능성 우려) 빅데이터 분석 및 활용에 있어 자체 빅데이터 분석 인력의 부재로 외부 업체 활용은 잠재적인 기업의 보안 및 영업비밀 노출 가능성 증대
 - 국내 대부분의 기업이 빅데이터 도입이 매우 낮은 상황에서 일부 도입한 기업들조차 빅데이터 분석 전문가가 부족하여 자체 분석 보다 외부 업체에 분석을 의뢰
 - 기업의 빅데이터 분석은 주로 자사의 중요한 전략적 의사결정에 활용하기 위한 것으로 자칫 외부 업체에 분석을 외뢰할 경우 기업의 중요한 영업상 비밀 내지 노하우가 유출 될 가능성 있음

[정책·제도]

- (빅데이터 관리·분석에 필요한 전문인력 부족) 데이터 분석을 통해 통찰력 있는 인사이트를 도출할 수 있는 빅데이터 전문가 부족
 - 빅데이터 분석은 복잡한 분석 과정, 의사결정에 유용한 정보생산, 숙련된 도메인 지식·경험 및 역량 등의 특성으로 인해 전문인력의 역할이 매우 중요
 - 그러나 빅데이터 기술 발전과 시장의 성장으로 빅데이터 전문가의 수요가 증가함에도 불구하고 빅데이터 전문 인력의 공급은 시장 수요를 충족하지 못하는 불균형을 초래
 - 국내 빅데이터 전문인력은 '17년 대비 '20년에 약 6,000명의 추가 인력이 필요할 전망이다. 특히 전문인력 중 빅데이터 분석가는 '17년 대비 '20년에 타 직무대비 상대적으로 높은 120.1% 증가한 약 2,400명의 신규 인력이 필요할 것으로 전망



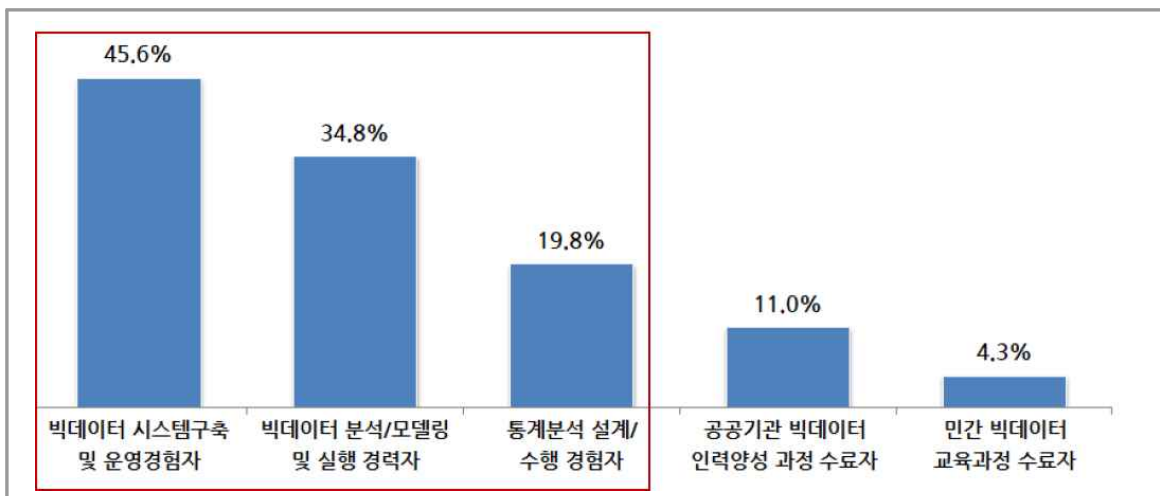
< 표-10 > 국내 직무별 빅데이터 전문인력 현황 및 전망(전체 산업) (단위: 명)

구분	2017년		2020년		'17년 대비 '20년 필요인력 증가	
	인력 수	비중(%)	인력 수	비중(%)	인력 수	비중(%)
빅데이터 개발자	2,908	29.2	5,258	32.9	2,350	80.8
빅데이터 엔지니어	1,679	16.9	2,586	16.2	907	54.0
빅데이터 분석가	1,108	11.1	2,439	15.3	1,331	120.1
데이터 사이언티스트	1,803	18.1	2,299	14.4	496	27.5
빅데이터 컨설턴트	1,724	17.3	2,233	14.0	509	29.5
빅데이터 기획/마케터	733	7.4	1,148	7.2	415	56.6
합계	9,955	100.0	15,963	100.0	6,008	60.4

* 출처 : 과학기술정보통신부(2018), 데이터산업진흥원(2018)

- 국내 빅데이터 수요기업이 선호하는 빅데이터 전문인력은 빅데이터 시스템 구축 및 운영경험자(45.6%), 빅데이터 분석·모델링 및 실행 경력자(34.8%), 통계분석 설계·수행 경험자(19.8%) 순으로 나타나, 단순 분석 업무보다 실제 관련 업무 경험이 있는 숙련된 전문가를 우선적으로 선호하는 것으로 나타남

(그림-6) 국내 빅데이터 전문인력 채용 요건(수요기업) (단위: %, 중복응답)



* 출처 : 과학기술정보통신부(2018), 데이터산업진흥원(2018)

- (개인정보 침해) 개인 데이터의 수집, 분석 및 활용이 확대되면서 개인 정보 유출로 인한 프라이버시 침해 우려가 커지고 있음
 - 스마트폰, 웨어러블 기기 및 소셜네트워크 서비스 등을 통해 생성된 방대하고 다양한 개인 데이터 기반 맞춤형 콘텐츠·서비스 및 광고 전송이 일상화되고 있음
 - 이는 통신사, 소셜네트워크 서비스 사업자, 쇼핑업체, 포털 사업자 등이 상시적으로 이용자의 온라인 활동을 모니터링하고 있음을 단적으로 보여주는 것으로 누군가에 의해 개인정보가 수집, 저장 및 분석되어 활용될 경우 언제든지 프라이버시 침해의 가능성 높음
 - 우리나라는 개인정보보호 관련하여 기본적으로 개인정보를 수집·처리·이용하고자 할 경우 이용자의 명시적인 동의를 얻는 옵토인(opt-in) 방식을 채택
 - 하지만, 엄격한 법 적용으로 인해 빅데이터 활용의 제약요인으로 작용함에 따라 이를 완화하기 위한 목적으로 '빅데이터 개인정보 가이드라인'(14년) 제정
 - ※ '빅데이터 개인정보 가이드라인'에 의하면, 개인정보 비식별화 조치를 할 경우 개인 동의 없이 데이터 활용이 가능
 - 개인정보에 대한 비식별화 조치를 취하더라도 정교해진 데이터 마이닝 알고리즘을 이용한 분석을 통해 개인 정보의 식별화가 현실적으로 가능해지는 문제 발생
 - ※ 예를 들어, 헬스케어 기업은 소셜네트워크서비스, 약국, 데이터 브로커 등으로부터 제공된 공개된 데이터의 조합을 통한 분석으로 개인의 의료정보 추론이 가능해짐
 - 결국 빅데이터는 근본적으로 개인정보 침해 문제 발생을 내포하고 있어 향후 개인정보 보호와 빅데이터 활용간 충돌을 최소화 시키며 빅데이터 산업성장을 위한 유연한 정책 개발이 요구됨

- (데이터 공유·활용 환경 미흡) 수요자 측면에서 실제 활용 가능한 공공데이터 부족
 - 공공데이터 활용률은 '17년 기준으로 12.5%로 낮은 수준이며, 공공데이터를 활용하지 않는 이유로 '필요한 데이터가 없음'이 53.4%로 조사됨

< 표-11 > 국내 공공데이터 미활용 이유 (복수응답)

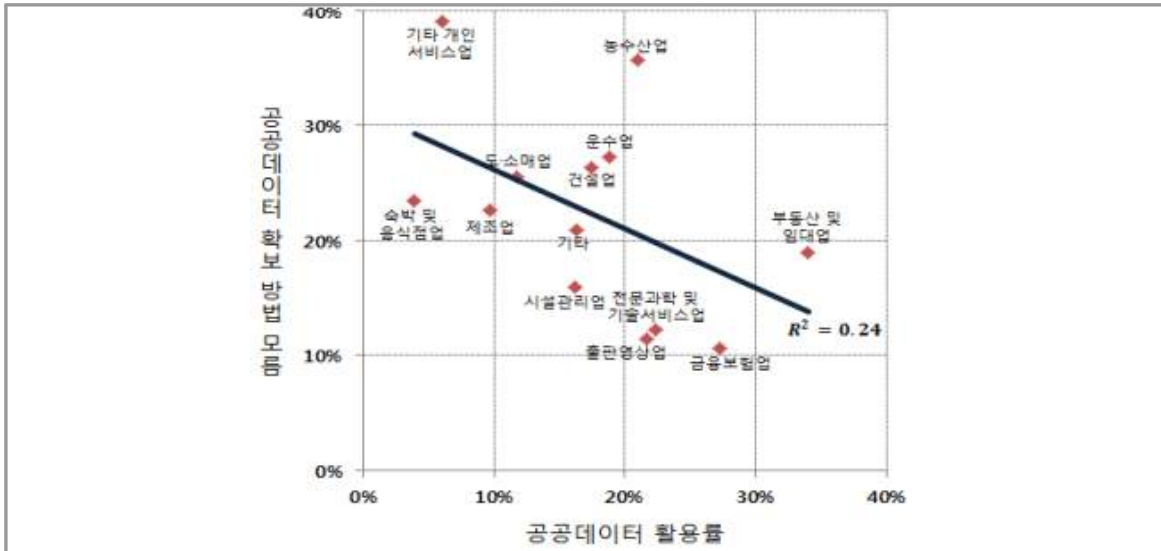
필요한 공공데이터가 없음	공공데이터 확보 방법 모름	공공데이터 인프라 부족	기타
53.4%	22.0%	2.6%	2.4%

* 출처 : 한국정보화진흥원(2018)



- 한편, 개별 기관마다 제공되는 공공데이터가 표준화되어 있지 않아 수요자 측면에서 공공데이터 활용이 제약됨
- 공공데이터 확보 방법을 모르는 비중이 클수록 공공데이터 활용률이 낮은 것으로 나타남

(그림-7) 빅데이터 활용과 미활용 이유 간 상관관계



* 출처 : 김태엽(2018), 한국정보화진흥원(2017)

[기술·R&D]

- (빅데이터 분석 알고리즘의 오류) 인공지능 기술과의 결합에 따른 인공지능 기술 자체의 오류로 인한 분석 결과의 객관성 결여 문제
 - 인공지능 SW의 잘못된 설계, 기술의 복잡도 등으로 인해 오류 발생에 따른 빅데이터 분석의 왜곡 우려
 - 검증되지 않은 빅데이터 분석은 정부, 기업 등에 심각한 손실을 초래할 수 있음
- (선진국 대비 낮은 기술 역량) 우리나라는 선진국 대비 낮은 기술수준으로 4차 산업 혁명을 선도하는 빅데이터 산업의 기술경쟁력 미흡
 - 최고 기술 보유국인 미국 대비 국내 빅데이터 수집·저장·처리·관리 기술은 약 80% 수준이며 분석·예측 기술은 70% 후반 수준으로 상당한 기술격차 존재

- 또한 중국의 기술경쟁력이 우리나라를 앞지른 상태로 글로벌 기술경쟁력 취약
- 초연결지능화, 분석 신뢰성 향상, 활용 제도 개선 등으로 빅데이터 플랫폼 기술 발전이 빠르게 전개될 예정으로 빅데이터 관련 기술역량 강화 필요
- (M&A를 활용한 빅데이터 관리·분석 기술 확보 미흡) 외부 자원을 활용하는 개방형 혁신을 통해 빅데이터 기술 확보 미흡
 - IBM, HP 등 글로벌 선진기업들은 빅데이터 시장을 선점하고 리딩하기 위해 자체 기술력 확보도 중요하지만 관련 분야 스타트업 및 벤처 기업 인수로 빅데이터 경쟁력 확보
 - ※ IBM은 비비시모, 웨더컴퍼니, 아스페라 등 빅데이터 유망기술을 보유한 업체를 인수하여 역량 강화 추진
 - 그러나, 선진국 대비 뒤쳐진 기술력 강화를 위해 선진기업의 사례처럼 빅데이터 수집·저장·분석 기술력을 갖춘 유망 벤처기업 대상 M&A·전략적 제휴를 통한 기술역량 확보는 부족한 상황임

나. 새로운 시도와 확장(융합) 가능성

- (빅데이터와 타 기술과의 융합 확대) 빅데이터와 인공지능, 사물인터넷, 클라우드, 블록체인 등 타 기술과의 융합을 통한 빅데이터 플랫폼 기술 발전 및 혁신적인 새로운 시장 기회 창출 기대
 - 양적인 측면의 빅데이터 기술 발전에서 인공지능 분석 방법론을 적용한 데이터 분석 지능의 다양화로 발전 전망
 - 초대용량 데이터의 고속 수집·처리·분석·예측 기술 및 빅데이터 기술과 4차 산업혁명 요소기술을 연계한 융합 선도기술을 통해 초연결, 지능화, 실시간성을 충족하는 새로운 비즈니스 기회 확대 전망



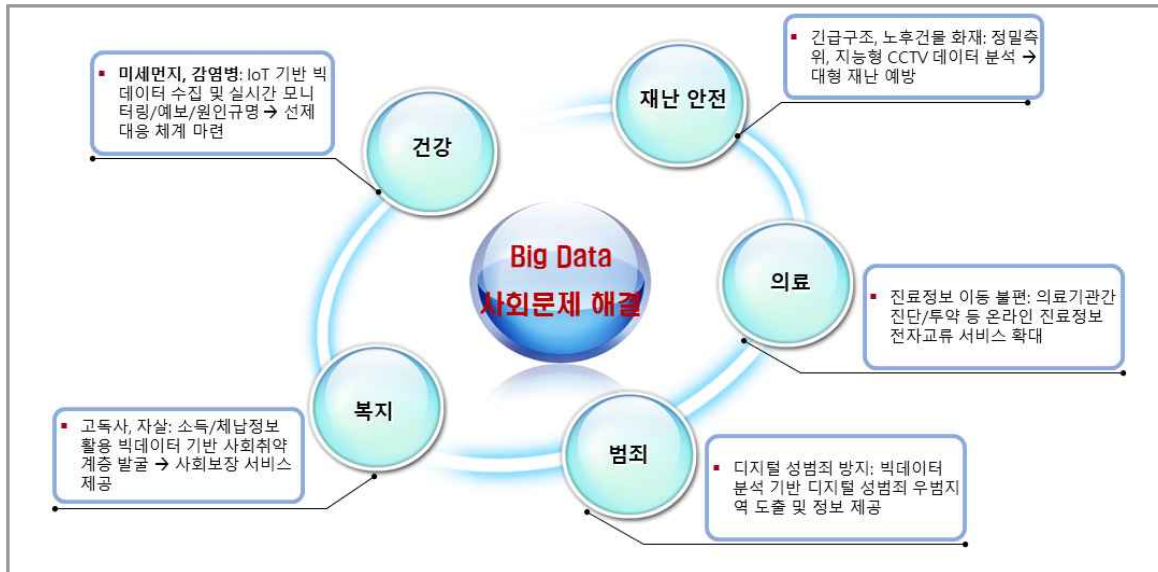
(그림-8) 빅데이터 및 타 기술과의 융합



* 출처 : 관계부처합동(2018)

- (사회문제 해결) 다양한 사회 현안 이슈 해결의 핵심 기술로 빅데이터 플랫폼 기술 활용
 - 환경, 복지, 안전, 의료 등에서 발생하는 사회문제 파악 및 해결 솔루션 제시를 위해 빅데이터 플랫폼 기술의 적극 활용
 - 정부 주도의 '공공빅데이터센터'를 설치하여 빅데이터 플랫폼 기반 사회현안 문제 관련 국가의 중요한 의사결정에 활용
 - 실시간 데이터 수집·모니터링, 데이터 분석 및 예측 등을 통해 발생(예정)된 다양한 사회문제 해결을 위한 정부 주도의 지속적인 '빅데이터 플래그십' 투자 확대 추진

(그림-9) 빅데이터 플랫폼 활용 사회문제 해결



* 출처 : 관계부처합동(2018) 참조하여 구성

2 핵심가치 및 R&D 시사점

가. 기술개발의 핵심 가치

- (산업생태계) 타 기술과의 융합을 통한 혁신 창출 및 활용성 높은 빅데이터 확보
 - 빅데이터와 인공지능을 포함한 다른 기술과의 융합 연계로 분석의 정확성 및 가치 있는 결과 도출로 새로운 혁신적인 비즈니스 기회 창출 가능
 - 기업에서 비즈니스 전략에 활용할 수 있는 데이터의 확보는 기업의 미래 비즈니스 성과에 의미있는 영향을 미침
- (정책·제도) 이용자 보호와 빅데이터 산업 진흥의 유연한 제도 정착 및 빅데이터 전문가 육성
 - (1) 개인정보 보호 관련 프라이버시 침해 대응 강력한 규제 적용과 (2) 데이터의 자유로운 사용 환경 조성을 위한 빅데이터 활용의 정책적 지원 방안 마련 등 두 마리 토끼



- 를 잡기 위한 사회적 합의 도출 필요
- 빅데이터 기술, 통계, 산업 등 다양한 분야의 통합적인 지식을 바탕으로 방대한 데이터 정보로부터 가치 있는 인사이트를 도출할 수 있는 데이터 과학자 또는 빅데이터 전문가 육성 필요
- (기술·R&D) 빅데이터 분석 알고리즘의 고도화 및 빅데이터 접근 및 분석 방법의 수월성 향상
 - 인공지능의 진화에 따른 빅데이터 분석 알고리즘의 지능화 및 다양한 데이터의 통합 분석으로 성능 향상
 - 빅데이터 분석 전문가 외 비전문가도 쉽게 데이터 수집에서부터 처리, 분석, 해석까지 가능한 기술개발

나. R&D에의 시사점

[빅데이터 관리 플랫폼]

- (초연결 지능형 기반 원천기술 개발) 초연결 지능화 추세에 대응한 빅데이터 관리, 인공지능 및 초연결 통신 결합의 핵심 기술력 강화
 - 인공지능 기술을 활용하여 실시간으로 데이터 상황에 대한 정확한 인지 및 모든 장치와 사람이 연결되는 디지털 기반 초연결 사회에서 새로운 가치를 창출하기 위한 상황 인지형·초연결 데이터 관리 기술 개발을 위한 R&D 지원 강화
- (데이터의 신뢰성 향상 기술개발) 빅데이터 분석시 활용되는 데이터의 신뢰성 제고를 위한 데이터 전처리, 정제, 학습용 데이터 셋 등 데이터 품질 향상위한 핵심기술 개발 강화
 - 방대하고 다양한 종류의 빅데이터 활용시 부정확한 데이터셋으로 인한 분석 오류 발생 가능
 - 효율적인 빅데이터 분석을 지원하기 위해 다양한 산업별 필요로 하는 데이터셋에 대한

사전수요 조사를 통해 인공지능 활용 학습용 데이터셋을 구축함으로써 산업현장에서의 빅데이터 활용성 강화

- 오류가 적고 실제 활용성이 높은 고품질의 데이터 확보를 위한 빅데이터 전처리 및 정제 기술 고도화 필요
- 금융, 통신, 에너지 등 분야별 데이터 표준체계 마련

- (사물인터넷 데이터의 실시간 처리 기술 역량 강화) 사물인터넷 환경에서 대규모의 방대한 데이터가 생성됨에 따라 실시간 빠른 데이터 처리 및 분석이 중요함에 따라 이에 최적화된 기술 개발 투자 강화

- 데이터 소스와 가장 가까운 곳에서 우선적으로 데이터를 처리·분석이 가능한 엣지컴퓨팅 기술 부각
- 인공지능 활용 데이터 배치 처리, 분석 및 스티리밍 데이터의 실시간 처리 분석을 위한 엣지컴퓨팅 기술의 R&D 투자 강화

- (빅데이터 보안 기술개발 강화) 빅데이터 플랫폼 기술개발 단계에서 개인정보 침해 및 데이터 위변조 관련 사전 대응 가능한 빅데이터 보안 기술 개발

- 빅데이터 보안 침해를 방지하기 위해 (1) 데이터 접속자의 엄격한 인증 시스템 개발 및 데이터 파손 대응 신속한 데이터 백업 기술 고도화, (2) 고도화된 개인 암호화 기술 개발을 통해 개인 데이터 유출 방지 필요

[빅데이터 분석 플랫폼]

- (빅데이터 분석 원천 기술 확보) 신뢰성 높은 빅데이터 분석 결과 도출을 위한 빅데이터 분석 원천 기술에 대한 R&D 지원 강화

- 정형화된 데이터뿐만 아니라 비정형 데이터의 증가로 실시간 의미 있는 분석을 지원하는 빅데이터 분석 기술에 대한 수요 증가
- 인공지능과 통신기술의 빠른 기술진보는 수집된 데이터의 분석 기술과의 결합을 통해 새로운 가치창출 촉진 가능



- 이에 인공지능 및 초연결 기술과의 결합을 통한 오픈 데이터 플랫폼, 엣지 분석 기술, 공간 빅데이터 분석 기술 등에 대한 원천기술력 확보 필요¹⁸⁾
- (산업별 특화된 빅데이터 분석 플랫폼 기술 개발) 다양한 산업의 특성을 반영한 빅데이터 분석 플랫폼 기술 개발
 - 산업별 빅데이터 분석·활용의 다양화: 실시간 교통 대응 지원(교통), 재난·안전 대응(재난·안전), 기업경쟁력 강화(제조), 의료 데이터 활용 기반 마련(보건의료) 등
 - 산업별 빅데이터 수요자의 구성 및 요구수준 등도 상이
 - 한편, 공공과 민간부문에서 빅데이터 분석 플랫폼 기반 서비스·제품 개발 등이 진행되고 있으나 교통, 재난·안전, 금융, 통신 등 일부 분야에서 부각, 전반적인 활용도 측면에서는 아직 미진한 상태¹⁹⁾
 - 이에 산업별 빅데이터의 활용도를 높여 이용자에게 새로운 가치 창출 제공이 가능한 맞춤형 서비스를 제공하기 위한 산업별 차별화된 빅데이터 플랫폼 분석 기술 개발 필요
- (지능화된 빅데이터 분석 툴(tool) 개발) 인공지능 분석 방법론을 적용한 다양한 빅데이터 분석 방법론 기술 개발 필요
 - 인공지능 및 초연결 기술의 발전과 병행하여 실시간성과 초연결 지능화를 달성하는 방향으로 기술이 발전할 것으로 전망
 - 증가하는 데이터, 빅데이터 분석에 대한 수요 증대, 보다 의미 있는 분석 결과 등에 대응하여 기존 빅데이터 분석 툴을 개선 및 확장 가능한 연구 개발이 필요한 시점

18) 관계부처합동(2018)

19) 관계부처합동(2019)



참고문헌

◆ 국내자료

- 과학기술정보통신부, 2017년 BIG DATA 시장현황 조사, 2018.
- 관계부처합동, 데이터 산업 활성화 전략, 2018.
- 관계부처합동, 데이터·AI경제 활성화 계획('19~'23년), 2019.
- 관계부처합동, 4차 산업혁명 대응계획, 2018.
- 김태엽, 공공데이터 개방정책의 현황 및 향후 과제, 2018.
- 데이터산업진흥원, 2017년 데이터산업 현황 조사, 2018.
- 인천발전연구원, 중국 빅데이터 산업 현황, 2018.
- 정도영, 김민창, 김재환, 빅데이터 정책 추진 현황과 활용도 제고방안, 국회입법조사처, 2018.
- 중소기업기술정보진흥원, 중소기업 전략기술로드맵 2018-2020; AI·빅데이터, 2018.
- 중소기업기술정보진흥원, 중소기업 전략기술로드맵 2019-2021; 빅데이터, 2019.
- 중앙일보, 1분마다 400시간 업로드되는 유튜브…구글은 어떻게 유튜브를 황금거위로 키웠나, 2019.
- 한국정보화진흥원, 2017 정보화통계집, 2017.
- 한국정보화진흥원, 2018 정보화통계집, 2018.
- 한국데이터진흥원, 2018 데이터산업 백서, 2018.
- IITP, 미국의 빅데이터 R&D 전략 계획, 해외 ICT R&D 정책동향 2016-12호, 2016.
- IITP, ICT 기술수준조사보고서, 2018.
- IITP, ICT R&D 기술로드맵 2023, 2018.



KISTEP, 빅데이터, KISTEP 기술동향브리프 2018-11호, 2018.

STEPI, 일본 “미래투자전략 2017” 대응 정책과 시사점, 동향과 이슈, 제31호, 2017.

◆ 국외자료

IBM, IBM 빅데이터 플랫폼, 2012.

IDC, Data age 2025, 2018.

IDC, Worldwide Internet of Things Forecast Update, 2018~2022, 2019.

BSA, 데이터는 왜 중요한가?-What's the Big Deal with Data?. 2015.

Economist Intelligence Unit, The Deciding Factor: Big Data & Decision Making, 2012.

Matt Turck, Firing on All Cylinders: The 2017 Big data Landscape.
Available from: <http://mattturck.com/bigdata2017>

<https://angel.co-big-data-analytics>

<https://www.amazon.com/>

◆ ETRI 내부자료

안춘모, 빅데이터 플랫폼 현황 및 이슈 분석, Insight Report 2017-33, 2017.



저자소개

김문구 ETRI 미래전략연구소 기술경제연구본부 기술경제연구그룹 책임연구원
e-mail: mkkim@etri.re.kr Tel. 042-860-1182

박종현 ETRI 미래전략연구소 기술경제연구본부 산업전략연구그룹 책임연구원
e-mail: stephanos@etri.re.kr Tel. 042-860-1081

빅데이터 플랫폼의 산업생태계 현황과 주요 이슈

발행인 한 성 수

발행처 한국전자통신연구원 미래전략연구소 기술경제연구본부

발행일 2019년 6월 28일



www.etri.re.kr

ETRI 한국전자통신연구원 미래전략연구소

34129 대전광역시 유성구 가정로 218
TEL.(042) 860-6114 FAX.(042) 860-6504

