



## 빅데이터를 비즈니스에 활용하려는 수요가 시장 성장 견인

### 중소기업 적합산업, 지식 집약적이며 노동집약적인 산업

- ▶ 중소기업 적합산업, 지식 집약적이며 노동집약적인 산업
- ▶ 빅데이터를 비즈니스에 활용하려는 수요가 시장 성장 견인
- ▶ 분석의 안정성, 확장성을 높이고, 사용자 인터페이스의 간소화 및 효율화 필요

빅데이터 분석 솔루션은 대용량의 정형 혹은 비정형 데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내어, 비즈니스 의사결정을 지원할 수 있는 인사이트를 발굴하고 예측하는 소프트웨어 또는 하드웨어를 의미한다.

빅데이터 분석 솔루션을 포함하는 빅데이터 산업은 오픈소스 중심의 소프트웨어 산업으로 중소기업에 적합한 산업이다. 또한, 하드웨어, 소프트웨어, 서비스가 모두 절묘하게 융합되어야만 성과를 낼 수 있을 정도로 고도의 지식 집약적이며 노동집약적인 산업이다.

## 빅데이터를 비즈니스에 활용하려는 수요가 시장 성장을 견인

시장조사기업인 MarketsandMarkets(2018.09)에 따르면, 전 세계 빅데이터 분석 시장은 2018년 약 53억 9,820만 달러에서 연평균 23.6% 성장하여 2023년 약 155억 5,450만 달러에 이를 것으로 전망되며, 과학기술정보통신부(2019.03)에 따르면 국내 빅데이터 분석 시장은 2014년 1,121억 원, 2015년 1,157억 원에서 2018년 1,584억 원으로 연평균 11.0% 성장한 것으로 나타났다.

데이터양의 폭발적 증가와 비즈니스 효율 향상을 위해 빅데이터를 비즈니스에 활용하려는 수요가 빅데이터 분석 솔루션 시장 성장을 견인하고 있다. 빅데이터 분석 솔루션은 산업의 투명성 증대, 소비자 니즈 발견, 트렌드 예측, 성과향상, 고객 세분화, 의사결정 지원과 대행, 비즈니스 모델 및 상품/서비스의 혁신 등을 위하여 사용된다.

## 분석의 안정성, 확장성을 높이고, 사용자 인터페이스의 간소화 및 효율화 필요

빅데이터 분석 솔루션의 핵심 기술에는 대용량 데이터 저장 및 처리 기술, 실시간 대용량 스트림 분석, 빅데이터 기반 고급 분석, 대용량 데이터 시각화 분석 및 인터페이스 기술 등이 있다. 대용량 데이터 저장 관리는 여러 대의 서버에 대용량 데이터를 나누어서 저장하고, 동시에 데이터를 읽어 처리하는 분산 시스템 기술로서, 데이터의 일관성, 사용성 등이 요구된다. 실시간 대용량 스트림 분석 기술은 방대한 규모로 연속적으로 생성되는 데이터를 생성 직후에 실시간으로 인-메모리 상태에서 감지, 처리 및 의사결정을 지원하는 기술이다. 데이터 분석 기술은 통계 분석을 기반으로 데이터 마이닝, 예측 분석 등의 기계학습 기술이 활용된다. 대용량 데이터 시각화 기술은 데이터 분석 결과를 효과적으로 전달하기 위한 기술로서 사용자 인터페이스의 간소화 및 효율화가 요구된다.

## I. 산업 생태계 분석

### 빅데이터 분석 솔루션의 정의

빅데이터 분석(Big Data Analytics)은 기존의 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 범위를 넘어서는 대용량의 정형 혹은 비정형 데이터에서 가치 있는 정보를 추출하고 의사결정 및 행동을 도출하기 위한 정보화 기술을 의미한다.

빅데이터 분석은 사회, 의료 및 보건, 금융, 교통, 재난, 국방 등 다양한 분야에서 생성되고 있는 데이터를 기반으로 하고 있으며, 데이터의 규모(Volume), 다양성(Variety), 생성속도(Velocity) 관점에서 과거의 데이터 분석에 비해 고도화된 특성이 있다.

또한 크게 대용량 데이터 저장 관리 기술(Large-Scale Data Management)과 고급 데이터 분석기술(Advanced Analytics)로 구분되며, 특정 기술이 단독으로 사용되기 보다는 복합적으로 사용된다.

빅데이터 분석 기술의 발달로 비정형 데이터의 활용성이 높아졌으며, 비정형 데이터의 분석 능력은 기업경쟁력의 핵심으로 인식되고 있다.

빅데이터 분석 솔루션은 대용량의 정형 혹은 비정형 데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내어, 비즈니스 의사결정을 지원할 수 있는 인사이트를 발굴하고 예측하는 소프트웨어 또는 하드웨어를 의미하며, 산업의 투명성 증대, 소비자 니즈 발견, 트렌드 예측, 성과향상, 고객 세분화, 의사결정 지원과 대행, 비즈니스 모델 및 상품/서비스의 혁신 등을 위하여 사용된다.

[그림 1] 빅데이터 개념도



\*출처: "혁신성장동력 시행계획", 관계부처 합동, 2018.05.28

## 빅데이터 분석과정 및 세부 기술

빅데이터 분석은 빅데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내기 위한 과정으로, 데이터수집, 저장, 처리, 분석, 표현 등의 기술요소를 순차적으로 사용한다. 또한 데이터를 수집하고 저장, 관리하는 기술을 필수적으로 포함하고 있으며, 빅데이터 분석이 완료된 결과물을 전송하고 공유하는 기술을 보편적으로 적용하고 있다.

[표 1] 빅데이터 분석 과정 및 주요 세부기술

분석과정	주요 세부기술
데이터 수집	<ul style="list-style-type: none"> <li>조직 내부와 외부의 분산된 여러 데이터 소스로부터 필요한 데이터를 검색하여 수동 또는 자동으로 수집하는 기술</li> <li>주요 세부 기술               <ul style="list-style-type: none"> <li>인터넷 데이터 수집(Crawling engine)</li> <li>응용 서비스 로그 수집</li> <li>센싱 데이터 수집</li> <li>RSS, Open API 연동 수집</li> </ul> </li> </ul>
데이터 저장	<ul style="list-style-type: none"> <li>모든 데이터를 저장하여 실시간으로 데이터를 처리하고, 처리된 데이터를 더 빠르고 쉽게 분석하도록 하여, 이를 비즈니스 의사결정에 바로 이용하는 기술</li> <li>주요 세부 기술               <ul style="list-style-type: none"> <li>병렬 데이터베이스 관리 시스템</li> <li>하둡(Hadoop), NoSQL</li> </ul> </li> </ul>
데이터 처리	<ul style="list-style-type: none"> <li>대규모 데이터의 수집·관리·유통·분석을 처리하는 일련의 기술</li> <li>주요 세부 기술               <ul style="list-style-type: none"> <li>실시간 처리, 분산 병렬 처리</li> <li>인-메모리(In-Memory), 인-데이터베이스(In-DB) 처리</li> </ul> </li> </ul>
데이터 분석	<ul style="list-style-type: none"> <li>데이터를 효율적으로 정확하게 분석하여 비즈니스 등의 영역에 적용하기 위한 기술</li> <li>주요 세부 기술               <ul style="list-style-type: none"> <li>통계 분석</li> <li>데이터 마이닝</li> <li>자연어 처리</li> <li>예측 분석</li> <li>기계 학습(Machine Learning)</li> </ul> </li> </ul>
데이터 표현	<ul style="list-style-type: none"> <li>자료를 시각적으로 묘사하는 기술</li> <li>주요 세부 기술               <ul style="list-style-type: none"> <li>데이터 편집</li> <li>정보시각화(Infographic)</li> <li>데이터 반응형 시각화</li> </ul> </li> </ul>

\*출처: 빅데이터 분석 솔루션, 한국신용정보원(2015), KED 재가공

## 빅데이터 분석 솔루션의 응용분야

빅데이터 분석 솔루션은 금융/보험, 의료, 정부 및 공공, 정보통신, 자동차, 에너지, 제조, 유통, 교통, 건설, 교육 등 거의 모든 산업 분야에서 응용 중이다.

금융/보험 산업에서는 거래데이터, 실시간 시장정보, 고객 서비스 이력, 소셜 미디어 데이터 등을 활용하여 상품 제안, 부정지급 발견 및 조사, 신용 리스크 관리 등에 빅데이터 분석이 사용된다. 고객의 니즈를 사전에 예측하고 필요시 고객 인사이

트를 상담원에게 제공하여 판매 및 고객 만족도 증대, 운영비용 절감 등 고객 지원 효율성을 향상시키는데 활용 가능하다. 금융사기, 범죄, 보안 침입 등 고비용 문제를 예방하고 완화하는데 사용가능하며, 거래 상대방의 리스크 평가를 위해 요구되는 방대한 데이터를 분석하는데 효과적으로 사용된다.

의료 산업에서는 환자에 대한 다양한 정보를 취합하여 지속가능한 헬스케어 시스템 구축 및 환자 치료 효율성 향상, 의료비용 절감 등에 빅데이터 분석이 사용된다. 의료 산업은 의료 환경의 변화로 인해 치료비용 절감, 효율적 인력관리, 1차 진료 접근성 향상 등의 요구에 직면해 있다. 당뇨병, 심부전증, 고혈압 등 만성질환에 의한 의료비용의 과중을 해결하고 예방 중심의 의료 서비스 제공을 위한 빅데이터 분석 솔루션 도입이 추진 중이다.

정보통신 산업은 타 산업에 비해 생성되는 데이터양이 비교할 수 없을 만큼 방대하며 이를 바탕으로 새로운 수익원 개발, 운영 효율성 향상, 고객경험 강화 등을 위해 빅데이터 분석을 활용 중이다. 위치 데이터, 지리 데이터, 소셜 미디어 데이터, 고객 접촉 데이터 등을 분석하여 지능형 마케팅 전략을 수립하고 시행하는데 활용한다. 네트워크상에서 생성되는 데이터를 분석하여 통화 불량 등의 성능 이슈를 신속히 해결하여 네트워크 품질을 향상시키고 운영비용을 낮추는데 사용한다.

정부 및 공공분야에서는 제공되는 다양한 서비스에서 생성되는 데이터를 바탕으로 범죄예방, 교통, 방위, 국가보안, 세입관리, 환경관리, 복지 등에 빅데이터 분석을 사용한다. 위협 및 범죄 예측과 예방을 위해 용의자, 범죄 및 사고 패턴, 자연 재해 이력 등 다양한 데이터 간의 연관성을 분석하여 고신뢰성의 비용 효율적 범죄 예방에 활용한다. 탈세 및 미납 등 부정행위 적발, 신종 사기기법 발견 등을 위해 내외부 데이터 간의 통합, 실시간 정보 공유 등에 빅데이터 분석을 활용한다.

자동차 산업에서의 빅데이터는 차량 센서 데이터, 품질보증 클레임, 여러 소스에서 추출되는 소비 심리 데이터 등이 활용되며, 드라이빙에 대한 고객의 기대 충족을 목표로 빅데이터 분석이 사용된다. 네트워크로 연결된 차량들의 정보와 특정 시점의 운행환경(날씨, 교통, 위험상황 등) 정보 등을 통합하여 차량 시스템 동작에 대한 실시간 인사이트 발굴이 가능하다. 차량 소유 고객의 소셜미디어, 접촉이력 등 다양한 정보를 통합 분석하여 고객감성 분석, 고객 이탈 알림 등에 활용한다.

에너지 산업에서 빅데이터 분석은 스마트 미터(Smart Meter), 스마트 그리드(Smart Grid) 운영 시 생성되는 데이터를 실시간으로 분석하여 발전 최적화와 설비 성능 향상에 사용된다. 센서 및 성능 데이터 등을 수집하여 부하(Load)를 받고 있는 기기에 대한 분석과 고장 발생 예측 등을 통해 상태기반 유지보수를 수행한다. 수요 반응(Demand Response) 요청에 대한 고객 행동, 소비 패턴 등을 분석하여 비정상적 에너지 소비 및 도난 여부 등을 판단하는데 활용되며, 그 외에도 유통, 여행, 교통 및 컨설팅 산업 등에서 빅데이터 분석을 도입하여 활용하고 있으며, 향후 대부분의 산업에서 활용될 것으로 전망된다.



## 빅데이터 분석 솔루션의 산업특징 및 구조

빅데이터 분석 솔루션을 포함하는 빅데이터 산업의 특징은 1) 규모의 경제, 2) 중소기업 적합 산업, 3) 기술·노동집약적 산업, 4) 경기 동행산업, 5) 라이프 사이클이 짧은 산업, 6) 진입장벽이 낮은 산업 등으로 요약할 수 있다.

[표 2] 빅데이터 산업 특징

산업 특징	세부 내용
규모의 경제	일반적인 상품의 경우 규모에 대한 수확체감의 특성을 보이지만 소프트웨어의 경우 생산이 증가할수록 소프트웨어 원본의 평균 개발비용이 감소함으로써 규모의 경제 특성을 가지는 바 사용자의 수가 생산비용 측면에서 매우 중요한 역할을 함.
중소기업 적합 산업	빅데이터 분석 솔루션 산업은 오픈소스 중심의 소프트웨어 산업으로 어플라이언스 형태의 고객 수요와 비싼 해외 기업 솔루션에 대한 대체 수요로 인해 중소기업에 적합함.
기술·노동집약적 산업	빅데이터는 하드웨어, 소프트웨어, 서비스가 모두 절묘하게 융합되어야만 성과를 낼 수 있을 정도로 고난도의 역량을 필요로 하는 고도의 지식 집약적이며 고급 노동집약적인 산업임.
경기 동행산업	일반적 패키지 SW 산업은 경기에 후행하는 특성을 보여 왔으나, 최근에는 동행하는 추세를 보임. SW에 대한 투자가 비용이라는 인식에서 경쟁력 강화의 수단으로 전환이 되는 추세임.
라이프 사이클이 짧은 산업	급속한 기술발전으로 제품의 라이프 사이클(Life Cycle)이 매우 짧으며, 새로운 시장이 지속적으로 창출되며 업그레이드를 통한 지속적인 유지 보수가 필요한 산업임.
진입장벽이 낮은 산업	비교적 대규모의 설비투자가 필요하지 않고 소자본으로 시장 진입이 가능하여 산업 내 신규 진출 기업이 지속적으로 증가하고 있어 동 산업 내 경쟁이 치열해지고 있음.

\*출처: 한국신용정보원(2016), KED 재가공

[표 3] 빅데이터 분석 솔루션 분야 산업연관구조

후방산업	빅데이터 분석 솔루션	전방산업
<ul style="list-style-type: none"> <li>인프라 산업               <ul style="list-style-type: none"> <li>- 데이터 서버</li> <li>- 데이터 저장기기</li> <li>- 네트워크</li> </ul> </li> <li>클라우드 컴퓨팅 산업               <ul style="list-style-type: none"> <li>- 컴퓨팅 인프라 대여</li> </ul> </li> <li>데이터 원천(Data Source) 산업               <ul style="list-style-type: none"> <li>- 통신 데이터</li> <li>- 미디어</li> <li>- 금융</li> <li>- 공공 기관(Public Data)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>대용량 데이터 저장관리 산업               <ul style="list-style-type: none"> <li>- 데이터 저장 시스템</li> <li>- 데이터 처리 통합 솔루션(플랫폼)</li> </ul> </li> <li>고급 데이터 분석 기술 산업               <ul style="list-style-type: none"> <li>- 데이터 분석 솔루션</li> <li>- 시각화 솔루션</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>서비스 산업               <ul style="list-style-type: none"> <li>- 인터넷 웹 서비스</li> <li>- 스마트 디바이스 어플리케이션</li> <li>- 통신, 금융(핀테크), 의료 등 응용산업</li> </ul> </li> <li>비즈니스 컨설팅 산업               <ul style="list-style-type: none"> <li>- 응용서비스 데이터 분석</li> <li>- 비즈니스 데이터 분석</li> <li>- 데이터 분석 교육</li> </ul> </li> <li>시스템 컨설팅 산업               <ul style="list-style-type: none"> <li>- 빅데이터시스템 설계/구축</li> <li>- 빅데이터시스템 유지/관리</li> </ul> </li> </ul>

\*출처: 중소기업청(2013), Wikibon(2014), KED 작성

빅데이터 분석 솔루션 산업은 인터넷 웹 서비스와 스마트 디바이스 어플리케이션 등 비교적 국가 경쟁력이 높은 분야의 지원 기술에 기반하고 있으며, 후방산업으로는 빅데이터 분석에 필요한 전통 컴퓨팅 인프라와 데이터 원천 산업 등이 있다.

따라서 빅데이터 분석 솔루션 산업의 가치사슬은 원천 데이터, 컴퓨팅 인프라, 빅데이터 저장 및 처리, 빅데이터 분석으로 구성되며, 인터넷 웹 서비스, 스마트 디바이스 어플리케이션, 통신, 금융, 의료 등 응용서비스 산업과 이를 활용하는 컨설팅 산업에서 최종적으로 빅데이터 분석을 실무에 적용하여 활용하고 있다.

## 해외시장 현황

시장조사기업인 MarketsandMarkets(2018.09)에 따르면, 전 세계 빅데이터 및 데이터 엔지니어링 서비스(Big Data and Data Engineering Services) 시장은 2018년 약 344억 7,000만 달러에서 연평균 17.6% 성장하여 2023년 약 773억 7,000만 달러에 이를 것으로 전망된다.

[표 4] 전 세계 빅데이터 및 데이터 엔지니어링 서비스 시장규모

(단위: 억 달러)

구분	2016년	2017년	2018년	2023년	CAGR (18~23)
데이터 모델링	67.2	74.8	84.4	162.8	14.0%
데이터 통합	100.6	113.2	130.0	275.3	16.2%
데이터 품질	55.3	64.3	76.2	180.1	18.8%
분석	34.4	42.7	54.0	155.5	23.6%
계	257.4	295.0	344.7	773.7	17.6%

\*출처: Big Data and Data Engineering Services Market - Global Forecast to 2023, MarketsandMarkets(2018.09) 연구개발특구진흥재단 제공

[표 5] 전 세계 빅데이터 분석 시장규모(지역별)

(단위: 백만 달러)

구분	2016년	2017년	2018년	2023년	CAGR (18~23)
북미	1,544.4	1,901.4	2,382.4	6,648.6	22.8%
유럽	836.6	1,018.0	1,262.9	3,437.8	22.2%
아시아태평양	617.8	799.3	1,054.1	3,429.2	26.6%
라틴아메리카	226.5	288.1	373.5	1,157.8	25.4%
중동 및 아프리카	216.2	262.7	325.3	881.1	22.1%
계	3,441.5	4,269.4	5,398.2	15,554.5	23.6%

\*출처: Big Data and Data Engineering Services Market - Global Forecast to 2023, MarketsandMarkets(2018.09) 연구개발특구진흥재단 제공

빅데이터 및 데이터 엔지니어링 서비스 시장은 데이터 모델링(Data Modeling), 데이터 통합(Data Integration), 데이터 품질(Data Quality), 분석(Analytics) 등

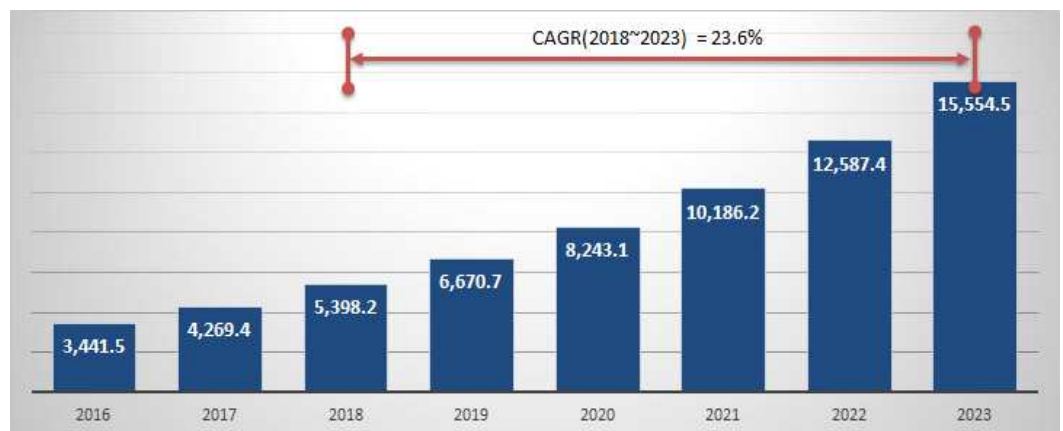
으로 구분되는데, 동 시장에서 가장 높은 비중을 차지하는 시장은 데이터 통합(Data Integration) 시장으로 2018년 기준 전체 빅데이터 및 데이터 엔지니어링 서비스 시장의 37.7%를 차지하고 있으며, 분석(Analytics) 시장은 15.7%의 비중을 보이고 있다.

동 보고서에 따르면, 전 세계 빅데이터 분석 시장은 2018년 약 53억 9,820만 달러에서 연평균 23.6% 성장하여 2023년 약 155억 5,450만 달러에 이를 것으로 전망된다.

빅데이터 분석 시장은 지역별로 가장 높은 비중을 차지하는 시장은 북미(North America) 시장으로 2018년 기준 전체 빅데이터 분석 시장의 44.1%를 차지하고 있으며, 아시아태평양(APAC) 시장은 19.5%의 비중을 보이고 있다.

[그림 2] 빅데이터 분석 세계 시장 규모

(단위: 백만 달러)



\*출처: Big Data and Data Engineering Services Market - Global Forecast to 2023, MarketsandMarkets(2018.09)(연구개발특구진흥재단 제공)

\*주: 2019~2022년은 CAGR(2018~2023) 23.6% 적용 추정

다양한 소스에서 생성된 다양한 비즈니스 데이터 양 증가, 비즈니스 데이터의 효율성 향상에 대한 필요성 증가, 정보 저장을 위한 공통 통합 플랫폼 요구 등이 빅데이터 분석 솔루션 시장 성장을 견인하고 있다.

## 국내시장 현황

과학기술정보통신부의 2018 데이터산업 현황 조사(2019.03)에 따르면, 빅데이터 분석 시장은 데이터산업 세부 시장 중 데이터솔루션 중분류별 시장에 속한다.

동 보고서에 따르면, 데이터솔루션 전체 시장은 2014년 1조 3,619억 원, 2015년 1조 4,124억 원에서 2018년 1조 7,561억 원으로 연평균 7.5% 성장한 것으로 나타났다.

데이터솔루션 시장에서 가장 높은 비중을 차지하는 시장은 DBMS 시장으로 2018년 기준 전체 데이터솔루션 시장의 36.5%를 차지하고 있으며, 데이터 분석 시장은 9.0%의 비중을 보이고 있다.



[표 6] 데이터솔루션 중분류별 시장규모

(단위: 억 원)

구분	2014년	2015년	2016년	2017년	2018년	증감률 (17~18)	CAGR (15~18)
데이터 수집	1,076	1,115	1,345	1,393	1,499	7.6%	10.4%
DBMS	5,502	5,727	6,148	6,121	6,403	4.6%	3.8%
데이터 분석	1,121	1,157	1,249	1,325	1,584	19.5%	11.0%
데이터 관리	4,531	4,699	5,367	4,628	4,887	5.6%	9.5%
데이터 보안				1,213	1,279	5.5%	
데이터 플랫폼	1,389	1,426	1,611	1,776	1,907	7.4%	10.2%
전체	13,619	14,124	15,720	16,457	17,561	6.7%	7.5%

\*출처: 2018 데이터산업 현황 조사, 과학기술정보통신부(2019.03)

국내 빅데이터 분석 시장은 2014년 1,121억 원, 2015년 1,157억 원에서 2018년 1,584억 원으로 연평균 11.0% 성장한 것으로 나타났으며, 2018년 빅데이터 분석 시장은 전년 대비로는 19.5% 성장한 1,584억 원의 시장을 형성하면서 괄목할 만한 성장세를 보였다.

최근 분석 성능의 고도화로 민간기업 수요에 적극 대응하고 있는 양상이며 범용 솔루션에서 업종별 특화 솔루션으로 변화하는 과정에서 높은 성장세의 기반을 제공하고 있다는 평가를 받고 있다.

주요 수요로는 대민 서비스 고도화를 위한 공공부문 분석 수요 증가, 대량의 데이터 분석 금융 특화 솔루션 공급, 자연어처리, 머신러닝, 언어학적 규칙을 활용한 비정형 데이터 분석 및 시각화 서비스 이용자 증가를 꼽아볼 수 있다.

[표 7] 빅데이터 분석 영역별 시장규모

(단위: 억 원)

구분		2015년	2016년	2017년	2018년	증감률 (17~18)
데이터 분석	라이선스	116 (10%)	190 (15%)	212 (16%)	238 (15%)	12.2%
	개발	846 (73%)	830 (66%)	861 (65%)	1,013 (64%)	17.7%
	유지보수	195 (17%)	229 (18%)	252 (19%)	333 (21%)	32.1%
계		1,157	1,249	1,325	1,584	19.5%

\*출처: 2018 데이터산업 현황 조사, 과학기술정보통신부(2019.03)

영역별로는 라이선스, 개발, 유지보수 등으로 구분되며, 2018년 기준 라이선스 15%, 개발 64%, 유지보수 21%의 비중을 보이고 있으며, 향후 2015~2018년과

동일한 성장률로 성장할 것으로 가정하면 국내 빅데이터 분석 시장은 2023년 2,674억 원의 시장을 형성할 것으로 전망된다.

[그림 3] 빅데이터 분석 국내 시장규모 및 전망

(단위: 억 원)



\*시장규모 산출근거: 2019~2023년은 CAGR(2015~2018) 11.0% 적용 KED 추정

[표 8] 빅데이터 분석 국내 시장규모 및 전망

(단위: 억 원)

구분	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	CAGR
계	1,121	1,157	1,249	1,325	1,584	1,759	1,953	2,169	2,408	2,674	11.0%

\*시장규모 산출근거: 2019~2023년은 CAGR(2015~2018) 11.0% 적용 KED 추정

## II. 업계 환경 분석

### 해외업체 동향

세계 빅데이터 분석 솔루션 산업은 2009년 세계 금융위기 이후 모바일 통신 환경이 보편화되고 효율적인 기업 경영활동의 중요성이 높아짐에 따라 글로벌 IT 솔루션 업체를 중심으로 성장하고 있다. 기존 데이터베이스 산업을 선도하고 있는 글로벌 IT 솔루션 업체들은 빅데이터 분석 솔루션 전문 업체들을 인수합병하여 경쟁력을 확보하고 있다. 해외의 주요 빅데이터 분석 솔루션 업체 현황은 다음과 같다.

[표 9] 빅데이터 분석 솔루션 해외 업체 현황

기업명	주요 제품	주요 특징
Teradata	AoTAs(Analytics of Things Accelerators), Aster Analytics, Teradata Everywhere, Borderless Analytics	<ul style="list-style-type: none"> <li>빅데이터 및 DW 관련 어널리틱스 솔루션과 컨설팅 서비스 제공(1979년 설립)</li> <li>아키텍처 컨설팅 등의 전략 제시</li> </ul>
GoodData	Enterprise Insights Platform-as-a-Service	<ul style="list-style-type: none"> <li>클라우드 컴퓨팅 기반의 빅데이터 어널리틱스와 BI 지향</li> <li>기업용 BI 플랫폼 서비스 시장에서 강점</li> </ul>
Datameer	DAS(Datameer Analytics Solutions)	<ul style="list-style-type: none"> <li>빅데이터 분석 및 시각화 회사</li> <li>스프레드시트 인터페이스 제공</li> <li>주요 하둡 배포판에서 사용 가능</li> </ul>
1010data	Insight Platform	<ul style="list-style-type: none"> <li>빅데이터 분석기관</li> <li>다양한 종류의 데이터 처리와 분석을 위한 플랫폼 제공</li> <li>머신러닝 등 개선된 빌트인 분석 기능 포함</li> </ul>
구글	BigQuery	<ul style="list-style-type: none"> <li>페타 바이트급의 데이터 저장 및 분석용 클라우드 서비스</li> <li>하둡, 스파크보다 상대적으로 사용이 용이</li> </ul>
Splunk	SIEM(통합로그분석시스템)	<ul style="list-style-type: none"> <li>'통합로그분석시스템(SIEM)' 시장에서 두각을 보이는 비정형 데이터 분석 솔루션 기업</li> <li>로그 데이터, 실시간 이벤트 데이터, 장비 데이터를 수집하고 다룰 수 있는 엔진 제공</li> </ul>
Mu Sigma	muOBI, muPDNA, muDSC 등	<ul style="list-style-type: none"> <li>기업들에게 전문적으로 빅데이터 분석을 의뢰 받아 의사 결정 지원 시스템을 제공</li> </ul>
Syncsort	DMX/DMEExpress, DMX-h 등	<ul style="list-style-type: none"> <li>빅데이터 통합, 처리, 정렬 SW 기업</li> <li>하둡, MS 윈도우, UNIX, Linux에 저장된 데이터 모두 지원</li> </ul>
Attivio	Active Intelligence Engine	<ul style="list-style-type: none"> <li>모든 데이터를 단일화된 인덱싱을 통해 검색, 분석이 가능한 정보 액세스 플랫폼</li> </ul>
FICO	Data Orchestrator, Decision Management Platform 등	<ul style="list-style-type: none"> <li>데이터 어널리틱스 기업</li> </ul>
Flytxt	NEON-dX	<ul style="list-style-type: none"> <li>통신서비스 사업자에게 빅데이터 분석 솔루션 공급</li> <li>인메모리 DB 사용</li> </ul>

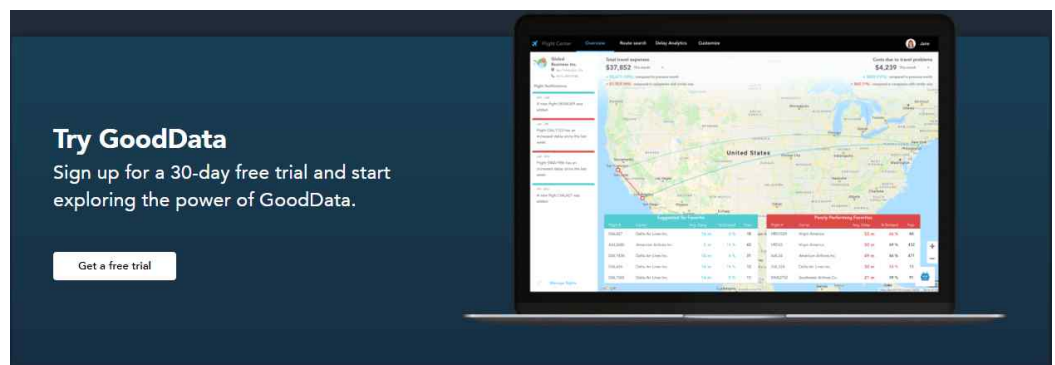
\*출처: 빅데이터 플랫폼 현황 및 이슈 분석, ETRI(2017.12) KED 재가공

[미국/Teradata]는 빅데이터 및 데이터 웨어하우스 관련 Analytic 솔루션과 컨설팅 서비스를 제공하는 글로벌 기업으로 미국 캘리포니아주 샌디에고에 본사를 두고 있다. 이 회사는 1979년 캘리포니아주 브렌트우드(Brentwood)에서 칼텍(Caltech)의 연구원과 씨티 은행(Citibank)의 첨단 기술 그룹 간의 협력으로 설립되었으며, 데이터웨어하우징 및 기업 인텔리전스 분야, DW/CRM 관련 플랫폼과 산업별 특화된 솔루션과 포괄적 전문 서비스에 역량을 보유하고 있다.

또한 2011년 애스터(Aster), Revelytix(2014), Hadapt(2014) 등의 인수합병을 진행하며 빅데이터 역량을 강화 중으로 최근에는 새로운 포트폴리오로 비즈니스 솔루션, 아키텍처 컨설팅, 기술 솔루션으로 구성된 새로운 전략을 제시하고 있고 주요 솔루션으로는 AoTAs(Analytics of Things Accelerators), Aster Analytics, Teradata Everywhere, Borderless Analytics 등을 공급하고 있다.

[미국/GoodData]는 클라우드 컴퓨팅 기반의 빅데이터 어널리틱스와 BI를 지향하는 플랫폼을 개발하는 기업으로 2007년 설립되었으며 캘리포니아주 샌프란시스코에 본사를 두고 있다. 자료의 수집, 저장, 연계, 분석, 시각화 등 비즈니스 인텔리전스(BI)의 모든 핵심 솔루션을 만들 수 있는 툴, API, 프레임워크 툴로 구성된다.

[그림 4] GoodData의 30-day free trial version

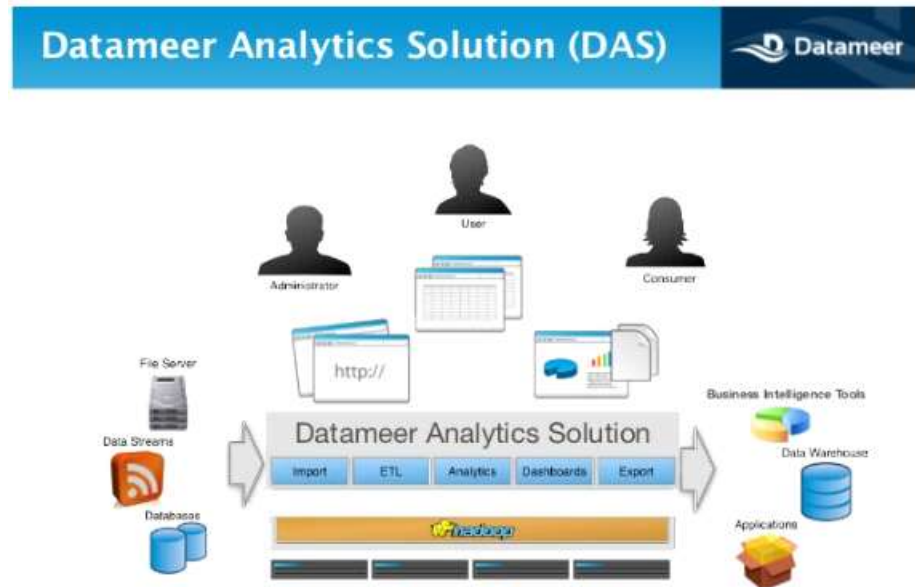


출처: <https://www.gooddata.com/>

GoodData의 모듈식 플랫폼은 높은 수준에서 데이터 수집, 준비, 변환, 분석 쿼리, 데이터 시각화 및 응용 프로그램 통합을 위한 도구, 런타임 및 스토리지를 제공한다. 이를 통해 기업은 기존 인프라와 원활하게 통합되고 직원, 고객 또는 비즈니스 파트너를 위한 분석 솔루션을 쉽게 구현할 수 있다.

[미국/Datameer]는 빅데이터 분석 및 시각화 회사로서 미국 캘리포니아주 샌프란시스코에 본사를 두고 빅데이터 어널리틱스 플랫폼을 제공하고 있다. 동사의 제품인 DAS(Datameer Analytics Solution)은 데이터 소스 통합, 200개가 넘는 분석 기능을 갖춘 스프레드시트 인터페이스가 제공되며, 아파치, 클라우데라, 맵알, 아마존 등 주요 하둡 배포판에서 사용 가능하다.

[그림 5] Datameer의 Datameer Analytics Solution(DAS)

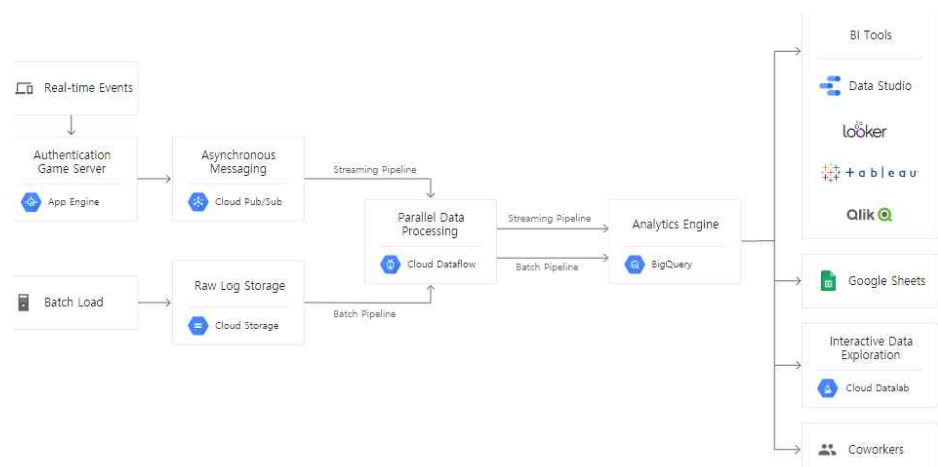


출처: <https://www.datameer.com/>

[미국/1010data]은 2000년 설립된 빅데이터 분석기관으로 미국 뉴욕에 본사를 두고 정형 및 비정형 등 다양한 종류의 데이터와 분석을 위한 툴을 하나의 플랫폼으로 제공하고 있다. 동사는 기본 비즈니스 쿼리에서 가장 정교한 데이터 과학 모델에 이르기까지 데이터에서 실행 가능한 전체 insight를 추출할 수 있는 도구를 제공하고 있다.

또한 강력한 시계열 분석, 통계 함수, 최적화, 기계 학습 등을 포함한 광범위한 함수 라이브러리를 사용하여 설명, 진단, 예측 및 규정 분석을 수행한다.

[그림 6] Google의 BigQuery 솔루션 아키텍처



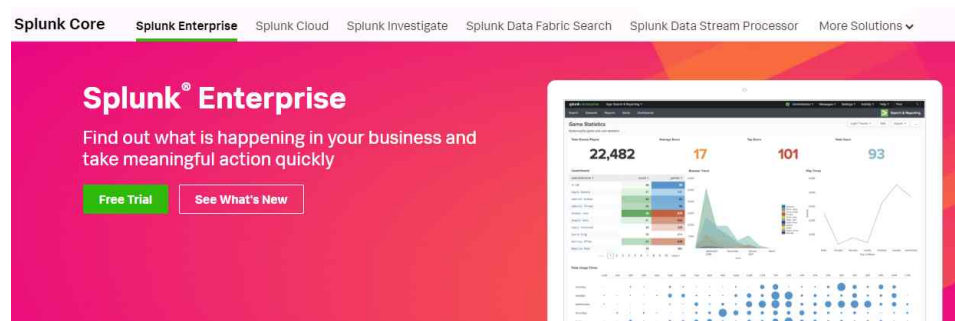
출처: <https://cloud.google.com/bigquery>



[미국/Google]는 1998년 세르게이 브린과 래리 페이지에 의해 설립된 업체로, 웹 검색, 포털 사이트, 또는 관련 사이트를 운영하고 광고를 주 사업 영역으로 하고 있다. BigQuery는 높은 확장성과 비용 효율성을 갖춘 서버리스 클라우드 데이터 웨어하우스로서 메모리 내 BI Engine과 머신러닝을 탑재하고 있다.

BigQuery의 특징은 클라우드 서비스이며(NoOps), 기존의 RDBMS에서 쓰는 SQL과 유사한 언어를 사용하고, 대용량 지원과 빠른 성능, 데이터 복제를 통한 안정성, 배치와 스트리밍 모두 지원, 저렴한 가격이다. 8,800개의 CPU와 3,600개의 디스크를 사용하는 대규모 인프라를 활용하여 1,000억 개의 레코드에 대한 질의를 30초 정도에 수행한다. Hadoop, Spark 등은 Map&Reduce 로직이나 SparkSQL 을 사용하기 위해선 일정 수준 이상의 전문성이 필요하지만, BigQuery는 상대적으로 실질적 사용과 유지보수 등이 매우 간편하다는 장점을 가지는 것으로 평가받고 있다.

[그림 7] Splunk의 Splunk® Enterprise



출처: [https://www.splunk.com/en\\_us/software/splunk-enterprise.html](https://www.splunk.com/en_us/software/splunk-enterprise.html)

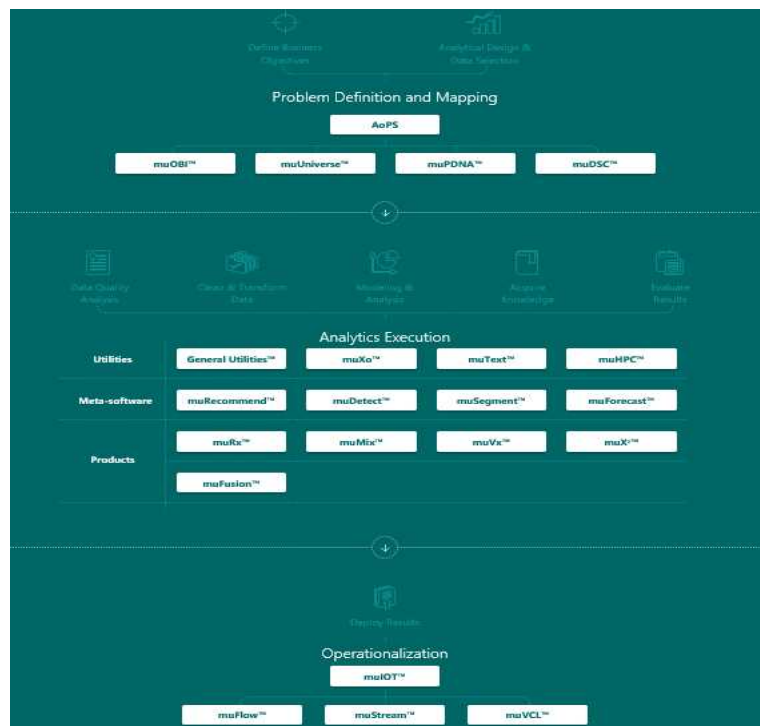
[미국/Splunk]는 IT분야에서 발생하는 로그 데이터, 실시간 이벤트 데이터 및 다양한 장비 데이터를 수집하고 모니터링하며 검색, 분류, 분석할 수 있는 엔진을 제공하는 미국의 다국적 기업으로 2004년 설립되어 미국 캘리포니아주 샌프란시스코에 본사를 두고 있다.

동사는 로그 데이터, 이벤트 데이터 등의 기계가 생성한 빅데이터를 웹 기반 인터페이스를 통해 검색, 모니터링, 분석하는 소프트웨어를 개발·공급하고 있다. 대부분의 Fortune 100대 기업을 비롯하여 전세계 100개국에 걸쳐 9,000여개 이상의 기업, 서비스 공급자 및 정부가 Splunk 소프트웨어를 사용하고 있으며, 국내에서는 인터넷 포털, 통신사, 제조사, 금융, 게임개발사 등 300여 개사에서 도입하여 사용하는 검증된 솔루션이다.

[인도/Mu Signal]는 주로 데이터 분석 서비스를 제공하는 인도 경영 컨설팅 회사로 2005년 설립되었으며, 인도 벵갈루루(Bengaluru)에 본사를 두고 있으며, 빅데이터 분석 소프트웨어, 통계 분석 소프트웨어, 빅데이터 시각화 및 리포팅 도구 등 많은 유틸리티를 개발하였다. 동사의 big data analytics software는 고객사가 문

제 정의 및 매핑에서 분석 솔루션 검색 및 운영에 이르기까지 데이터 분석 솔루션을 확장할 수 있도록 도와준다.

[그림 8] Mu Sigma의 analytics value chain



출처: <https://www.mu-sigma.com/our-platform/big-data-analytics-software>

[미국/Syncsort]는 1968년 설립된 글로벌 소프트웨어 회사로 미국 뉴욕 주 펠리버에 본사를 두고 있으며, Big Data, 초고속 정렬 제품, 데이터 통합과 서비스를 제공하는 SW 회사로서, Hadoop, MS windows, UNIX, Linux를 모두 지원하고 있다. 동사는 Hadoop, cloud 및 Splunk와 같은 차세대 분석 플랫폼을 모두 지원하며, 각 플랫폼에 특화된 솔루션을 제공하고 있다.

[미국/Attivio]는 최고의 인지 검색 및 인사이트 플랫폼 회사로 미국 매사추세츠주 보스턴에 본사를 두고 있으며, Attivio Cognitive Search and Insight Platform는 자연 언어 처리, 기계학습 및 지식 그래프와 같은 자체 학습 기술을 결합한 안전한 기업용 플랫폼으로 형식에 관계없이 엔터프라이즈 데이터 및 콘텐츠를 색인화하고 분석하며 행동 데이터를 통합하여 사용자 경험을 개인화한다.

[미국/FICO]는 선도적인 분석 소프트웨어 회사로 미국 캘리포니아주 산호세에 본사를 두고 있다. FICO Big Data Analyzer는 비즈니스 사용자, 분석가 및 데이터 과학자가 Hadoop에서 모든 유형 및 크기의 데이터를 탐색하고 분석하여 귀중한 인사이트를 얻을 수 있는 맞춤형 분석 환경이다. 데이터 프로파일링을 활용하여 한 번에 한 변수씩 데이터를 즉시 시각화한 다음 머신러닝을 적용하여 향후 결과에 대한 예측을 수행한다.

[네덜란드/Flytxt]는 네덜란드 뉴베게인(Nieuwegein)에 본사를 두고 전 세계 CSP와 모바일 업체들을 위한 빅 데이터 분석 솔루션을 제공하는 업체로 2008년 설립되었다. 동사의 NEON-dX는 분석, 인공지능, 마케팅 자동화를 클라이언트의 CRM 환경과 원활하게 통합되는 즉시 사용가능한 솔루션으로 패키지화한 것으로, 기업의 마케팅 팀이 쉽고 빠르게 측정 가능한 수익증가를 창출할 수 있도록 실시간으로 학습, 예측, 권고 및 행동한다.

## 국내업체 현황

국내 빅데이터 분석 솔루션 분야 주요 업체로는 그루터, (주)다음소프트, (주)레브웨어, (주)비아이매트릭스, (주)사이람, (주)사이버다임, (주)소프트센, (주)솔트룩스, (주)아이지에이웍스, (주)와이즈넷, (주)이씨마이너, (주)케이티넥스알(kt NexR), (주)코난테크놀로지, SK텔레콤(주) 등이 있다.

[표 10] 빅데이터 분석 솔루션 국내 업체 현황

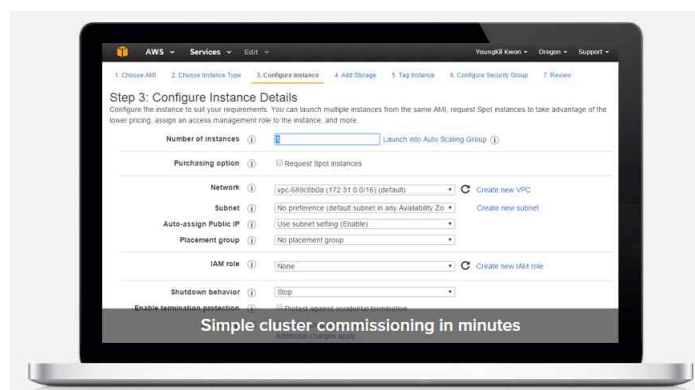
업체	사업 내용	주요 제품
그루터	빅데이터 수집/저장/분석	TaaS, Qrytica, Tajo Enterprise Gear, Cloumon
(주)다음소프트	언어처리, 텍스트 마이닝, 소셜미디어 분석, 검색솔루션	Contextual Finder™, Social Big Data Mining, Contextual IR™
(주)레브웨어	실시간 예측 분석 OLAP 솔루션	LevSPA
(주)비아이매트릭스	비즈니스 인텔리전스, 분석 리포팅, 비정형 분석 솔루션, 데이터 마이닝 솔루션	i-MATRIX, i-CANVAS, i-STREAM, i-BIG, G-MATRIX
(주)사이람	데이터 분석 솔루션	NetMiner, NetExplorer
(주)사이버다임	데이터 수집 및 가공, 데이터 분석, 시각화 솔루션	Process Analyzer
<b>(주)소프트센</b>	빅데이터 솔루션	Research Assistant, Big Cen TA
(주)솔트룩스	데이터 수집, 데이터시각화, 자연어처리 솔루션	Bigdata Suite, Tornado, DISCOVERY, TMS, BlueBolt, CAS, Rainbow
(주)아이지에이웍스	앱 광고성과 측정 및 유저 분석 솔루션	adbrix
(주)와이즈넷	데이터 수집 솔루션, 데이터 검색 솔루션, 텍스트 마이닝 솔루션, 소셜 분석 솔루션	WISE BICrawler, Search Formula-1 V5, WISE TEA V2, WISE BIC ANALYZER, BuzzInsight
(주)이씨마이너	데이터 분석 솔루션	ECMiner™, ECMinerIMS™, ECMinerAML™
(주)케이티넥스알	빅데이터 분석 솔루션, 빅데이터 플랫폼	Lean Stream, NDAP
(주)코난테크놀로지	빅데이터 분석 솔루션	Konan Analytics 4, Konan Log Analytics, pulse-K
SK텔레콤(주)	실시간 빅데이터 분석 솔루션	Metatron

\*볼드 및 밑줄 친 기업은 코스닥 기업임

\*출처: 2018 데이터산업 현황 조사, 과학기술정보통신부(2019.03), 각사 홈페이지, KED 작성

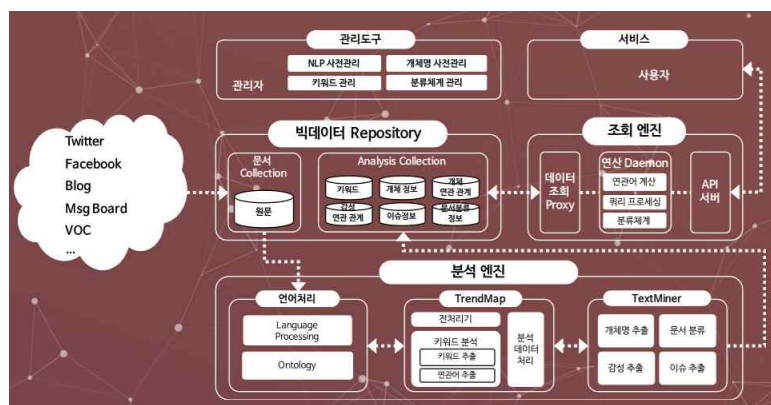
[그루터]는 빅데이터 플랫폼 및 솔루션 전문업체로, 주요 제품으로 TaaS(Tajo-as-a-Service), Qrytica, Tajo Enterprise Gear, Cloumon 등이 있다. TaaS는 하둡 에코 시스템 기반 오픈소스 빅데이터 웨어하우스 솔루션이고, Qrytica는 표준 SQL로 빅데이터 분석을 실행하는 빅데이터 분석 도구, Tajo Enterprise Gear는 빅데이터 처리를 위해 설계된 강력한 엔드-투-엔드 빅데이터 웨어하우스 플랫폼, Cloumon은 Hadoop 통합 관리 도구이다.

[그림 9] Gruter Enterprise Tajo on AWS



\*출처: <http://gruter.com/products/taas/>

[그림 10] 다음소프트의 Contextual Finder™ TECHNICAL ARCHITECTURE



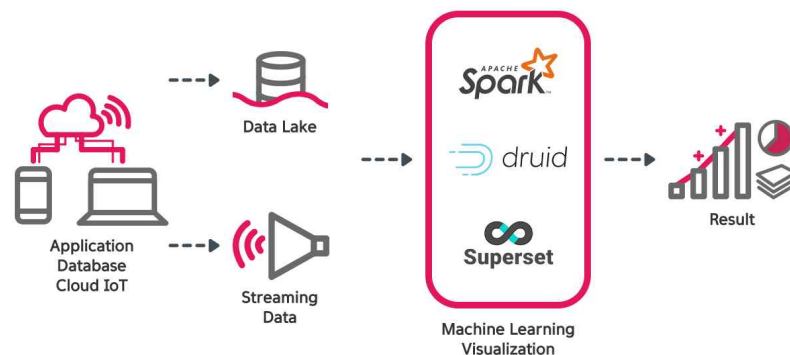
\*출처: <http://www.daumsoft.com/contextualFinder.html>

[(주)다음소프트]는 인공지능 기반 빅데이터 전문기업으로 인공지능 대화기술 및 자연어 처리 기술을 보유하고 있다. 주요 제품에는 빅데이터 마이닝 솔루션인 Contextual Finder™, 소셜 빅데이터 분석 서비스 Social Big Data Mining 등이 있다. Contextual Finder™는 자연어처리기술과 텍스트마이닝기술을 이용하여 기업 내부 및 외부에 존재하는 대량의 문서로부터 인사이트를 발굴하고 이슈를 모니터링할 수 있는 하둡기반의 빅데이터 마이닝 솔루션이고, Social Big Data Mining은 소셜미디어로부터 소비자들의 살아있는 이야기가 담긴 대량의 데이터를 수집/

분류하여 언어에 대한 이해를 기반으로 텍스트를 분석하고 의미 있는 정보를 추출해내는 분석 서비스이다.

[(주)레브웨어(Levware)]는 1999년 한국후테로시스템 상호로 설립되어, 2017년 현 사명으로 변경한 빅데이터 솔루션 전문기업으로 스마트 제조원가 솔루션 제공 및 구축 서비스, 빅데이터 기반 실시간 예측분석 솔루션 제공 및 구축 서비스 등을 주요 사업으로 영위하고 있다. 동사의 LevSPA는 빅데이터 기반의 실시간 예측 분석 OLAP 솔루션으로 과거 데이터로부터 머신러닝 모델을 개발하고, 개발된 모델을 PMML 형태로 저장/활용 가능함으로써 실시간 스트리밍 데이터를 분석/예측 가능한 특징이 있다.

[그림 11] 레브웨어(Levware)의 LevSPA 프로세스



\*출처: <http://www.levware.co.kr/toPageProducts?path=levSPA>

[그림 12] 비아이매트릭스의 통합 데이터 분석 솔루션 i-STREAM



\*출처: <http://www.bimatrix.co.kr/products/i-stream/>

[(주)비아이매트릭스]는 2005년 설립된 BI 및 빅데이터 분석 솔루션 전문업체로 BI와 Big Data 핵심 기술을 보유하고 있다. 동사의 i-CANVAS는 Look & Feel이 우수한 미려한 분석 보고서 제작에 최적화된 BI 솔루션으로 웹에서 DB에 연결하고, 보고서 생성 / 편집 / 수정 / 공유 및 배포가 가능한 강력한 보고서 개발도구를 제공하고, i-STREAM는 데이터 추출에서 분석 결과의 시각화까지 데이터



분석 전 과정을 지원하는 통합 데이터 분석 솔루션으로 분석을 위한 ETL, 이기종 데이터 소스의 통합 분석, 데이터마이닝, 통계, 예측 분석 등 다양한 기능을 수행하며 복잡한 데이터 처리 요구 사항에 빠르게 대응하고 데이터 분석가가 쉽게 사용할 수 있도록 사용자 중심의 쉬운 UI를 제공한다.

[(주)사이람]은 소셜 네트워크 분석 전문기업으로, 2000년에 창립되었다. 동사는 2001년에 세계 최초의 상용 소셜 네트워크 분석(Social Network Analysis) 소프트웨어 NetMiner를 출시하였다. NetMiner는 GUI 기반의 편리한 이용법과 유연한 데이터 분석 환경을 제공해 세계적인 명성을 확보하고 있다. 그외 제품으로는 2003년에 출시된 NetExplorer가 있다. NetExplorer는 범죄 네트워크 분석 및 공모사기 적발 솔루션으로, 급증하고 있는 보험사기 및 범죄 수사 등의 분야에서 혐의자들 간의 직·간접적 연관관계를 시각적으로 분석해 공모집단을 적발하는데 효과적으로 활용될 수 있다.

[그림 13] 사이람의 소셜 네트워크 분석 소프트웨어 NetMiner



\*출처: [http://www.cyram.com/product/product\\_01.jsp](http://www.cyram.com/product/product_01.jsp)

[(주)사이버다임]은 1998년 비정형 콘텐츠 관리의 국산화를 목표로 설립된 업체로 문서중앙화 시스템을 주력으로 문서관리, 기업용 클라우드 솔루션 등을 보유하고 있는 전문기업이다. 동사의 Process Analyzer는 그룹웨어, ERP, 제조정보시스템 등 조직의 다양한 업무 처리 시스템에 기록되는 이벤트 로그를 분석하여 의미 있는 정보를 찾아내는 제품으로, Process Analyzer를 사용하면 데이터베이스, Hadoop, 클라우드 서비스에 이르기까지 모든 데이터를 탐색할 수 있다. 기본적인

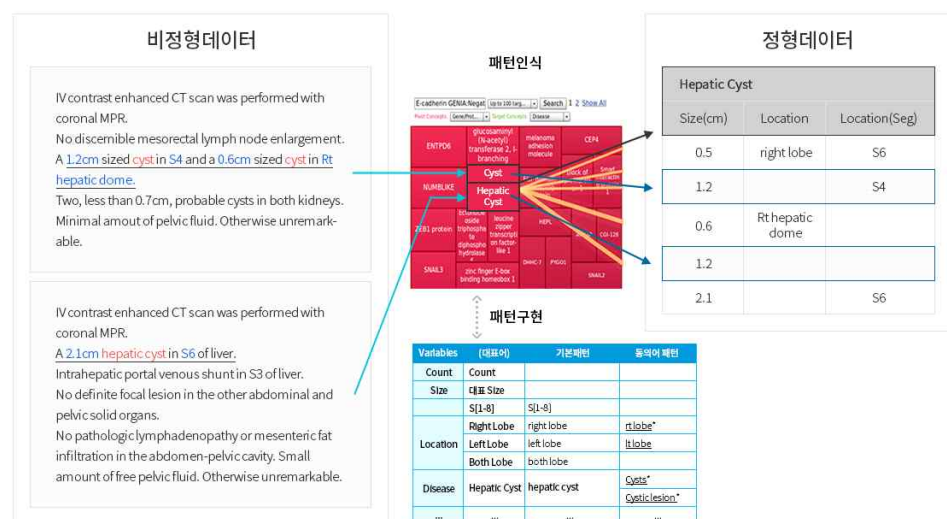
Data Dashboard부터 데이터 흐름을 표현하는 Process Map, Animation까지 다양한 시각화 도구를 제공하며, 웹 기반 프로세스 마이닝 제품으로 어디에서나 서버에 접속하여 통계 및 분석 정보를 확인할 수 있다.

[그림 14] 사이버다임의 Process Analyzer 시각화 기술



\*출처: <https://www.cyberdigm.co.kr/products/process-analyzer/>

[그림 15] 소프트센의 Big CEN TA 시스템 예시



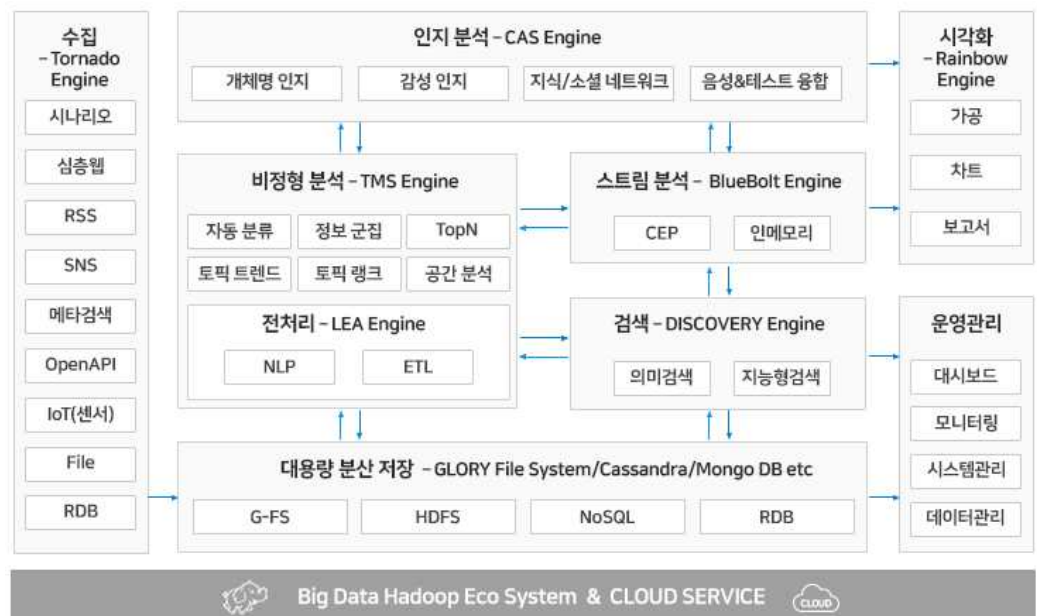
\*출처: <http://www.softcen.co.kr/app/view/business/department>

[(주)소프트센]은 1988년 4월 11일 (주)비티씨코리아로 설립되었으며, 수차례 상호변경을 거쳐 2014년 4월 현 상호로 변경한 업체로, 1997년 7월 코스닥시장에 상장되었다. 동사는 빅데이터 솔루션, 빅데이터 서비스, IT 솔루션, IT인프라 등의 사업을 영위하고 있다. 빅데이터 솔루션 관련 주요 제품은 Research Assistant, Big Cen TA 등이다. Research Assistant는 임상연구 방법론 절차에 따른 사용

자 중심의 화면 구성과 복합 검색(정형조건 검색+키워드 검색)지원, 동의어/유사어 사전 제공, 다차원 분석정보 제공을 통해 의료진 중심의 연구/임상정보 제공 솔루션이고, Big CEN TA(Text Analysis)는 고객사의 기간제 시스템에 다양한 형태로 존재하는 비정형 텍스트로부터 의미 있는 항목을 자동으로 추출하여 정형화하며, 정형과 비정형이 결합된 하이브리드 검색기능의 제공이 가능한 솔루션이다.

[(주)솔트룩스]는 기계학습과 자연어처리 스타트업으로 2000년에 창업된 (주)시스메타에 그 뿌리를 두고 있으며, 창업 이래 인공지능과 빅데이터 기술에 전념해 왔다. 동사의 Bigdata Suite는 End-to-End One stop 지능형 실시간 빅데이터분석 통합 플랫폼으로, 수집 엔진(Tornado), 저장/검색 엔진(DISCOVERY), 비정형 분석 엔진(TMS), 스트림분석 엔진(BlueBolt), 인지분석 엔진(CAS), 시각분석 엔진(Rainbow) 등 6개의 엔진을 제공한다.

[그림 16] 솔트룩스의 Bigdata Suite 구성도



\*출처: <http://www.saltlux.com/bigdata/summary.do?menuNumber=2>

[(주)아이지에이웍스]는 2006년 설립된 모바일 빅데이터 업체로 모바일 데이터분석, 마케팅 자동화, AI와 머신 러닝 기반의 DMP(Data Management Platform), 트레이드데스크, 마켓인텔리전스 등 모바일 데이터와 플랫폼을 제공해주는 국내 유일의 종합 Data-Tech SaaS(서비스형 소프트웨어) 기업이다.

동사의 애드브릭스(adbrix)는 앱 광고 성과 측정부터 사용자 분석, 타겟팅 광고, 운영을 동시에 해결할 수 있는 솔루션으로 고객사는 동사의 애드브릭스를 통하여 광고성과를 측정하고 유입된 이용자를 자유롭게 분석하고, 결과 데이터를 활용하여 광고와 운영을 고도화할 수 있다.

[(주)와이즈넷]은 1999년 미국 실리콘밸리에서 출발한 벤처기업으로, 자체 인공지능 기반 기술을 바탕으로 기업용 인공지능 SW(챗봇 및 빅데이터 수집, 분석, 검색) 제품을 제공한다. 동사의 주요 빅데이터 제품으로는 빅데이터 정보수집 솔루션 WISE BICrawler, 빅데이터검색솔루션 Search Formula-1 V5, 텍스트마이닝 솔루션 WISE TEA V2, 빅데이터 의미분석솔루션 WISE BIC Analyzer, 소셜분석 서비스 BuzzInsight 등이 있다.

[그림 17] 이씨마이너의 데이터마이닝SW ECMiner™ 제품 개요



\*출처: [http://ecminer.com/?page\\_id=132](http://ecminer.com/?page_id=132)

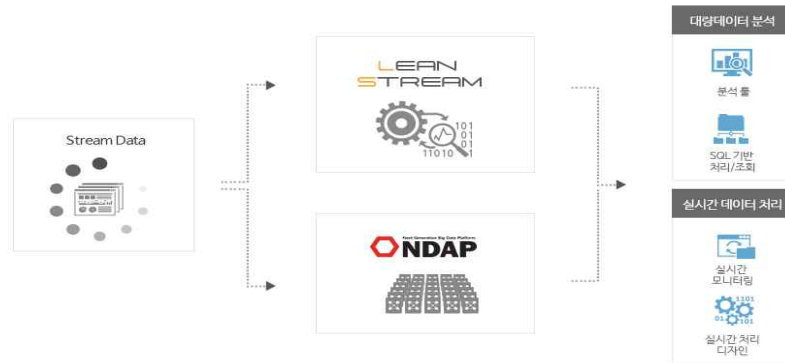
[(주)이씨마이너]는 빅데이터 분석에 특화된 솔루션을 개발하는 업체로, 2000년 창업 이래 Data Mining 기술력을 바탕으로 통합 분석 솔루션 공급 및 시스템 구축, Data Mining S/W 판매 및 응용 프로젝트 수행, Data Warehouse/Data Mart 구축 등으로 그 사업 영역을 확장하고 있다.

동사의 주요 제품으로는 데이터마이닝SW ECMiner™, 데이터마이닝 기반 지능형 모니터링시스템 ECMinerIMS™, 자금세탁방지 솔루션 ECMinerAML™ 등이 있다. ECMiner™는 통계 기반의 실시간 빅데이터 분석 솔루션으로, 지능형 모니터링, 연관분석, 텍스트 및 이미지 마이닝 등이 가능하다.

[(주)케이티넥스알(kt NexR)]은 빅데이터 End-to-End 제품/서비스를 제공하는 전문기업으로 빅데이터 글로벌 표준 기술인 하둡(Hadoop)에 있어 국내 최고수준의 기술력을 보유하고 있으며, 주요 제품으로는 실시간 빅데이터 처리 및 분석을 위한 엔터프라이즈 솔루션 Lean Stream, 빅데이터 플랫폼 NDAP(kt NexR Data Analytics Platform)이 있다.

[(주)코난테크놀로지]는 인공지능 기반 비정형데이터 마이닝 전문기업으로, 주요 분석 제품으로 비정형 데이터 통합 분석 플랫폼 Konan Analytics 4, 검색로그 분석 솔루션 Konan Log Analytics, 시맨틱 검색 기술 기반의 텍스트 마이닝 분석 서비스인 pulse-K가 있다.

[그림 18] kt NexR의 빅데이터 제품군



\*출처: <https://www.ktnexr.com/product/product.html>

[SK텔레콤(주)]은 SK그룹 계열의 이동통신업체이자 유가증권시장상장법인으로 H/W와 S/W를 결합하여 실시간 Big Data 분석을 위한 최적화된 솔루션을 제공한다. 동사에서 제공하는 Metatron은 실시간 빅데이터 분석 솔루션으로, 대용량의 로그 데이터 수집부터 실시간 분석을 통한 시각화까지 한번에 제공하는 End-to-End 분석 솔루션이다.

### 포트폴리오 분석

빅데이터 분석 솔루션 산업 관련 종목군으로 포트폴리오를 구성하였을 때의 주식 수익률 변화를 살펴보기 위해, 빅데이터 분석 솔루션 관련 코스닥 상장 기업을 5개사 선정하였으며, 2019년도 수익률은 2019.10.28. 종가를 기준으로 산정하였다.

KOSDAQ 지수를 벤치마크 지수로 분석한 결과 최근 4년간 빅데이터 분석 솔루션 관련 종목의 평균 수익률의 변동성이 심한 편이고, 2019년 들어 전년대비 대폭 개선된 수익률을 보이고 있다.

[표 11] 빅데이터 분석 솔루션 관련 주식 연도별 수익률 현황

단위: %				
구분	2016년	2017년	2018년	2019년
빅데이터 분석 솔루션 관련 종목 평균	-7.6	-6.4	-35.4	6.6
코스닥 평균	-6.8	26.3	-16.8	-1.6

출처: 한국기업데이터



### Ⅲ. 기술 심층 분석

#### 빅데이터 분석기술 기술동향

빅데이터 분석 기술은 다양한 출처로부터 폭발적으로 데이터양이 증가함에 따라, 비즈니스 영역에서 필수도구이자 수단으로 활용되고 있으며, 빅데이터 분석 분야의 중심이 기반 기술 영역에서 응용 기술 영역으로 빠르게 전환 중이다.

고도화되고 예측 가능한 분석기능들이 빅데이터 분석 기술 중 하나인 기계 학습과 통합되어 응용 기술의 새로운 성장세를 가져오고 있으며, 특히 의사결정 플랫폼과 비디오, 오디오, 이미지를 포괄하는 리치미디어 분석에 빠르게 접목되고 있다.

클라우드 기반의 빅데이터 분석 솔루션은 클라우드 서비스 형태로 빅데이터 분석 자원을 제공받을 수 있어 초기 투자비용이 저렴하고 활용성이 좋기 때문에 중소 규모의 조직을 중심으로 그 수요가 증가하고 있다.

빅데이터 분석 기술은 자연어(Natural Language) 처리 기반 텍스트·음성 데이터 분석 기술, 영상 데이터 내 콘텐츠 분석 기술, 데이터 마이닝 알고리즘 구현 기술, 데이터 특성 및 패턴 기반 시각화 기술로 구분된다.

#### 자연어 처리 기반 텍스트·음성 데이터 분석

자연어(Natural Language) 처리 기반 텍스트·음성 데이터 분석 기술은 자연어 기반의 방대한 텍스트 데이터의 패턴, 규칙, 관계 등의 정보를 추출하기 위한 기술을 의미한다.

[표 12] 자연어 처리 기반 텍스트·음성 데이터 분석 세부기술

세부분야	주요 내용
응용분야 대규모 말뭉치 구축	<ul style="list-style-type: none"> <li>응용분야의 분석 품질 향상을 위해 다양한 빅데이터 소스로부터 지속적으로 말뭉치를 자동으로 구축하고 갱신하는 기술</li> <li>활용 목적에 맞게 말뭉치의 헤더 및 본문에 자동으로 태그를 부착</li> </ul>
컨텍스트 기반 정보 분류 및 분석	<ul style="list-style-type: none"> <li>비정형 텍스트 및 음성 데이터로부터 구문 분석, 의미 분석을 통하여 응용분야에 따라 의미있는 정보를 추출하는 기술</li> <li>감성 및 오피니언 분석 등 목적에 맞는 기계학습(Machine Learning), 텍스트마이닝(Text-mining)</li> </ul>
다국어 처리	<ul style="list-style-type: none"> <li>다국어 형태소 분석 및 품사 태깅 기술을 포함해서 문서 언어 감지, 다국어 문장 분리 등 다양한 언어(한국어, 영어, 일본어 등)가 복합적으로 구성된 문서를 처리하기 위한 기술</li> <li>구조화되지 않은 텍스트 본문에서 이름, 장소 및 주요 개념을 찾아 주고 외래어를 식별</li> </ul>

\*출처: 빅데이터 분석 솔루션, 한국신용정보원(2015), KED 재가공

자연어 처리 기반 텍스트·음성 데이터 분석 기술은 응용분야별 대규모 말뭉치 구축 기술, 컨텍스트 기반 정보 분류 및 분석 기술, 다국어 처리 기술 등으로 구분되며, 자연어 처리 기반 텍스트·음성 데이터 분석은 기계학습 알고리즘을 활용하여 말뭉치 구축 자동화, 텍스트 내 개념 추출 정교화 등을 목표로 하고 있다.

## 영상 데이터 내 콘텐츠 분석 기술

영상 데이터 내 콘텐츠 분석은 영상을 분석하여 내포된 특성을 파악하고 패턴을 추출하여 객체 인식, 상황 감지, 모션 추적 및 검색을 가능하게 한다.

영상 데이터 내 콘텐츠 분석 기술은 특정 객체 검출 및 어노테이션(Annotation), 컨텍스트 기반 상황 감지 및 인식, 객체 추적 및 예측 기술 등으로 구분되며, 영상 데이터 내 콘텐츠 분석 기술은 안전 및 보안, 검색 및 관리 등 다양한 분야에서 적용 가능하다.

[표 13] 영상 데이터내 콘텐츠 분석 세부기술

세부분야	주요 내용
특정 객체 검출 및 어노테이션	<ul style="list-style-type: none"> <li>요청한 영상과 비슷한 영상을 검색해 오는 유사영상 검색 기술</li> <li>사람의 얼굴이나 자동차 등 특정 객체를 영상으로부터 검출</li> <li>영상이 표현하는 내용에 대응하는 메타 데이터(Metadata)를 자동으로 부여</li> </ul>
컨텍스트 기반 상황 감지 및 인식	<ul style="list-style-type: none"> <li>특정 컨텍스트 하에서, 안면 근육의 움직임을 분석하여 감성을 인식하거나, 객체의 이동 경로의 변화 등 행동 패턴의 변화를 발견하기 위하여 모니터링하고 감지</li> </ul>
객체 추적 및 예측	<ul style="list-style-type: none"> <li>이동 경로 및 속도 등의 정보를 획득하는 객체 추적 기술</li> <li>객체 추적을 통해 범죄 징후, 고장 징후 등 향후 예상되는 객체의 행동을 사전에 예측</li> </ul>

\*출처: 빅데이터 분석 솔루션, 한국신용정보원(2015), KED 재가공

## 데이터 마이닝 및 알고리즘 구현기술

데이터 마이닝 알고리즘 구현 기술은 일반적인 마이닝 기술로 처리하기 어려운 특성을 갖고 있는 빅데이터에서 효과적으로 데이터 마이닝을 할 수 있는 기반 기술로, 대용량 분산 데이터 기반 마이닝, 실시간 스트림 데이터 기반 마이닝, 불확실/불완전 데이터 처리 알고리즘 등으로 구분된다.

[표 14] 데이터 마이닝 및 알고리즘 구현기술 세부기술

세부분야	주요 내용
대용량 분산 데이터 기반 마이닝 구현	<ul style="list-style-type: none"> <li>노드별 편향된 데이터 해결 및 전체 최적화를 위한 데이터 교환 및 모델 융합 알고리즘</li> <li>로컬 데이터 기반의 통계량 계산 및 노드간 공유 기술</li> <li>로컬 마이닝 수행 및 글로벌 패턴 합성, 로컬 생성 모델 연관 분석 및 통합 인사이트 도출</li> </ul>
실시간 스트림 데이터 기반 마이닝	<ul style="list-style-type: none"> <li>저장된 데이터가 아닌 스트림 데이터에서 마이닝 결과를 얻기 위한 메커니즘 및 알고리즘</li> <li>메모리 사용 및 처리소요 시간을 고려하고 시간 경과에 따라 데이터의 생성 방식의 변화에 대응</li> </ul>
불확실/불완전 데이터 처리 알고리즘	<ul style="list-style-type: none"> <li>희소 데이터 처리를 위한 차원축소 및 샘플 데이터 추가 기술</li> <li>불확실 데이터 처리를 위한 평균 및 분산 활용 마이닝, 불완전 데이터 처리를 위한 결측치 보정</li> </ul>

\*출처: 빅데이터 분석 솔루션, 한국신용정보원(2015), KED 재가공

데이터 마이닝 알고리즘 구현 기술은 대용량 데이터 처리 기술과 연계하여 활용되며, 특정 응용 분야에 적합하도록 고도화되어 활용된다.

## 데이터 특성 및 패턴 시각화 기술

데이터 특성 및 패턴 기반 시각화 기술은 기존의 테이블 혹은 단순 차트 형태의 데이터 레포팅(Reporting) 기술로 처리가 어려운 대용량의 데이터를 사용자들이 쉽게 인지할 수 있도록 효과적으로 가시화한다.

데이터 특성 및 패턴 기반 시각화 기술은 대용량 데이터 분석을 위한 인터랙티브 시각화, 다양한 데이터소스 통합 매쉬업(Mash up), 패턴 기반 시각화 기법 구현 기술 등으로 구분된다.

데이터 특성 및 패턴 기반 시각화 기술은 업체 내 빅데이터 분석 전문 기술 인력이 부족한 환경에서 빅데이터 분석을 활용할 수 있도록 지원하며 저비용의 효율적 빅데이터 분석 기술 활용이 가능하다.

[표 15] 데이터 특성 및 패턴 시각화 기술 세부기술

세부분야	주요 내용
대용량 데이터 분석 위한 인터랙티브 시각화	<ul style="list-style-type: none"> <li>대용량 데이터를 사용자가 인식할 수 있는 형태로 효과적으로 시각화하기 위한 데이터 축소 및 표현 기술</li> <li>데이터 크기 및 유형에 따라 가장 적합한 시각화 방식 선정하며, 효과적인 시각화를 통해 새로운 인사이트 발굴 유도</li> <li>대용량 데이터에 대한 사용자의 질의를 실시간으로 처리하여 시각화</li> </ul>
다양한 데이터 소스 통합 매쉬업	<ul style="list-style-type: none"> <li>사용자가 별도 교육 필요 없이 데이터를 탐색할 수 있는 직관적인 화면 구성</li> <li>이질적(Heterogeneous) 데이터 소스로부터 데이터를 통합할 수 있는 데이터 구조</li> <li>대량 데이터 쿼리(Query), 집계 및 사전 계산 등 데이터를 신속히 추출</li> </ul>
패턴 기반 시각화 기법 구현	<ul style="list-style-type: none"> <li>로그 데이터, 객체, 공간정보 간 상관관계 등 도메인별 조회 패턴에 따라 사용자들이 직관적으로 이해할 수 있는 형태의 시각화 기법을 개발하고 구현</li> </ul>

\*출처: 빅데이터 분석 솔루션, 한국신용정보원(2015), KED 재가공

## 빅데이터 분석 기술 전망

시장조사기관 가트너는 2019년 10대 데이터 및 분석 기술 트렌드를 발표했다. 가트너는 향후 3~5년간 큰 영향력을 미칠 데이터 및 분석 기술 트렌드로 증강 분석(augmented analytics), 지속적 지능화(continuous intelligence), 그리고 설명 가능한 인공지능(artificial intelligence)을 꼽았다.

[트렌드 1: 증강 분석] 가트너는 증강 분석을 데이터 및 분석 시장의 차세대 혁신의 물결이라고 예측했다. 증강 분석은 머신러닝(ML)과 AI 기술을 사용해 분석 콘텐츠가 개발, 소비, 공유되는 방식을 혁신한다. 증강 분석은 2020년까지 분석과 비즈니스 인텔리전스, 데이터 사이언스와 머신러닝 플랫폼, 그리고 임베디드 분석

이 신규 구매를 촉진하는 주요인이 될 것으로 예상된다.

[트렌드 2: 증강 데이터 관리] 증강 데이터 관리는 머신러닝 기능과 AI 엔진을 활용해 기업 정보 관리 카테고리를 생성하는 것이다. 여기에는 데이터 품질, 메타데이터 관리, 마스터 데이터 관리, 데이터 통합, 데이터베이스 관리 시스템(DBMS) 자체 구성 및 자체 조정 등이 포함된다. 이를 통해 많은 수작업이 자동화되고, 기술 숙련도가 부족한 사용자들은 데이터를 사용해 더욱 자율적으로 작업할 수 있으며, 숙련된 기술자들은 보다 가치 있는 작업에 집중할 수 있다.

증강 데이터 관리를 통해 메타데이터는 회계 감사와 계보, 보고용으로만 사용되는 것이 아니라, 동적 시스템을 실행하는 것으로 전환된다. 메타데이터는 수동에서 능동으로 바뀌고 있으며, 모든 인공지능과 머신러닝을 실행하는 주 동력이 되고 있다. 2022년 말에 이르면 머신러닝과 자동화 서비스 수준 관리가 추가돼 데이터 관리 수작업이 45%가량 줄어들 것으로 전망된다.

[트렌드 3: 지속적 지능화] 2022년에 이르면 주요 신규 비즈니스 시스템의 절반 이상이 실시간 컨텍스트 데이터를 사용하는 지속적 지능화를 통합하여 의사 결정을 개선할 것으로 보인다. 지속적 지능화는 실시간 분석이 비즈니스 운영에 통합되는 설계 패턴으로, 현재와 과거의 데이터를 처리해 사건에 대응하는 행동을 정한다. 이는 의사 결정을 자동화하거나 지원하게 된다. 지속적 지능화는 증강 분석, 이벤트 스트림 프로세싱, 최적화, 비즈니스 규정 관리, 머신러닝 등 다양한 기술을 활용한다.

[트렌드 4: 설명 가능한 AI] 점점 더 많은 AI 모델이 인간의 의사 결정을 강화하고 대체하는 데 사용되고 있다. 그러나 기업은 때때로 이러한 모델이 의사 결정에 어떻게 도달하는지를 정당화해야 한다. 사용자 및 이해당사자들과 신뢰를 구축하기 위해, 애플리케이션 리더들은 AI 모델을 한층 해석하기 쉽고 설명 가능한 방식으로 만들어야 한다. 안타깝게도 이러한 고급 AI 모델의 대다수는 복잡한 블랙박스로, 특정 권고사항이나 결정사항에 도달한 이유를 설명하지 못한다. 예를 들어, 데이터 사이언스와 머신러닝 플랫폼에서 설명 가능한 AI는 모델의 정확성, 속성, 모델 통계, 기능 등의 측면에서 모델에 대한 설명을 자연어로 자동 생성한다.

[트렌드 5: 그래프] 그래프 분석은 조직, 사람, 거래 등 이해 주체 간 관계를 탐색할 수 있는 일련의 분석 기법이다. 그래프 처리 애플리케이션과 그래프 데이터베이스 관리 시스템은 2022년까지 매년 100%씩 성장하며 데이터 준비를 지속적으로 가속화하고, 더욱 복잡하고 적응력 뛰어난 데이터 사이언스를 구현할 것이다. 복잡한 데이터 전반에 걸쳐 복잡한 질문을 제기할 필요성이 대두됨에 따라, 그래프 분석은 향후 몇 년 내 성장할 것이다. SQL 쿼리를 사용하는 방법은 항상 실용적이지 않을 수 있으며, 대규모 실행은 불가할 수 있다.

[트렌드 6: 데이터 패브릭] 데이터 패브릭은 분산된 데이터 환경에서 마찰 없는

액세스와 데이터 공유를 가능케 한다. 이를 통해 일관된 단일 데이터 관리 프레임워크를 구축할 수 있으며, 사일로화된 저장소를 설계단에서 바꾸어 원활한 데이터 액세스와 프로세싱을 지원한다. 2022년까지 맞춤형 데이터 패브릭 설계는 주로 정적인 인프라로 구축될 것이며, 기업들은 보다 동적인 데이터 메시 접근방식을 완전히 재설계하기 위해 새로운 비용을 지불해야 할 것이다.

[트렌드 7: NLP 및 대화형 분석] 2020년까지 분석 쿼리의 50%가 검색, 자연어 처리(NLP), 음성을 통해 생성되거나 자동 생성될 것이다. 복잡한 데이터 조합을 분석하고 조직 내 모든 이들이 분석에 액세스해야 할 필요성이 늘어나면서 해당 기술에 대한 폭넓은 채택이 이뤄질 것이며, 이를 통해 분석 톨은 검색 인터페이스나 가상 비서와의 대화만큼 용이해질 것이다.

[트렌드 8: 상용 AI와 머신러닝] 가트너는 2022년에 이르면 AI와 머신러닝 기술을 활용하는 새로운 최종 사용자 솔루션의 75%가 오픈소스 플랫폼이 아닌 상용 솔루션으로 구축될 것으로 전망한다. 상용 업체들은 현재 오픈소스 생태계에 커넥터를 구축했으며, AI와 머신러닝을 확장하고 대중화하기 위해 오픈소스 기술에 결여된 프로젝트 및 모델 관리, 재사용, 투명성, 데이터 계보, 플랫폼 응집력, 통합 등의 엔터프라이즈 기능을 제공한다.

[트렌드 9: 블록체인] 블록체인 기술이 핵심적으로 제안하는 가치는 신뢰할 수 없는 참여자 네트워크 전반에 분산된 신뢰를 제공하는 것이다. 분석 사용 사례에 대한 잠재적 영향은 중요하며, 참여자 관계나 상호 작용을 활용하는 경우에는 특히 중요하다.

그러나 4, 5가지의 주요 블록체인 기술이 주류로 자리잡기까지는 수년이 걸릴 것이다. 그 때까지 최종 기술 사용자들은 주요 고객이나 네트워크가 요구하는 블록체인 기술과 표준을 통합할 수밖에 없을 것이다. 여기에는 기존 데이터 및 분석 인프라와의 통합도 포함된다. 통합 비용은 잠재적 이익보다 더 클 수도 있다. 블록체인은 데이터베이스가 아닌 데이터 소스이며, 기존 데이터관리 기술을 대체하지는 않을 것이다.

[트렌드 10: 퍼시스턴트 메모리 서버] 새로운 퍼시스턴트-메모리 기술을 이용하면 인메모리 컴퓨팅(IMC) 기반 아키텍처를 채택하는 데 드는 비용과 복잡성을 줄일 수 있다. 퍼시스턴트 메모리는 고성능 워크로드에 비용 효율적인 대용량 메모리를 제공할 수 있는, D램과 낸드 플래시 메모리 사이 새로운 메모리 계층을 의미한다. 이는 애플리케이션 성능, 가용성, 부팅 시간, 클러스터링 방법, 보안 방식 등을 향상시키면서 비용을 제어할 수 있는 잠재력을 갖췄다. 또한, 데이터 복제 필요성을 줄여 기업이 애플리케이션과 데이터 아키텍처 복잡성을 줄일 수 있도록 지원한다.



## 빅데이터 분석 핵심 기술 확보 경쟁력 확보를 위한 제언

빅데이터 분석 솔루션의 핵심 기술에는 대용량 데이터 저장 및 처리 기술, 실시간 대용량 스트림 분석, 빅데이터 기반 고급 분석, 대용량 데이터 시각화 분석 및 인터페이스 기술 등이 있다.

대용량 데이터 저장 관리는 여러 대의 서버에 대용량 데이터를 나누어서 저장하고, 동시에 데이터를 읽어 처리하는 분산 시스템 기술로서, 데이터의 일관성, 사용성 등이 요구된다. 실시간 대용량 스트림 분석 기술은 방대한 규모로 연속적으로 생성되는 데이터를 생성 직후에 실시간으로 인-메모리 상태에서 감지, 처리 및 의사결정을 지원하는 기술로서 실시간성, 보안성, 안정성, 확장성 등이 요구된다. 데이터 분석 기술은 통계 분석을 기반으로 데이터 마이닝, 예측 분석 등의 기계학습 기술이 활용된다. 대용량 데이터 시각화 기술은 데이터 분석 결과를 효과적으로 전달하기 위한 기술로서 사용자 인터페이스의 간소화 및 효율화가 요구된다.

빅데이터 분석 기술은 글로벌 기업들이 국내 빅데이터 비즈니스 분야와 공공 빅데이터 처리 분야에서 이미 초기 시장을 선점하고 있고, 국내 빅데이터 예측·분석 기술은 최고 기술국 대비 40% 미만으로 초기단계에 머물러 있다는 점은 국내 빅데이터 기술경쟁력 강화 정책을 추진할 필요성을 시사하고 있다.

혁신성장동력 시행계획에 따르면, 정부는 초연결 지능화를 위한 빅데이터 기초원천 기술 연구로 빅데이터 기술 강국 도약을 목표로, 2022년까지 인공지능을 위한 빅데이터(Bigdata for AI) 기술 확보와 기술선진국 대비 국내기술 수준 90% 달성(2016년 78.2%→2022년 90%)하고, 2027년까지 초연결 사회를 위한 빅데이터(BigData for Hyper-Connected Society) 기술 확보와 기술선진국 대비 국내기술 수준 93% 달성(2016년 78.2%→2027년 93%)을 세부 목표로 제시하고 있다.

정부는 이와 더불어 공공부문을 필두로 국내 우수 빅데이터 분야에 대한 SW 활용연계를 강화할 필요가 있다. 국산 빅데이터 SW를 검증하여 정부사업에서 활용할 수 있도록 지원하고, 공공조달 시장에서 국내 기업의 기술력을 적용한 SW 개발이 우선적으로 이루어질 수 있도록 점진적인 국내 SW 연계 정책을 통해 자생력을 갖출 필요가 있다.