

실무에 적용 가능한 Big Data 분석 개론

빅데이터 발전 방향

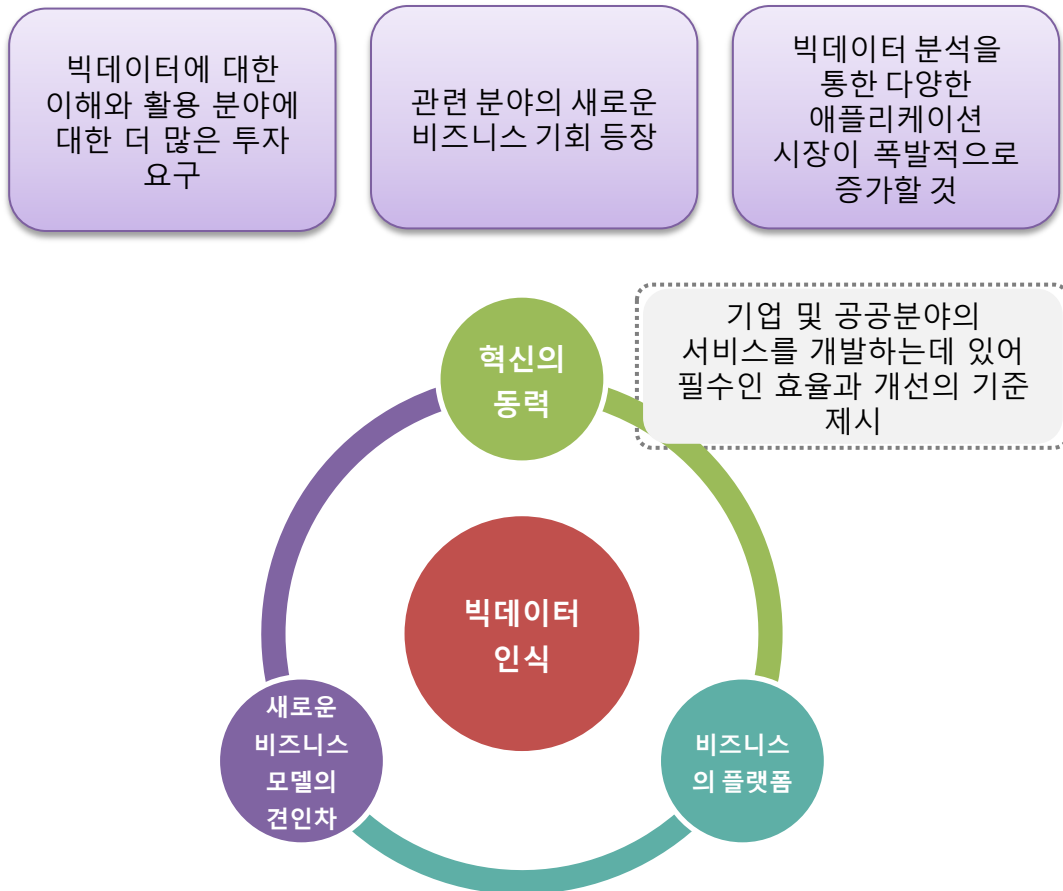


한국기술교육대학교
온라인평생교육원

■ 빅데이터 시장 동향

1. 빅데이터 시장

지속적인 성장 전망



빅데이터는 신규 진입 장벽을 낮추고 기존 서비스를 개선할 수 있는 혁신 플랫폼으로 작용함

■ 빅데이터 시장 동향

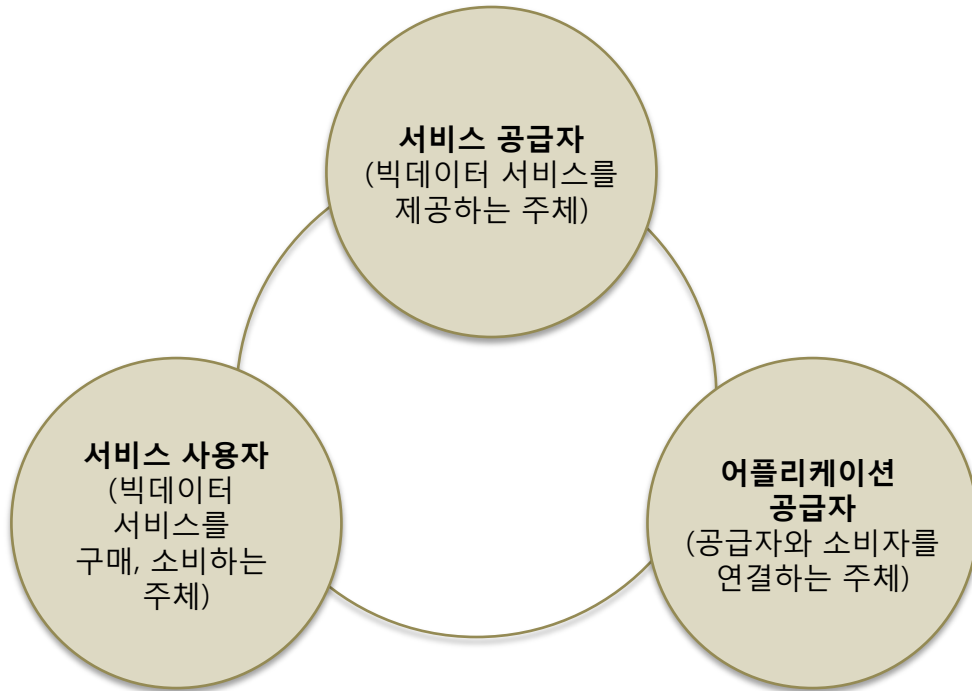
1. 빅데이터 시장

빅데이터 중요성의 심화



■ 빅데이터 시장 동향

2. 빅데이터 에코시스템



<한국정보화진흥원, 빅데이터 시대 : 에코시스템을 둘러싼 시장경쟁과 전략분석, 2012>

유형에 따른 빅데이터 공급자

하드웨어 공급자

- 자체적인 데이터 센터 및 클라우드 시스템을 통해 빅데이터 서비스를 위한 인프라를 생성하거나 실행 및 유통시킴

처리 소프트웨어 공급자

- 사용자가 저장한 빅데이터를 효과적으로 저장하고 처리할 수 있는 소프트웨어를 제공함

분석 소프트웨어 공급자

- 사용자의 빅데이터를 사용자의 요구에 맞게 또는 시장 요구에 맞는 데이터 분석틀을 제공함

■ 빅데이터 시장 동향

2. 빅데이터 에코시스템

서비스 레벨에 따른 빅데이터 공급자

인프라 공급자

- 데이터를 저장, 처리하는 등 빅데이터를 위한 **기초 자원**을 담당하는 하드웨어 및 운영 체제를 제공함
- 자체 인프라를 구축하거나, 클라우드 컴퓨팅 인프라를 구축함

플랫폼 공급자

- 빅데이터를 위한 클라우드 컴퓨팅 서비스와 하드웨어에 종속되지 않은 **처리소프트웨어, 분석 소프트웨어**를 공급함

애플리케이션 공급자

- 사용자가 주로 웹 브라우저를 통해 빅데이터와 소통하는 **메커니즘**을 제공하고, 빅데이터 처리 결과를 바탕으로 소비자가 원하는 분석결과를 시각화 등을 통해 제공함

■ 빅데이터 시장 동향

2. 빅데이터 에코시스템

주요 글로벌 빅데이터 서비스 기업의 특성

	공급자	서비스 수준	사용자	개방	범위
IBM	데이터센터	인프라, 플랫폼, 애플리케이션	기업	비공개	로컬, 커뮤니티
Microsoft	클라우드 데이터센터	인프라, 플랫폼, 애플리케이션	소비자, 기업	공개, 비공개, 하이브리드	글로벌
Google	클라우드	인프라, 플랫폼, 애플리케이션	소비자, 기업	공개	글로벌
Amazon	클라우드	인프라, 플랫폼, 애플리케이션	소비자, 기업	공개	글로벌
HP	데이터센터	인프라, 플랫폼	소비자, 기업	비공개	로컬, 커뮤니티
SAS		플랫폼, 애플리케이션	기업	비공개	로컬, 커뮤니티
Cloudera		플랫폼, 애플리케이션	기업	비공개	로컬, 커뮤니티

	공급자	서비스 수준	사용자	개방	범위
IBM	데이터센터	인프라, 플랫폼, 애플리케이션	기업	비공개	로컬, 커뮤니티
Microsoft	클라우드 데이터센터	인프라, 플랫폼, 애플리케이션	소비자, 기업	공개, 비공개, 하이브리드	글로벌
Google	클라우드	인프라, 플랫폼, 애플리케이션	소비자, 기업	공개	글로벌
Amazon	클라우드	인프라, 플랫폼, 애플리케이션	소비자, 기업	공개	글로벌
HP	데이터센터	인프라, 플랫폼	소비자, 기업	비공개	로컬, 커뮤니티
SAS		플랫폼, 애플리케이션	기업	비공개	로컬, 커뮤니티
Cloudera		플랫폼, 애플리케이션	기업	비공개	로컬, 커뮤니티

<한국정보화진흥원, 빅데이터 시대 : 에코시스템을 둘러싼 시장경쟁과 전략분석, 2012>

■ 빅데이터 시장 동향

2. 빅데이터 에코시스템

애플리케이션 레벨

실제 빅데이터와 사용자가 만나게 되는
커뮤니케이션 레벨

빅데이터를 통한 새로운 비즈니스의 기회, 신규서비스 적용 등 빅데이터의
활용에 해당함

인프라 및 데이터 처리기술 분야보다 진입장벽이 상대적으로 낮음

빅데이터 시장에서도 애플리케이션 분야의 확장과 참여가 기대되고 있음

빅데이터 서비스의 경쟁력

데이터의 처리와 분석 능력

하드웨어는 빅데이터를 구성하는 중요한 요소 중 하나임

데이터 처리속도, 데이터 저장 용량 등
빅데이터를 구성하는 인프라

- ➡ 클라우드 컴퓨팅 기술의 발전에 따라 과거 개별 사용자 또는 기업이 할 수 있는 범위를 크게 증가하는 하드웨어 인프라가 클라우드 서비스 공급자를 통해 이용이 가능해짐

소프트웨어 기술 분야에 대한 관심이 높아지고 있음

- ➡ 수집된 빅데이터를 실시간으로 분석할 수 있는 컴퓨팅 기술

예 분산 병렬 컴퓨팅 기술 : 하둡, 맵리듀스, NoSQL 등

실무에 적용 가능한 Big Data 분석 개론

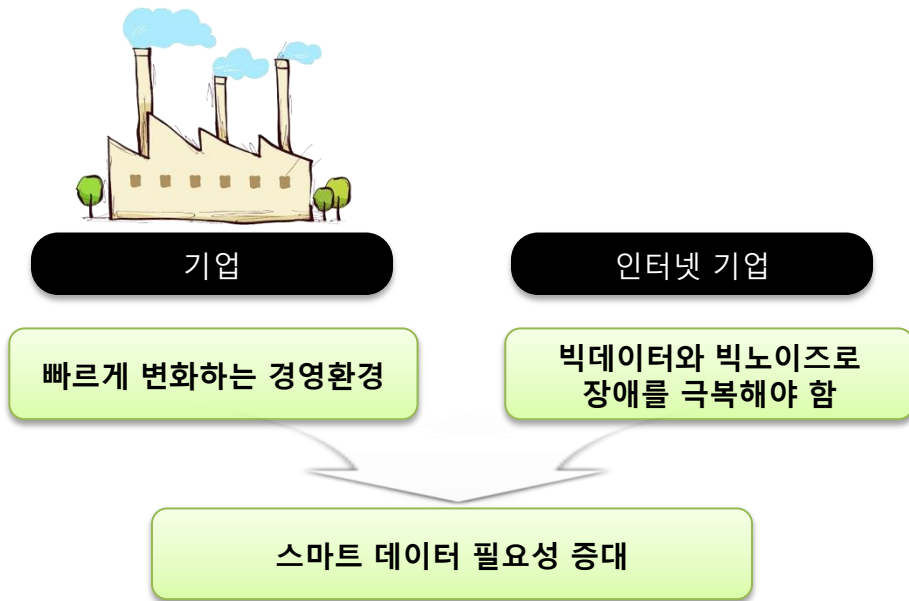
빅데이터 발전 방향



한국기술교육대학교
온라인평생교육원

■ 빅데이터의 품질 요구

1. 스마트 데이터란?



스마트 데이터의 시대 도래를 전망함



■ 빅데이터의 품질 요구

1. 스마트 데이터란?

스마트 데이터의 3가지 특성(3A)



정확성(Accurate)



행동성(Actionable)



민첩성(Agile)

정확성
(Accurate)

행동성
(Actionable)

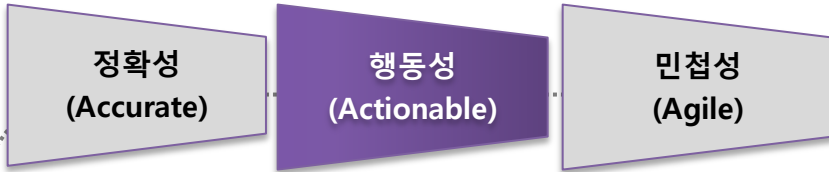
민첩성
(Agile)

- 빅데이터의 노이즈로부터 정확하고 양질의 정보를 전달할 수 있는 것을 의미함
 - ✓ 스마트 데이터가 정확하려면 지속적으로 제3자의 벤치마크를 통해 유효성이 입증
 - ✓ 스마트 데이터 이용자는 지속적으로 투자 대비 효과성을 검증할 수 있어야 함
- ➡ 데이터는 새로운 가치를 창출할 수 있을 정도로 정확성을 가지고 직설적이어야 함

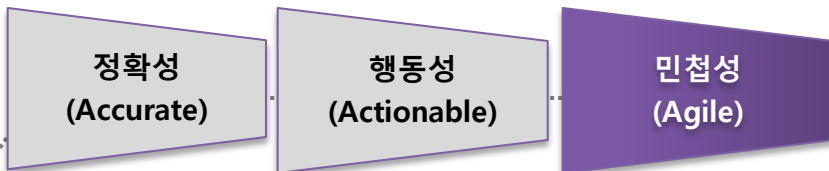
■ 빅데이터의 품질 요구

1. 스마트 데이터란?

스마트 데이터의 3가지 특성(3A)



- 스마트 데이터는 조직이 바로 행동으로 옮길 수 있는 특성을 보유함으로써 가치를 창출하는 원동력으로서 작동함
 - ✓ 시시각각 다양하고 복잡한 데이터가 생성되는 환경에 필요한 즉각적 의사결정은 상당한 시간을 필요로 하는 보고서 분석 등 과거 행동 방식으로는 빅데이터 시대에 부적절함
 - ✓ 기업 고객의 활동 영역을 최대화하고 고객이 바로 행동과 서비스를 할 수 있는 데이터의 확보는 수익이나 시장 점유율로 직결되는 **성장 동력으로 작동함**
- ➡ 데이터는 즉각적으로 계량이 가능한 행동의 원동력이 되어야 함

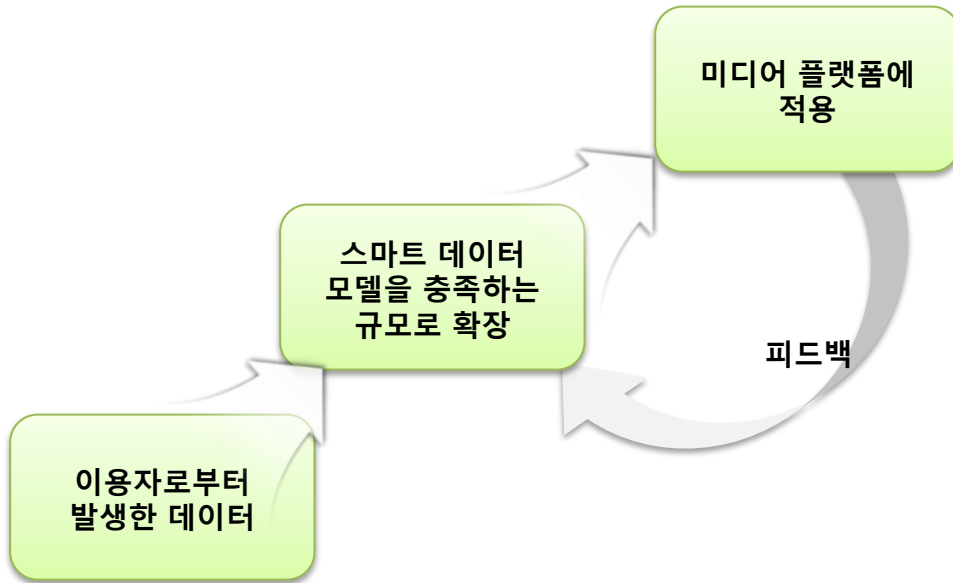


- 급변하는 비즈니스 환경에서 실시간으로 데이터 분석이 가능하도록 함
 - ✓ 비즈니스 결정을 내리기 위해 몇 주를 필요로 하는 업무 환경은 실시간으로 의사 결정을 필요로 하는 환경으로 변화되고 있음
 - ✓ 고객과 의사결정자가 실시간으로 결정을 내릴 수 있도록 데이터도 민첩성을 지원하는 형태로 변화해야 함
 - ✓ **모든 것이 즉각적인 행동을 취할 수 있도록** 변화하는 기술적 요구 조건과 비즈니스 조건에 따라 데이터의 속성도 변화를 필요로 함
- ➡ 변화하는 비즈니스 환경에 대한 유연성이 중요함

■ 빅데이터의 품질 요구

2. 스마트 데이터 환경

스마트 데이터의 환경



스마트 데이터의 환경은
순환 과정을 거칠 수 있도록 구현되어야 함

■ 빅데이터의 품질 요구

2. 스마트 데이터 환경

빅데이터와 스마트 데이터의 구분 기준

데이터 품질 평가

- 기업 등 조직에서 지속적으로 데이터 품질과 관련된 문제 발생 여부
- 관계자들이 데이터의 정확도에 대해 의문을 가지는지 여부

데이터 범위

- 데이터를 기반으로 차별화된 비즈니스 사례를 만들기 위한 설득력 있는 근거를 확보하였는지 여부
 - ✓ 빅데이터는 본질적으로 비즈니스 목적으로 확장되었거나 기업을 성장시키기 위한 목적으로 탄생되지 않았으므로, 근거를 확보하는 것이 중요한 척도
- 데이터가 기업의 수익이나 시장 점유율의 원동력으로 활용되고 있는지 여부
- 비용 절감을 이루고 있는지 여부

실시간성 판단

- 빅데이터가 묶음(batch)으로 처리되고 분석 결과로부터 행동을 취하기 위해 수 시간을 기다려야 하는지에 대한 여부

데이터의 객관적 유효성 입증

- 제 3자 혹은 독립적 시각에서 지속적으로 데이터의 정확함을 평가하는지에 대한 여부

■ 빅데이터의 품질 요구

2. 스마트 데이터 환경

빅데이터와 스마트 데이터의 구분 기준

행동성

- 매일 데이터가 어떠한 비즈니스 행동의 근거 혹은 원동력으로 작동되는지 설명할 수 있는 지에 대한 여부

실행 자원 구축

- 빅데이터 전략을 성취하기 위해서 다수의 컨설턴트, 플랫폼 및 서비스 벤처들을 필요로 하는지에 대한 여부

유연성 확보

- 데이터를 핸드폰과 같은 새로운 플랫폼에 적용시키거나 변화하는 비즈니스 조건에 맞춰 변화시키는 것에 대한 어려움의 여부

■ 빅데이터의 품질 요구

2. 스마트 데이터 환경

빅데이터와 스마트 데이터에 대한 점검 사항

구분	점검 사항
데이터 품질평가	지속적으로 데이터 품질과 관련된 문제의 발생 여부
	관계자들이 데이터의 정확도에 대해 의문을 가지는지 여부
데이터 범위	데이터를 기반으로 차별화된 비즈니스 사례를 만들기 위해 노력하지만 설득력 있는 근거를 확보하였는지 여부
	데이터가 기업의 수익이나 시장 점유율의 원동력이 되는지 여부
	비용 절충을 해야 하는지 여부
실시간성 판단	빅데이터가 묶음(Batch)으로 처리되고 분석 결과로부터 행동을 취하기 위해 수 시간을 기다려야 하는지에 대한 여부
데이터의 객관적 유효성 입증	제 3자 혹은 중립적 시각에서 지속적으로 데이터의 정확함을 평가하는지에 대한 여부
행동성	매일 데이터가 어떠한 비즈니스 행동의 근거 혹은 원동력으로 작동되는지 설명 가능 여부
실행 자원 구축	빅데이터 전략을 성취하기 위해서 다수의 컨설턴트, 플랫폼 및 서비스 벤처들을 필요로 하는지에 대한 여부
유연성 확보	데이터를 핸드폰과 같은 새로운 플랫폼에 적용시키거나 변화하는 비즈니스 조건에 맞춰 변화시키는 것에 대한 어려움의 여부

<한국정보화진흥원, 빅데이터의 진화: 스마트 데이터, 20132>