

실무에 적용 가능한 Big Data 분석 개론

---

# 빅데이터 이해와 동향



한국기술교육대학교  
온라인평생교육원

## ■ 빅데이터의 정의와 특성

### 1. 빅데이터의 정의



기존의 일반적인 기술로는 다루기 힘든  
대용량의 데이터

#### 2011년 정의

맥킨지(McKinsey)

IDC(International  
Data  
Corporation :  
세계적인  
시장조사기관)

일반적인 데이터베이스 관리 시스템(DBMS ; Data Base  
Management System)으로 저장, 관리, 분석할 수 있는  
범위를 초과하는 대규모 데이터

## ■ 빅데이터의 정의와 특성

### 1. 빅데이터의 정의

#### 2011년 정의

맥킨지(McKinsey)

IDC(International  
Data  
Corporation :  
정보 통신 기술  
시장조사 기관)

다양한 조율의 대규모 데이터로부터 가치를 추출하고  
데이터의 초고속 수집, 발굴, 분석을 지원하도록  
고안된 차세대 기술

#### 2012년 정의

가트너(Gartner)

삼성경제연구소

함유근, 채승병

대용량, 빠른 속도, 다양성 높은 정보자산

## ■ 빅데이터의 정의와 특성

### 1. 빅데이터의 정의

2012년 정의

가트너(Gartner)

삼성경제연구소

함유근, 채승병

- 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합을 지칭함
- 대규모 데이터와 관계된 기술 및 도구를 포함함

가트너(Gartner)

삼성경제연구소

함유근, 채승병

- 보통 수십에서 수천 테라바이트(TB) 정도의 거대한 크기를 가짐
- 여러 가지 다양한 비정형 데이터를 포함하고 있음
- 생성 - 유통 - 소비가 몇 초에서 몇 시간 단위로 일어나 기존의 방식으로는 관리와 분석이 매우 어려운 데이터 집합

## ■ 빅데이터의 정의와 특성

### 2. 빅데이터(Big Data)의 특성

#### 빅데이터의 3V



#### 규모(Volume)

1 처리해야 할 데이터의 크기를 말하는 속성

데이터 규모가  $10^{12}$  바이트

2 **테라바이트(Terabyte, TB)**급 이상의 데이터군을 빅데이터로 통칭함

3 정보통신 기술의 발달로 최근에는 더 큰 규모의 빅데이터를 접할 수 있음

#### 디지털 데이터 단위

데이터 용량 구분(Memory unit)	크기(Size)
Kilobyte(KB)	$10^3$ 바이트
Megabyte(MB)	$10^6$ 바이트
Gigabyte(GB)	$10^9$ 바이트
Terabyte(TB)	$10^{12}$ 바이트
Petabyte(PB)	$10^{15}$ 바이트
Exabyte(EB)	$10^{18}$ 바이트
Zettabyte(ZB)	$10^{21}$ 바이트
Yottabyte(YB)	$10^{24}$ 바이트

## ■ 빅데이터의 정의와 특성

### 2. 빅데이터(Big Data)의 특성

#### 다양성(Variety)

1 처리해야 할 데이터의 유형이 다양함을 말하는 속성

2 빅데이터는 다양한 데이터 유형을 가짐

정형 데이터 (Structured data)	비정형 데이터 (Unstructured data)	반정형 데이터 (Semi-structured data)
고정된 필드에 저장된 데이터 <b>예</b> 관계형 데이터베이스, 스프레드시트	형태와 구조가 복잡한 데이터 <b>예</b> 소셜 데이터, 문서, 이미지, 오디오, 비디오, 동영상	값과 형식이 다소 일관성이 없는 데이터 <b>예</b> HTML, XML, 웹문서, 웹로그, 센서 데이터

3 데이터의 양과 더불어 유형의 복잡성이 증대함

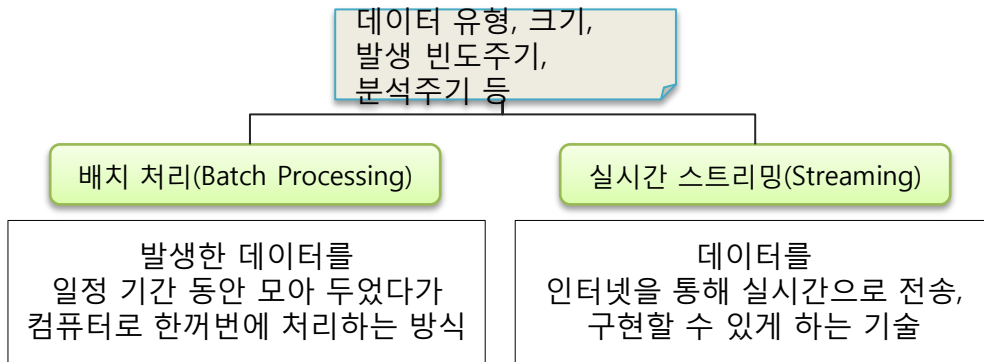
정보통신 기술의 발전과 더불어 **관측 데이터**가 의미 있는 분석의 대상으로 포함됨

소셜 데이터, 센서 데이터 등

#### 속도(Velocity)

1 대용량의 데이터를 빠르게 처리하고 분석할 수 있는 속성

2 데이터의 빠른 처리 및 분석을 위해 다양한 방식을 적용함



실무에 적용 가능한 Big Data 분석 개론

---

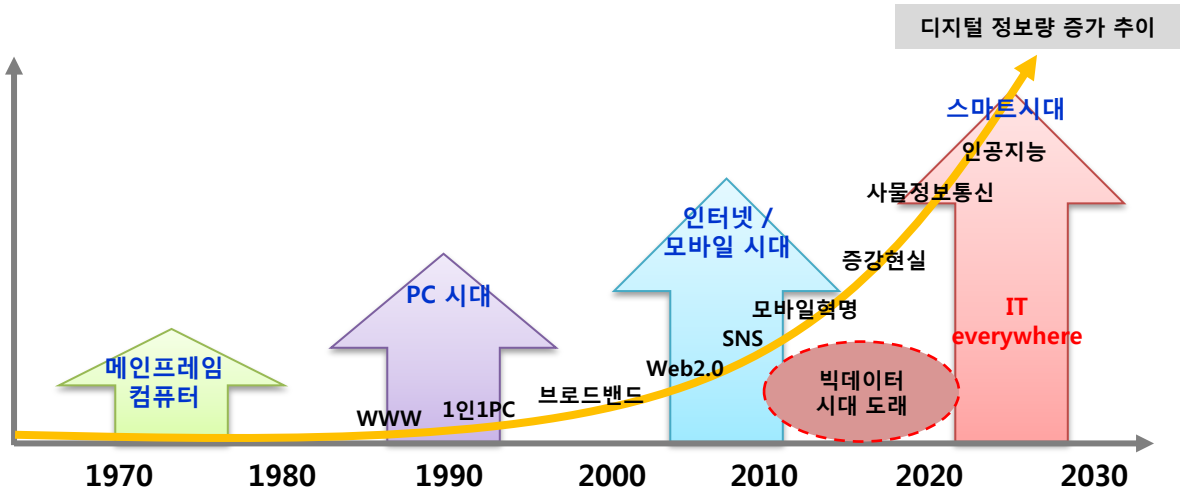
# 빅데이터 이해와 동향



한국기술교육대학교  
온라인평생교육원

## 빅데이터 시대의 도래와 빅데이터 가치

### 1. 빅데이터 시대의 도래



<정지선, 빅 데이터의 새로운 가능성과 대응전략, 한국정보화진흥원, 2012>

	메인프레임 컴퓨터		PC 시대		인터넷/모바일 시대		스마트 시대	
	1970	1980	1990	2000	2010	2020	2030	
데이터 규모		EB(Exa Byte) (90년대 말 = 100EB)		→	ZB(Zetta Byte)진입 (2011년 = 1.8ZB)		→	ZB 본격화 시대 (‘20년 = ‘11년 대비 50배 ↑)
데이터 유형		정형 데이터 (데이터베이스, 사무정보)		→	비정형 데이터 (이메일, 멀티미디어, SNS)		→	사물정보, 인지정보 (RFID, Sensor, 사물통신)
데이터 특성		구조화		→	다양성, 복합성, 소셜		→	현실성, 실시간성

<정지선, 빅 데이터의 새로운 가능성과 대응전략, 한국정보화진흥원, 2012>



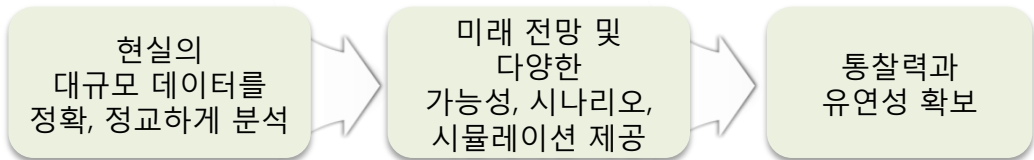
## ■ 빅데이터 시대의 도래와 빅데이터 가치

### 2. 빅데이터의 가치

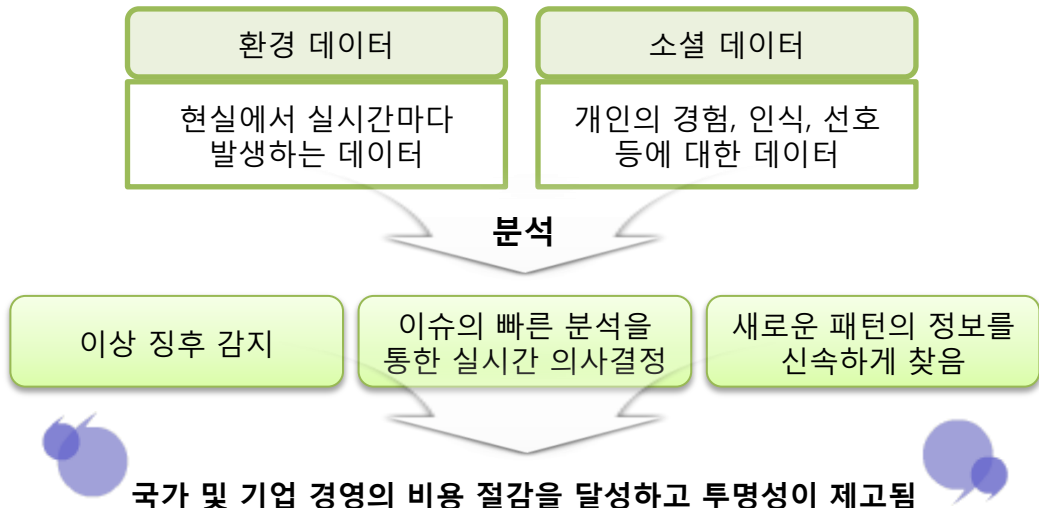
1 경제적 자산이 되고, 가치창출의 원천이 됨

- 새로운 기회를 창출하고, 위험을 해결하여 사회 및 경제 발전의 엔진 역할을 수행할 수 있음

2 불확실한 미래에 대한 통찰력을 제공할 수 있음



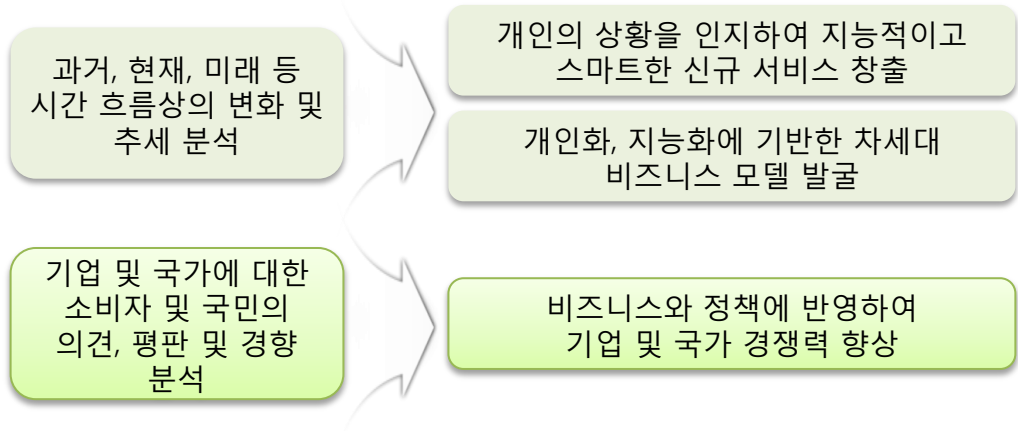
3 미래의 리스크에 대응력을 키울 수 있음



## ■ 빅데이터 시대의 도래와 빅데이터 가치

### 2. 빅데이터의 가치

#### 4 미래 사회의 스마트한 경쟁력을 제고할 수 있음



#### 5 스마트한 창조력을 제고할 수 있음

- 이종데이터들을 상호 결합하여 새로운 지식을 발견함

다른 분야에서  
발생하여 서로 다른  
데이터들

- 많은 요인간의 상관관계를 파악하여 결과에 대한 시행착오를 최소화함
- 방대한 데이터 분석을 통해 新융합 시장을 창출함

## ■ 빅데이터 시대의 도래와 빅데이터 가치

### 2. 빅데이터의 가치

#### 빅데이터의 사회 · 경제적 가치

구분	기관명	주요내용
산업 경제성	Economist (2010)	• 데이터는 자본이나 노동력과 거의 동등한 레벨의 경제적 투입 자본, <b>비즈니스의 새로운 원자재</b> 역할
	Garter (2011)	• 데이터는 <b>21세기 원유</b> , 데이터가 미래 경쟁 우위를 좌우 • 기업은 다가올 ' <b>데이터 경제 시대</b> '를 이해하고 정보 고립(Information Silo)을 경계해야 성공 가능
	Mckinsey (2011)	• 빅데이터는 <b>혁신, 경쟁력, 생산성의 핵심 요소</b> • 의료, 공공행정 등 5대 분야에서 6천억 불 이상의 가치 창출
국가 경쟁력	美 대통령 과학기술자문위	• 미국 정부기관들이 <b>데이터를 지식으로, 지식을 행동으로</b> 변환하는 전략에 집중해야 함을 주장
	싱가포르	• 데이터를 기반으로 싱가포르를 위협하는 <b>리스크</b> 에 대한 평가와 환경변화를 탐지

<김현곤, 빅데이터 시대 전망과 대응전략, 한국정보화진흥원, 2012>

실무에 적용 가능한 Big Data 분석 개론

---

# 빅데이터 이해와 동향



한국기술교육대학교  
온라인평생교육원

## ■ 빅데이터 활용

### 1. 주요국가의 빅데이터 활용

#### 미국



모든 연방정부기관에  
빅데이터 전략이 필요하다.

2012년 대통령의 과학기술정책실에서 6개의 주요 연방정부기관이 협력함



빅데이터 관련 R&D에 2억 달러를 투입하기로 결정함

#### 국립과학재단(NSF)



- 데이터 관리, 분석, 시각화, 지식 추출 등을 위한 핵심 연구/기술 개발 지원 및 인력양성
- 빅데이터를 이용한 정보 및 지식 추출 방법론을 위한 포괄적이고 장기적인 전략 수립에 투자

#### 국립보건원(NIH)



- 생물, 건강 및 질병에 대한 데이터 확보 및 공개를 통한 빅데이터 이용 확산

## ■ 빅데이터 활용

### 1. 주요국가의 빅데이터 활용

#### 미국

##### 국방부(DoD)



- 데이터를 이용한 국방 분야 의사결정 체계 구축

##### 방위고등연구계획국(DARPA)



- 대용량 데이터를 분석하는 컴퓨팅 기술 및 SW 개발
- 다양한 미션에 쉽게 적용 가능한 HCI tool 개발

##### 에너지부(DoE)



- 고등 컴퓨팅(Advanced computing) 기술 활용을 통한 에너지 과학 발견

##### 지질조사원(USGS)



- 지구 시스템 과학을 위한 빅데이터 기술 개발

## ■ 빅데이터 활용

### 1. 주요국가의 빅데이터 활용

#### 미국

##### 美 의료부문의 빅데이터 활용

1 연간 \$3,300억의 직간접적 비용 절감 효과를 기대하고 있음

2 의료산업 분야의 개선방법 예시

R&D	• 빅데이터는 통계활용, 모델링을 가능하게 하여 \$250억 비용감소와 \$1,000억의 산업효과를 유발함
임상분야	• 의료기관 별 진료방법, 효능, 비용 데이터를 분석하여 보다 효과적인 진료방법을 파악함
가격 / 회계	• 투약효능을 기록한 데이터에 근거하여 가격을 결정함 • 환자는 고비용 저효능의 약물을 구매하여 리스크를 감소시킴
New BM	• 환자 데이터의 온라인 플랫폼화 - 의료협회 간 데이터 공유로 치료 효과 제고
공중보건	• 전국의 의료데이터를 연계하여 전염병 발생과 같은 긴박한 순간에 빠른 의사결정을 함

<McKinsey, Big Data : The Next Frontier for Innovation, Competition, and Productivity, 2011>

#### 영국



과학부 장관 및 내각사무처 장관이  
공동책임으로 공공 데이터 그룹과  
협력하여 공공 데이터에 대한  
공개 및 접근에 대해 개선해야 한다.

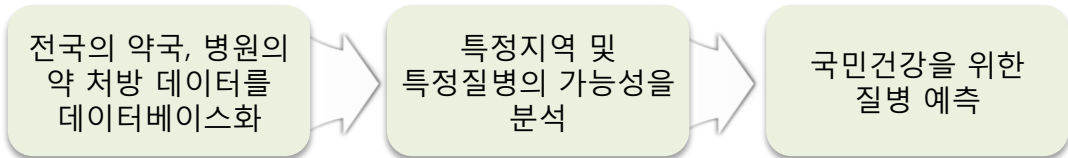
사회 및 경제적인 이익을 위해 빅데이터의 활용을 극대화하려고 노력함

## ■ 빅데이터 활용

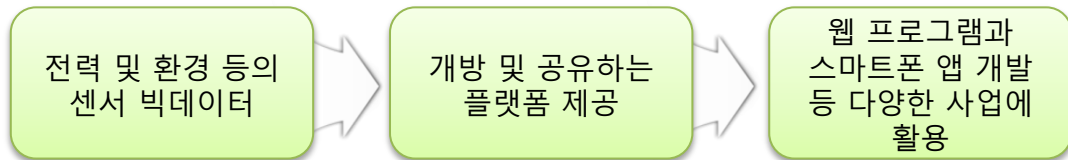
### 1. 주요국가의 빅데이터 활용

#### 영국

##### 국영 보건복지 체제(NHS)



##### 패치베이(Pachube)



패치베이의 데이터 공유 화면

<김현곤, 빅데이터 시대 전망과 대응전략, 한국정보화진흥원, 2012>



## ■ 빅데이터 활용

### 1. 주요국가의 빅데이터 활용

#### 한국



빅데이터를 활용하여 지식과 정보를  
개방하고 상호 협력할 수 있는  
스마트한 정부를 구현한다.

정부 및 공공기관, 민간 전문기업,  
연구기관, 빅데이터 보유기관

- 1 빅데이터 국가전략 포럼을 통해 여러 기관 및 전문가간의 협력을 활성화 함
- 2 공공부문의 빅데이터 활용의 성공사례를 조기 발굴하려 노력함
- 3 공공 및 민간 데이터의 연계 활용토록 유도하고 있음
- 4 범정부적인 빅데이터 전략로드맵을 수립하고 빅데이터 분석 전문인력 양성에 노력함

## ■ 빅데이터 활용

### 2. 빅데이터 활용 분야

#### 2011년 맥킨지가 제시한 빅데이터 활용분야

활용분야	분석 대상 데이터
의료 산업	<ul style="list-style-type: none"><li>• 제약사 연구 개발 데이터</li><li>• 환자 치료·임상 데이터</li><li>• 의료 산업의 비용 데이터</li></ul>
공공 행정	<ul style="list-style-type: none"><li>• 정부의 행정 업무에서 발생하는 데이터</li></ul>
소매업	<ul style="list-style-type: none"><li>• 고객의 거래 데이터</li><li>• 구매 경향</li></ul>
제조업	<ul style="list-style-type: none"><li>• 고객 취향 데이터</li><li>• 수요 예측 데이터</li><li>• 제조 과정 데이터</li><li>• 센서 활용 데이터</li></ul>
개인 위치 데이터	<ul style="list-style-type: none"><li>• 개인과 차량의 위치 데이터</li></ul>

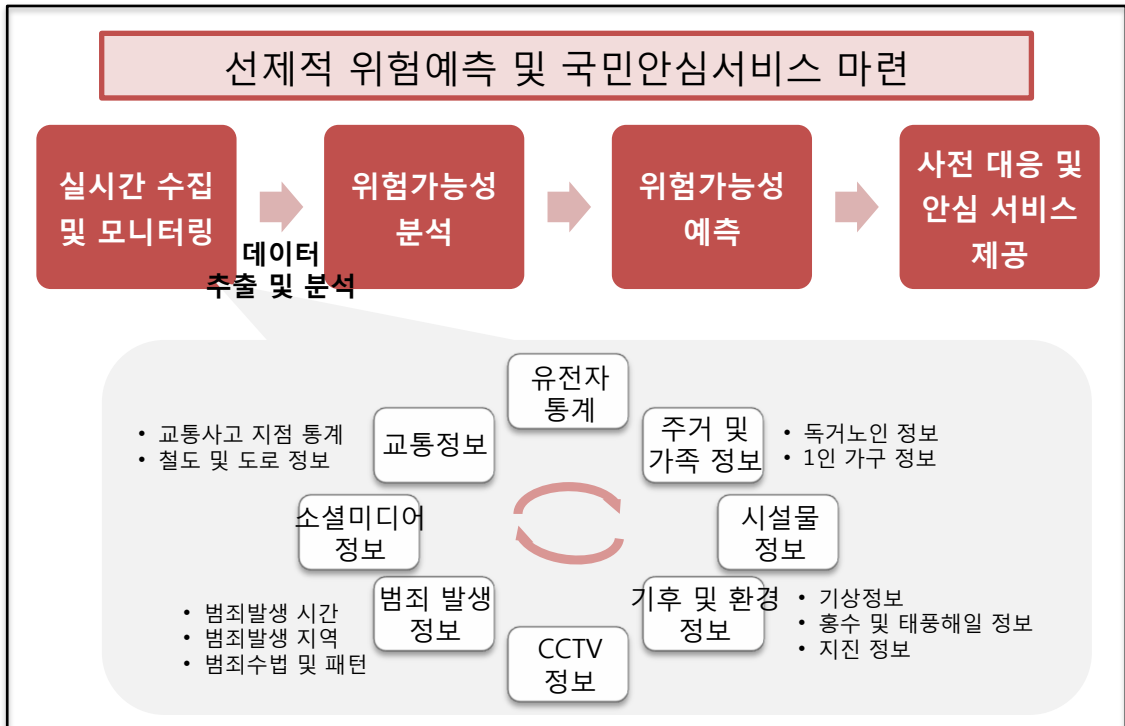
<McKinsey, Big Data : The Next Frontier for Innovation, Competition, and Productivity, 2011 재구성>

## ■ 빅데이터 활용

### 2. 빅데이터 활용 분야

#### 공공분야

➔ 안전한 사회 구축을 위한 빅데이터 활용



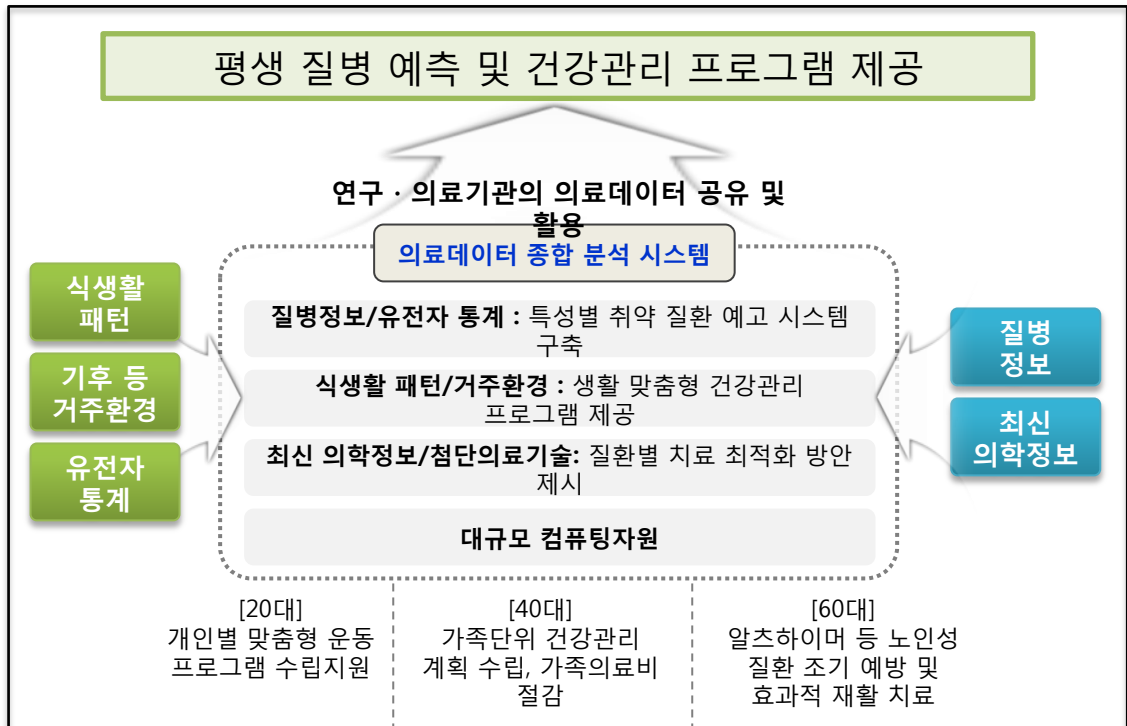
<김현곤, 빅데이터 시대 전망과 대응전략, 한국정보화진흥원, 2012>

## ■ 빅데이터 활용

### 2. 빅데이터 활용 분야

#### 의료분야

➔ 건강한 사회 구축을 위한 빅데이터 활용



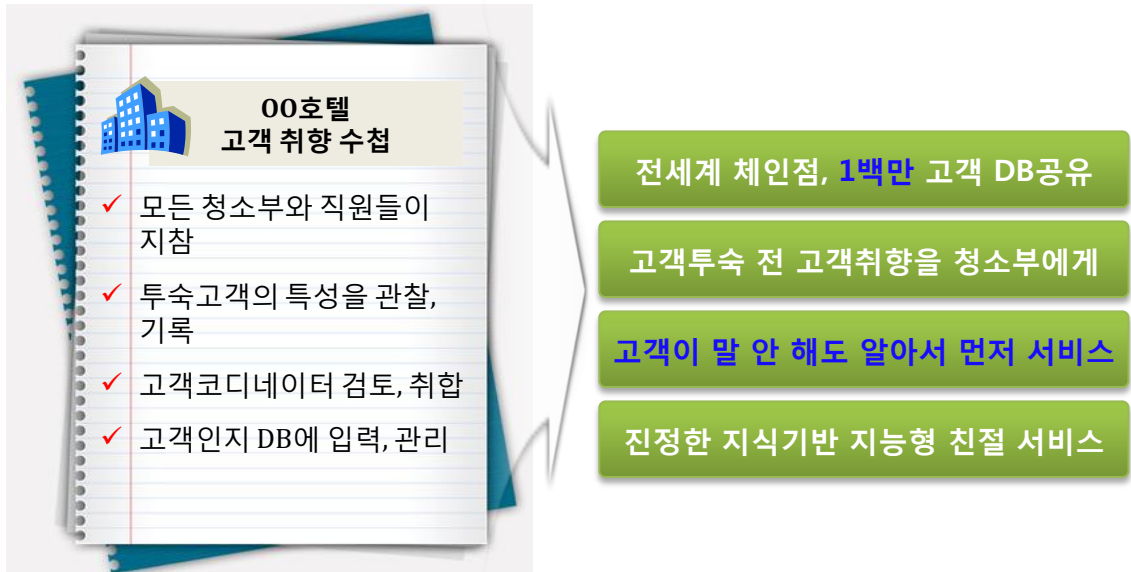
<김현곤, 빅데이터 시대 전망과 대응전략, 한국정보화진흥원, 2012>

## ■ 빅데이터 활용

### 2. 빅데이터 활용 분야

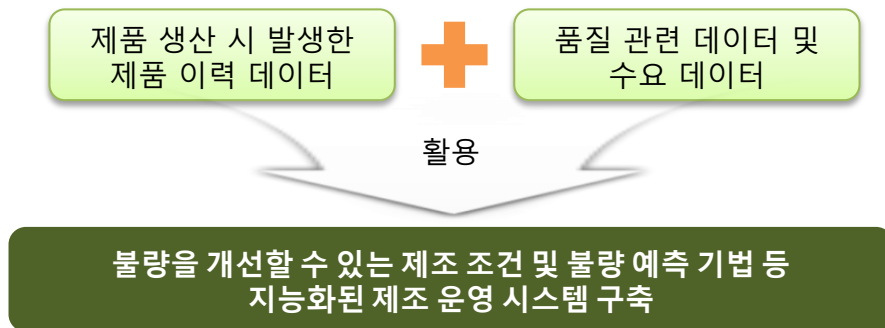
#### 소매 또는 유통 분야

➔ 호텔에서의 빅데이터 활용



<KBS신년기획, 무엇이 세계경제를 움직이는가? 제2편 신성장동력 : 서비스를 설계하라, 2008>

#### 제조분야

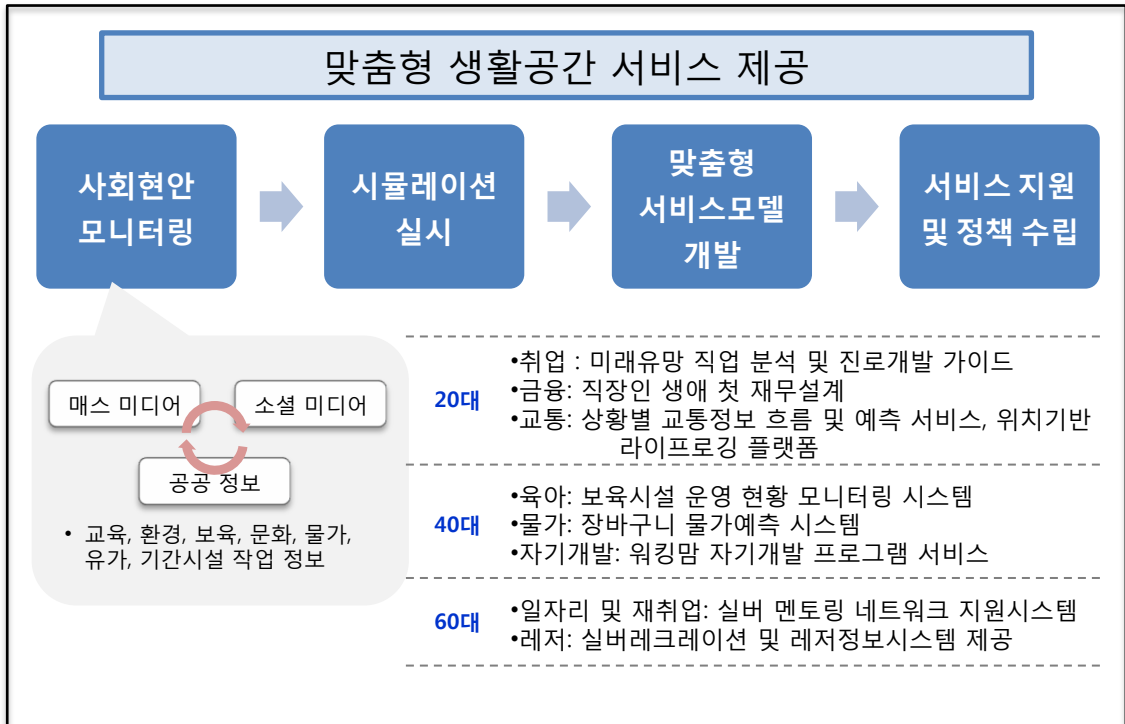


## ■ 빅데이터 활용

### 2. 빅데이터 활용 분야

#### 정보 서비스 분야

➡ 행복한 사회 구축을 위한 빅데이터 활용



<김현곤, 빅데이터 시대 전망과 대응전략, 한국정보화진흥원, 2012>