

빅데이터의 이해와 활용

Understanding and Using Big Data

13

산업체 및 공공기관 빅데이터 적용 사례

학습 내용

- 01 주요 산업 빅데이터 적용 사례
- 02 공공기관 빅데이터 분석시스템 사례

학습 목표

- 업무 도메인별 빅데이터 분석업무의 특징을 이해할 수 있다.
- 빅데이터 분석시스템 구축을 위한 업무 이해를 적용할 수 있다.
- 빅데이터 시스템 아키텍처, 프로세스, 구축 과정을 이해할 수 있다.
- 공공기관의 대국민 서비스 업무의 구성 현황을 이해할 수 있다.
- 공공기관의 빅데이터 분석시스템의 주요 구성요소를 설명할 수 있다.

주요 용어

스마트 팩토리

- 설계 및 개발, 제조 및 유통 등 생산과정에 디지털 자동화 솔루션이 결합된 정보통신기술(ICT)을 적용하여 생산성, 품질, 고객만족도를 향상시키는 지능형 생산공장
- 공장 내 설비와 기계에 사물인터넷(IoT)을 설치하여 공정 데이터를 실시간으로 수집하고, 이를 분석해 스스로 제어할 수 있게 만든 미래의 공장

〈출처 : 위키백과〉

주요 용어

BI(비즈니스 인텔리전스)

- 비즈니스 인텔리전스(Business Intelligence, BI)는 기업에서 데이터를 수집, 정리, 분석하고 활용하여 효율적인 의사결정을 할 수 있는 방법에 대해 연구하는 학문
- 기업의 비전을 달성하기 위하여 비즈니스의 전략을 효율적이고 효과적으로 지원하여 각 조직의 구성원(종업원, 중간 관리자, 의사결정자 등)에게 적시에 의사결정을 할 수 있도록 지원하는 정보체계

〈출처 : 위키백과〉

주요 용어

ERP(전사적 자원관리)

- 전사적 자원 관리(Enterprise Resource Planning : ERP)는 경영 정보 시스템(MIS)의 한 종류
- 회사의 모든 정보 뿐만 아니라, 공급 사슬관리, 고객의 주문정보까지 포함하여 통합적으로 관리하는 시스템
- ERP를 도입 시 생산부분이 마케팅을 실시간으로 조회하여 생산일정을 조회 및 변경할 수 있는 등 비용낭비나 생산 지연 요인을 사전에 제거하는 일이 가능함

〈출처 : 위키백과〉

사전 퀴즈

01

기본통계를 확인하고 데이터 분포를 분석하고
변수 간의 관계를 확인하는 업무는
데이터 분석업무 단계 중 “빅데이터 처리”
과정에서 이루어진다.

Q1

Q2

Q3

O

X

사전 퀴즈

01

기본통계를 확인하고 데이터 분포를 분석하고
변수 간의 관계를 확인하는 업무는
데이터 분석업무 단계 중 “빅데이터 처리”
과정에서 이루어진다.

Q1

Q2

Q3

O

X

정답

X

해설

기본통계를 확인하고 데이터 분포를 분석하고
변수 간의 관계를 확인하는 업무는,
데이터 분석업무 단계 중, "데이터 탐색" 과정에서
이루어집니다.

사전 퀴즈

02

빅데이터의 처리 과정은 데이터 생성, 수집,
저장, 처리, 분석, 표현(시각화) 으로
구성되어 있습니다.

Q1

Q2

Q3

사전 퀴즈

02

Q1

Q2

Q3

빅데이터의 처리 과정은 데이터 생성, 수집, 저장, 처리, 분석, 표현(시각화) 으로 구성되어 있습니다.



정답

O

해설

- 빅데이터의 처리 과정은 데이터 생성, 수집, 저장, 처리, 분석, 표현(시각화) 으로 구성되어 있습니다.
- 내부 및 외부 데이터가 생성되고, 수집 기술에는 크롤링, ETL 등이 있으며, 처리 기술에는 맵리듀스, 프로세싱 등이 있으며, 분석 기술에는 자연어 처리, 기계학습 등이 있습니다.

사전 퀴즈

03

Q1

Q2

Q3

데이터 분석을 위한 대용량 처리 시스템은 다음 단계를 거쳐 진행한다.

- 기간계 시스템: 데이터 입출력
- 데이터 연결: 복사, 이전, 변환, 로딩
- 정보계 시스템: 데이터 집계 -> 데이터 마트 -> 시각화

사전 퀴즈

03

Q1

Q2

Q3

데이터 분석을 위한 대용량 처리 시스템은 다음 단계를 거쳐 진행한다.

- 기간계 시스템: 데이터 입출력
- 데이터 연결: 복사, 이전, 변환, 로딩
- 정보계 시스템: 데이터 집계 -> 데이터 마트 -> 시각화

정답

O

해설

- 기간계 시스템에서 정보계 시스템으로 데이터를 옮기는 방법은 다양합니다.

데이터에 대한 추출, 복사, 이전, 이행, 변환, 로딩 및 적용 등의 방법을 사용합니다.

- 정보계 시스템에서 데이터의 집계한 이후에, 시각화로 가는 길목에, 데이터 마트 접근법이 있습니다.

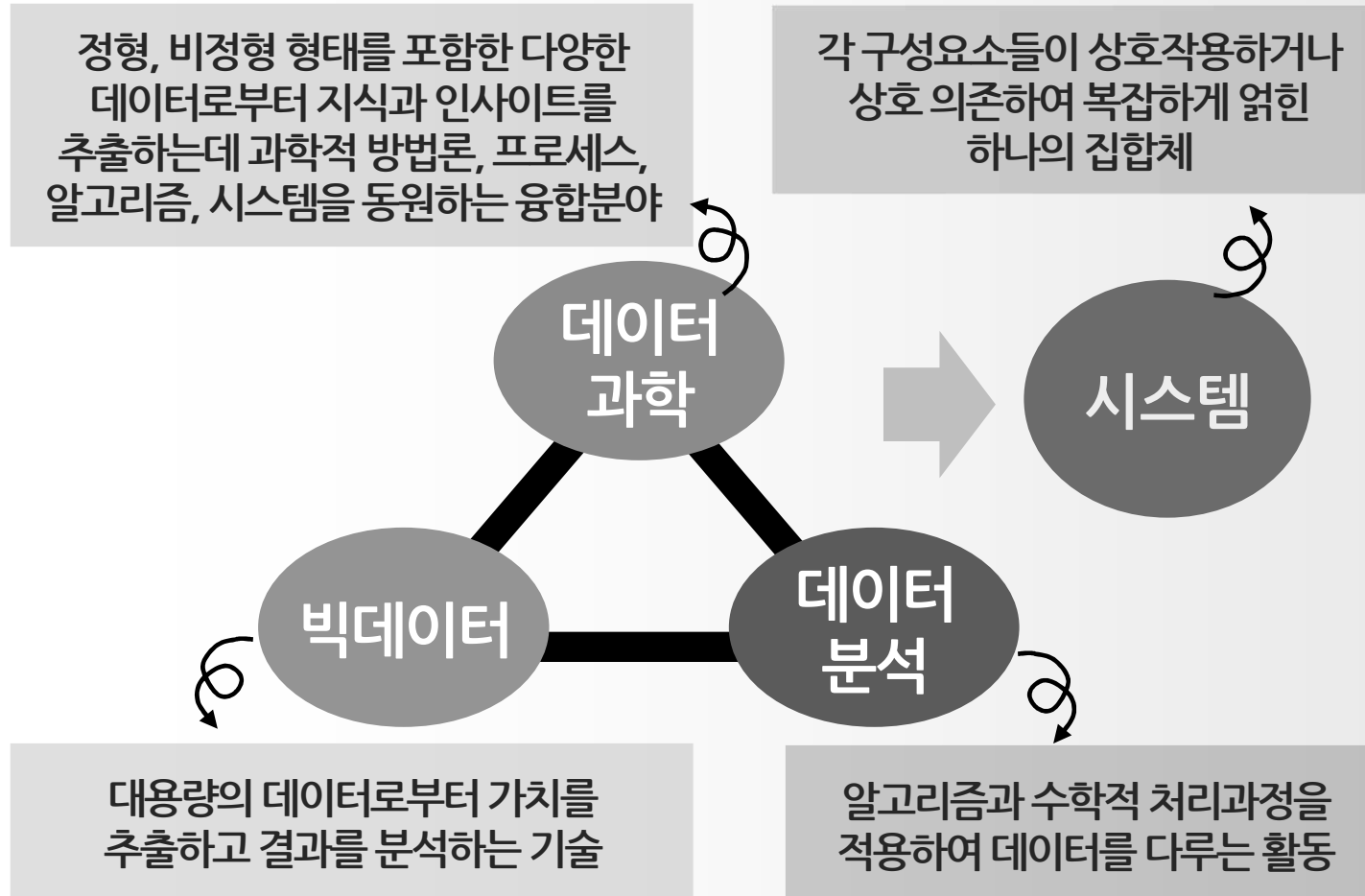


01 주요 산업 빅데이터 적용 사례

-
- 1) 빅데이터 분석시스템 이해
 - 2) 산업별 빅데이터 시스템 적용 사례

1) 빅데이터 분석시스템 이해

● 빅데이터 분석시스템 의미



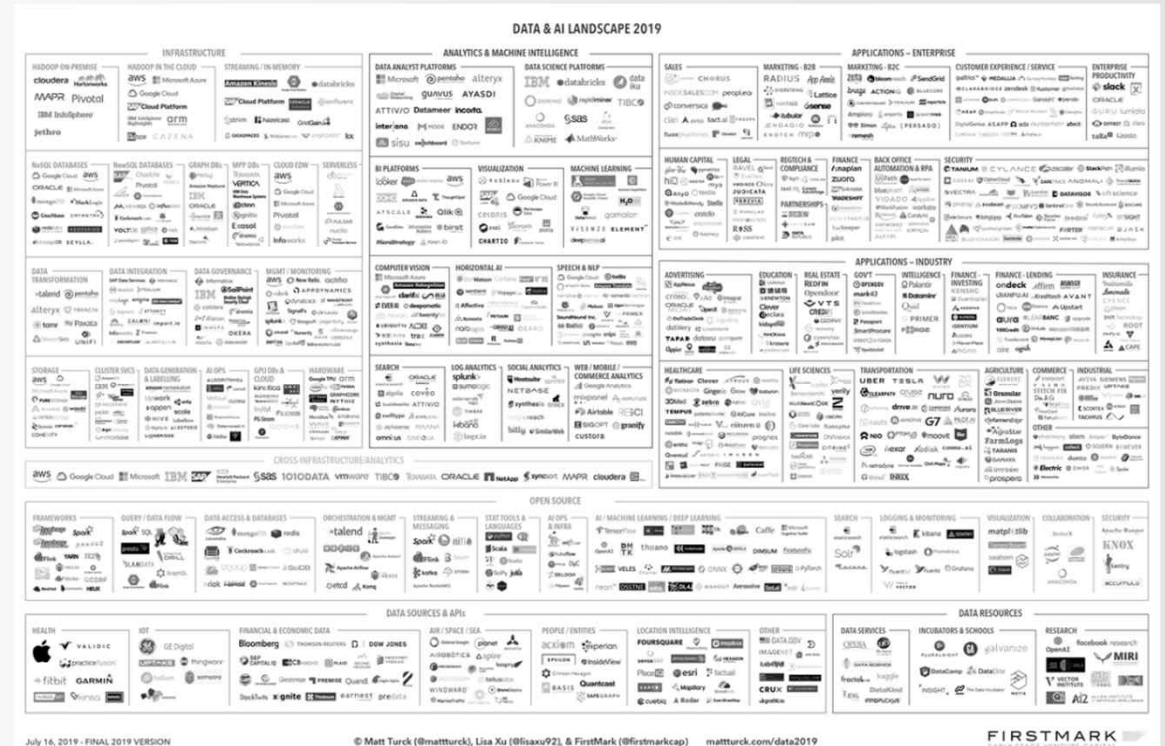
1) 빅데이터 분석시스템 이해

● 진화하는 빅데이터 분석시스템

빅데이터는...

데이터만을 가리키지 않음

관련된 기술을 의미하지 않음



끊임없이 문제해결을 추구하는데 그 의의가 있음

1) 빅데이터 분석시스템 이해

● 진화하는 빅데이터 분석시스템 - cont.

{ 빅데이터 구성 }

분석

- 데이터 분석 플랫폼
- 데이터 사이언스 플랫폼
- BI 플랫폼
- 시각화
- 머신 러닝
- 스피치 & NLP
- 검색
- 로그분석
- 소셜분석
- 웹/모바일/커머스 분석

인프라스트럭처

애플리케이션

- Enterprise
- Industry

오픈소스

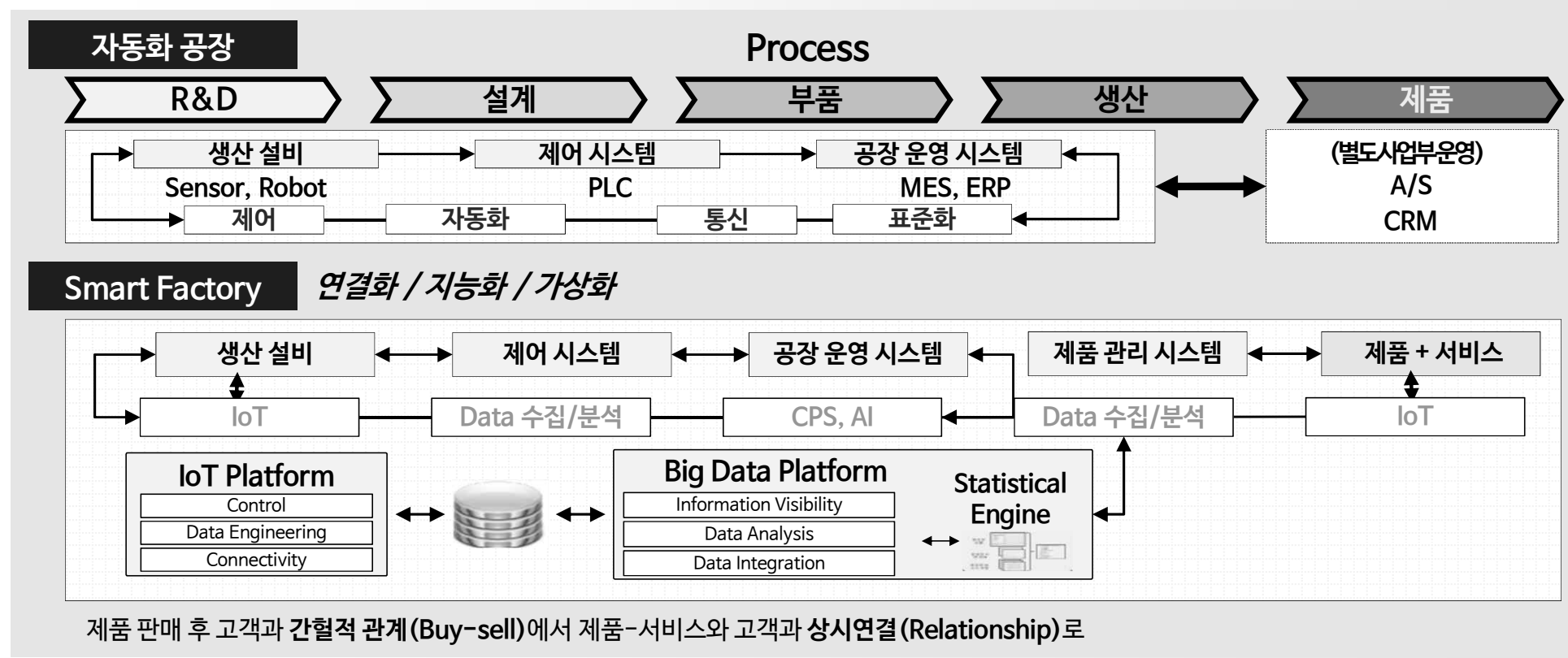
데이터 소스 및 API

데이터 리소스

2) 산업별 빅데이터 시스템 적용 사례

- 빅데이터 활용 제조공정 개선시스템(예시)

- ▶ 자동화 공장과 스마트 팩토리의 차이



〈출처 : 창조경제연구회(KCERN)〉

2) 산업별 빅데이터 시스템 적용 사례

- 빅데이터 활용 제조공정 개선시스템(예시)

- ▶ 재조공정 분야의 빅데이터 분석

{ 4차 산업혁명 관점 }

인더스트리 4.0

- 사물인터넷, 빅데이터 등 IT와 물리분야 기술을 활용한 생산방식의 전면적 개편
- 사이버물리시스템에 의한 패러다임 변화



2) 산업별 빅데이터 시스템 적용 사례

- 빅데이터 활용 제조공정 개선시스템(예시)

- ▶ 제조공정 품질불량 분석

(테스팅 결과 분석에 따른 개선사항 발굴)

{ MES 도전과제 }

프로세스 제약 제어
(Process Limit Control)

오류탐지와 분류
(Fault Detection)

손실발생에 대한
예측 (Prediction)

Historical
데이터 분석

실시간 작업 분배
(Real Time Dispatching)

〈출처 : MES 빅데이터 분석(활용)을 통한 생산성 및 품질향상 사례, 김용한, 2013〉

2) 산업별 빅데이터 시스템 적용 사례

- 빅데이터 활용 제조공정 개선시스템(예시)

- ▶ MES 최적화 시뮬레이션 분석

(설비관리의 상태분석, 공정재고의 최소화)



공정 실시간 감시, 제품 불량 분석, 이상점 Detection
처리를 위한 설비 관련 데이터 분석



생산관리 및 6시그마 관점의 과학적 분석기법의
테스트 및 발굴

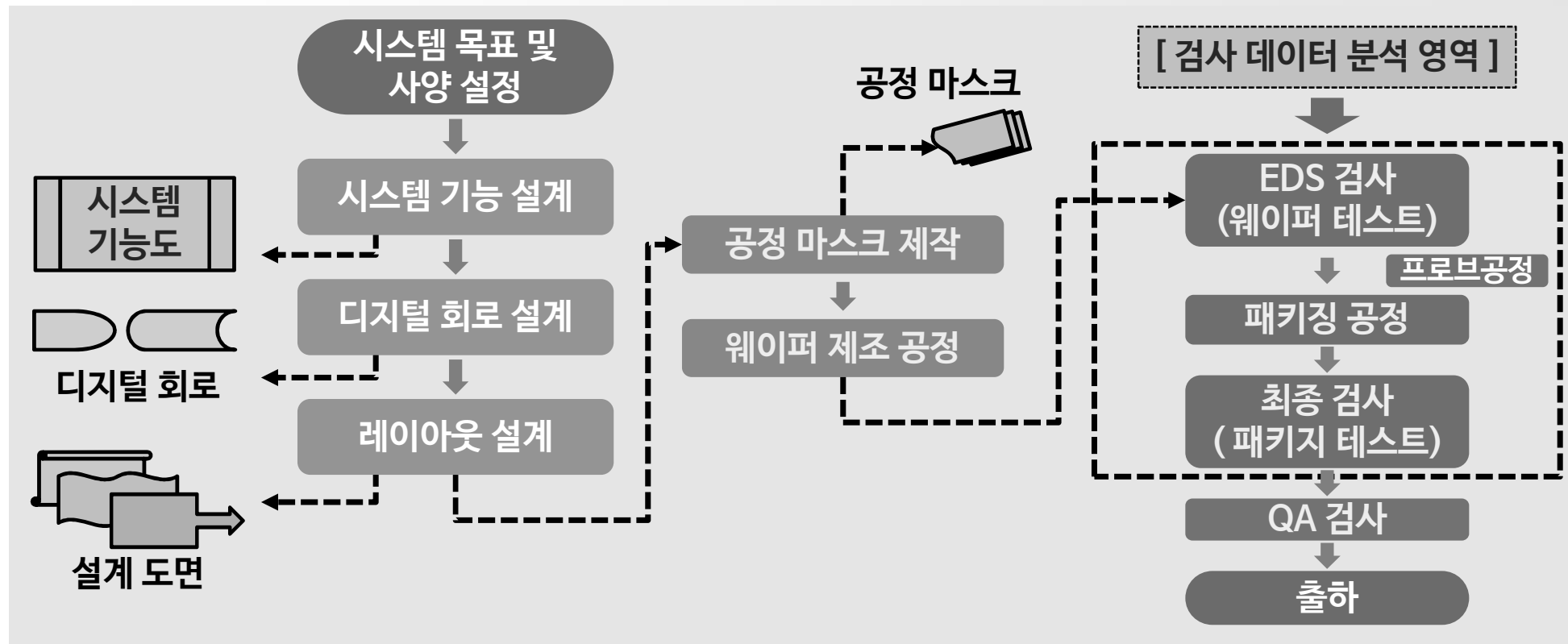


기존의 10억 건 정도의 표본추출에서 수 천억 건
이상의 전수조사를 통한 혁신적인 분석 방향 제시

2) 산업별 빅데이터 시스템 적용 사례

● 빅데이터 활용 제조공정 개선시스템(예시)

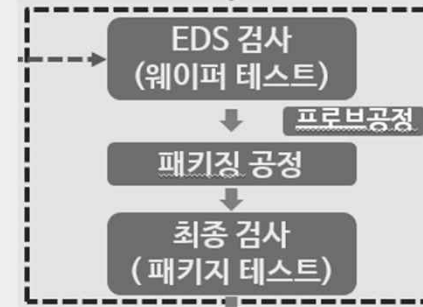
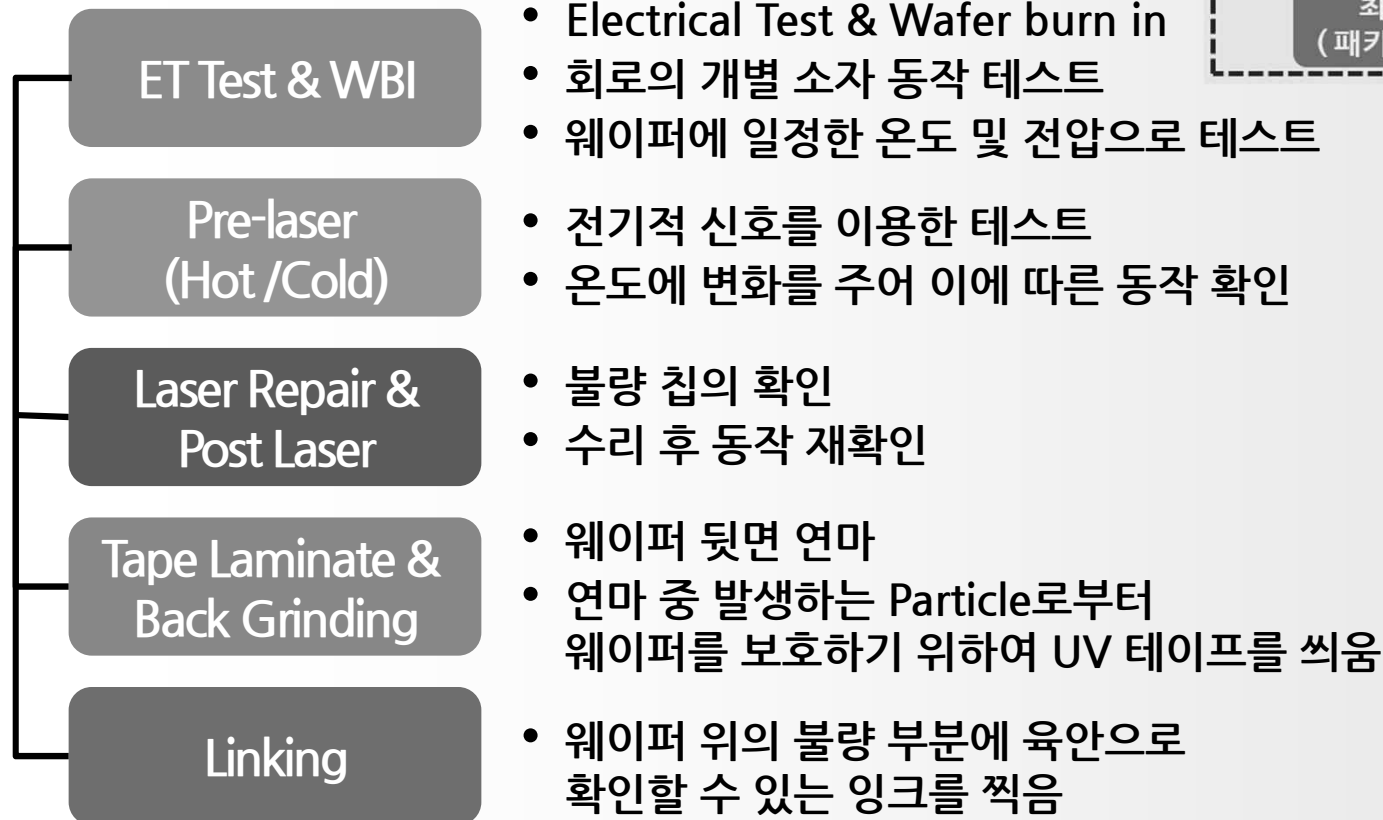
▶ 반도체 제조과정 흐름도



2) 산업별 빅데이터 시스템 적용 사례

● 빅데이터 활용 제조공정 개선시스템(예시)

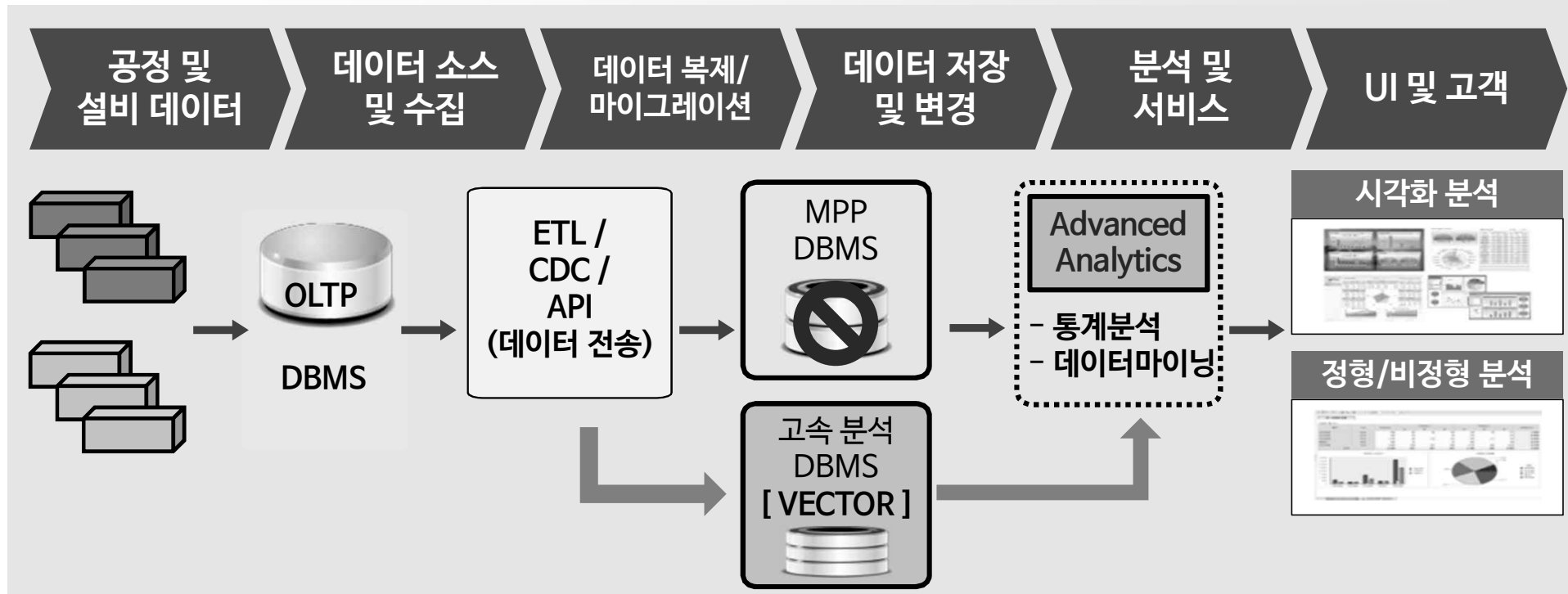
▶ 반도체 제조과정 흐름도 : 프로브공정



2) 산업별 빅데이터 시스템 적용 사례

- 빅데이터 활용 제조공정 개선시스템(예시)

- ▶ 제조공정 분야의 빅데이터 분석 플랫폼



2) 산업별 빅데이터 시스템 적용 사례

- 금융업무의 고객 정보 및 리스크 관리 이해

“ 데이터의 수집과 정제 이후에,
데이터와 정보 분석으로 고도화하기 ”



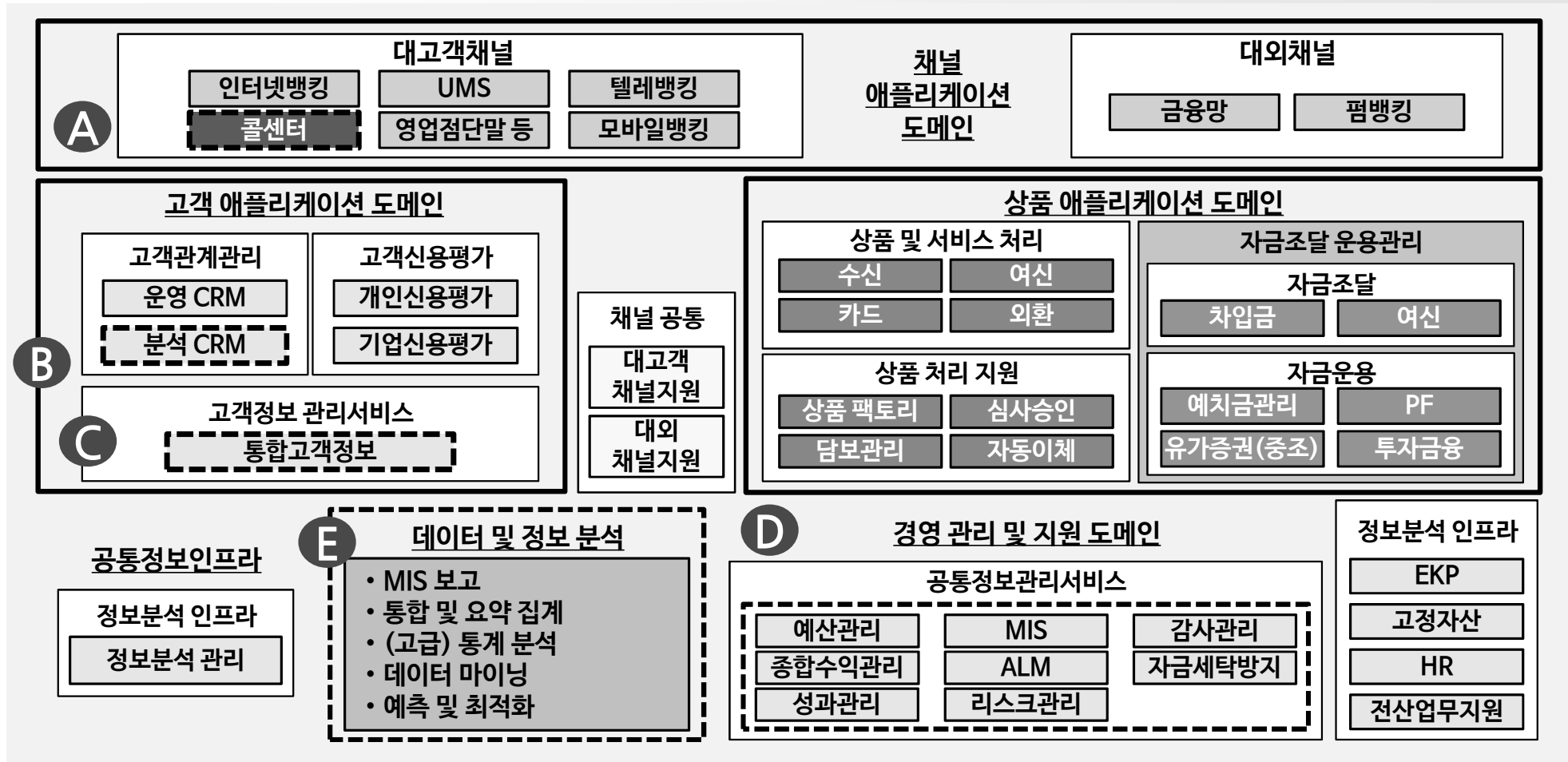
데이터 정합성, 데이터 추적성 관점의 리스크 관리



고객 정보 수집 및 관리 관점의 위험 요소 확인사례

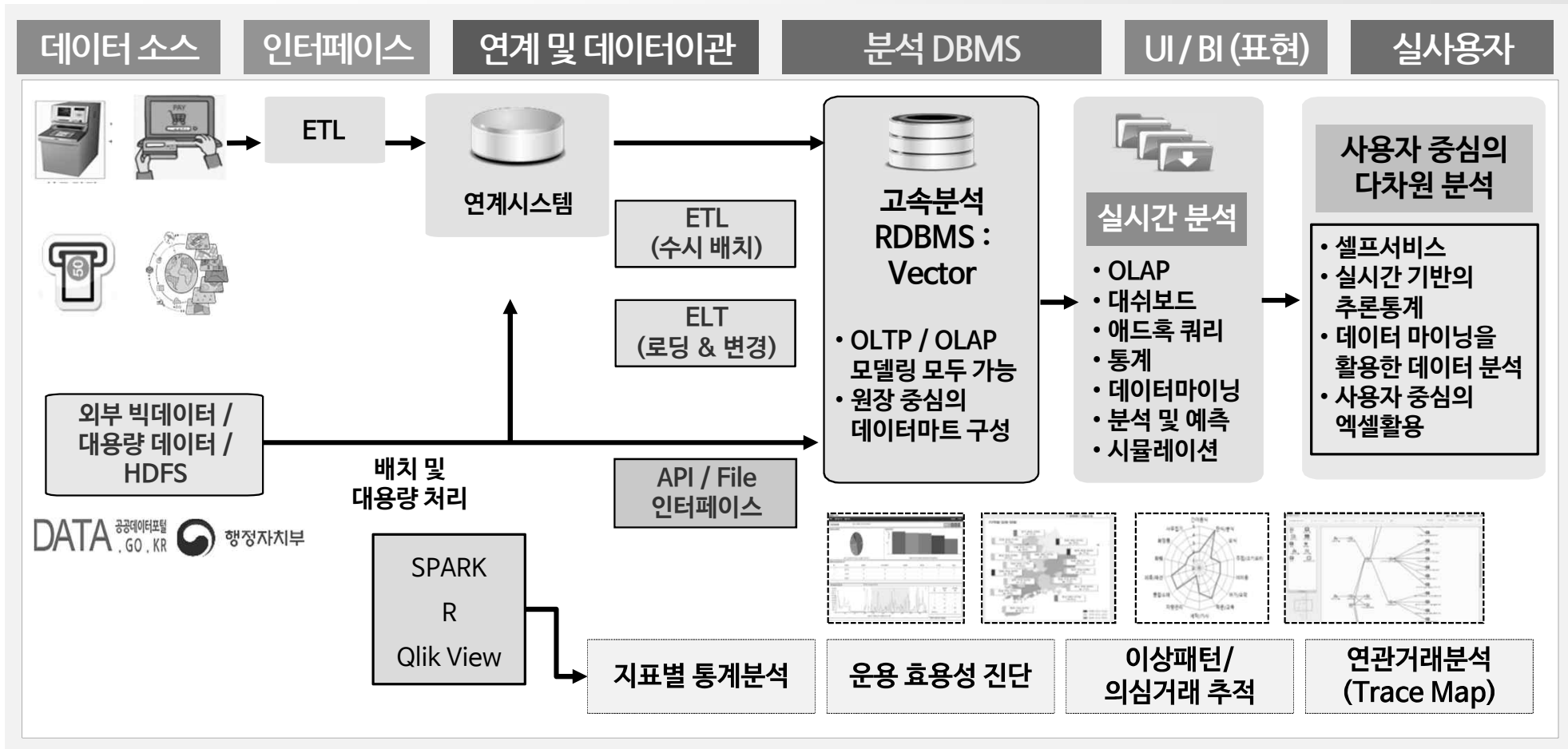
2) 산업별 빅데이터 시스템 적용 사례

● 금융업무 고객 정보 및 리스크 관리 이해(예시)



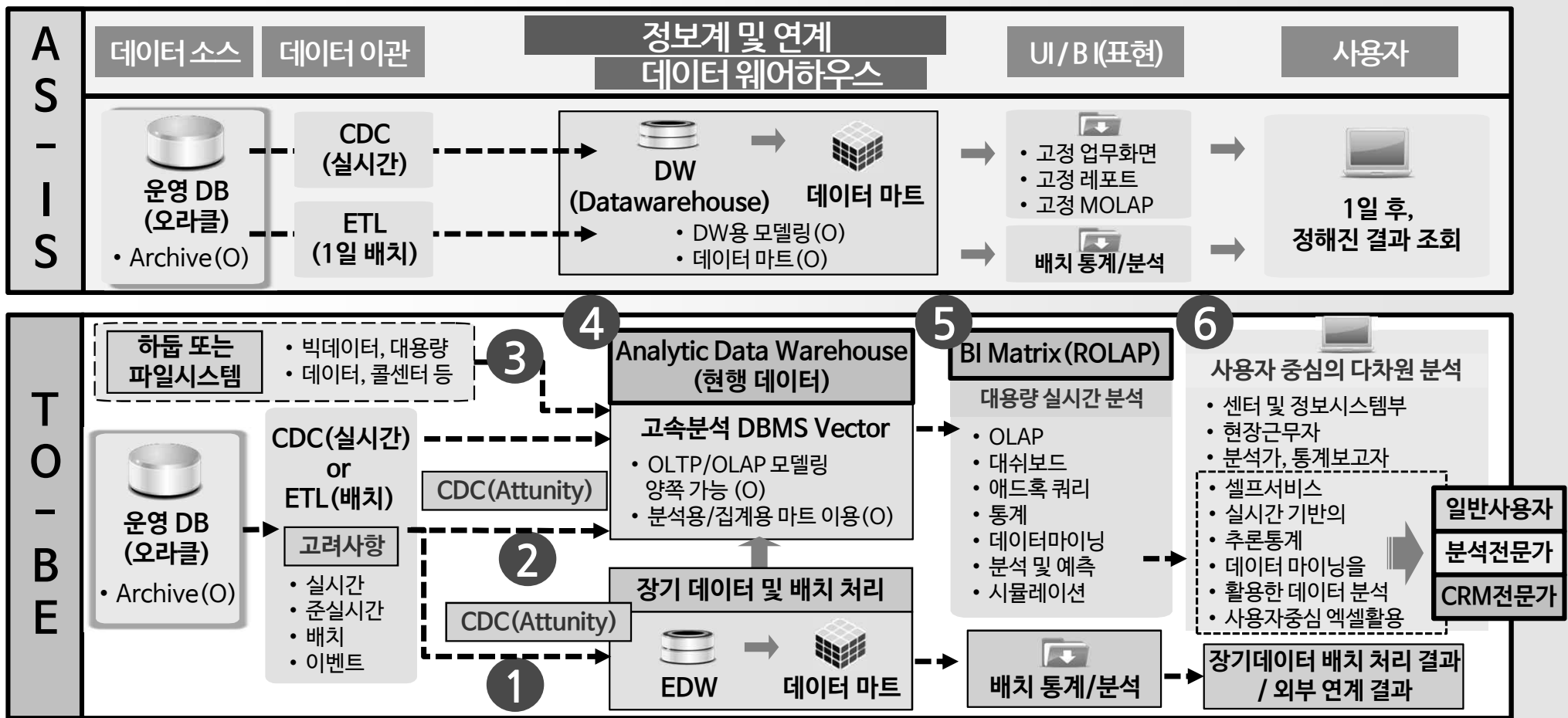
2) 산업별 빅데이터 시스템 적용 사례

● 금융업무 사기탐지 시스템(예시)



2) 산업별 빅데이터 시스템 적용 사례

● 금융업무 정보계 시스템을 위한 대용량 고속 분석시스템(예시)



2) 산업별 빅데이터 시스템 적용 사례

● 프로세스 개선과 SW 안전(예시)

01 분석 업무 중심의 프로세스 단순화

- 상위 의사결정자, 중간보고자, 현장업무자 간 동일한 형태의 데이터를 공유하는 환경이 단순화되고 편리성 증가
- 일원화된 분석 결과 공유를 통한 의사결정구조의 신속한 판단

2) 산업별 빅데이터 시스템 적용 사례

- 프로세스 개선과 SW 안전(예시) - cont.

02 실시간으로 현장담당자의 데이터 획득 및 분석 능력 강화

- 현장업무자의 입장에서 본인이 원하는 데이터의 통계, 분석결과를 본인이 직접 작성하고 확인하기
- 업무개선 능력과 효과는 지시사항에서 발생하는 지침과 방침이 아닌, 현장중심의 시각과 결과를 중심으로 보완, 개선 및 지속되어야 함

2) 산업별 빅데이터 시스템 적용 사례

● 프로세스 개선과 SW 안전(예시) - cont.

03

데이터 공유를 통한 고객 케어 시간 단축 및 고객 대응 비용 감소

- Analytic Datawarehouse를 통한 통계업무, 분석업무, 정기 레포팅 업무 등을 순차적으로 적용할 수 있음
- 업무 분석 능력은 분석 중심의 고도화된 시스템을 지속적으로 활용함으로써 성공할 수 있음

2) 산업별 빅데이터 시스템 적용 사례

- 프로세스 개선과 SW 안전(예시) - cont.

- ▶ SW 안전을 고려한 데이터분석과 프로세스 개선

{ 프라이버시 / 보안기술 고려사항 }

01 빅데이터에서의 보안 역할 및 책임

02 모바일 인터넷 서비스에서의 보안 위협

03 서비스로서 빅데이터 환경에서의 보안 역할,
책임 및 대책

2) 산업별 빅데이터 시스템 적용 사례

- 프로세스 개선과 SW 안전(예시) - cont.

- ▶ SW 안전을 고려한 데이터분석과 프로세스 개선

{ 안전 수명주기 구성 }



기능안전표준 IEC61508과 SW 수명주기
표준 ISO12207을 참고하여 안전수명주기 구성

〈출처 : "ICT 표준화전략맵 Ver.2019", TTA, 2018〉

〈출처 : SW안정성 공통개발가이드", NIPA, 2016〉

2) 산업별 빅데이터 시스템 적용 사례

- 프로세스 개선과 SW 안전(예시) - cont.

- ▶ SW 안전을 고려한 데이터분석과 프로세스 개선

{ 안전 수명주기 구성 }

안전 수명주기 구성

- Analysis 부문
(위험분석 단계)
- Realization 부문(우측 참고)
- Operation 부문
(SW 운영 단계)

Realization 부문

- SW 계획 단계
- SW 요구 분석
- SW 설계 단계
- SW 구현 단계
- SW 통합 단계
- SW 확인 검증 단계

<출처 : SW안정성 공통개발가이드", NIPA, 2016>

The background features a dark, abstract design with glowing white lines and geometric shapes. A prominent white hexagon is located in the upper center, with lines radiating from it. To the left, there are several overlapping, glowing white diamond shapes. The overall aesthetic is futuristic and technological.

02

공공기관 빅데이터 분석시스템 사례

- 1) 자동차보험 교통사고 다발생 환자 탐지 및 예측
- 2) 시설물 운영 및 보건의료 빅데이터 분석시스템

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

● 교통사고 다발생 환자 및 의료기관 탐지 및 예측

▶ 사회적 이슈와 자동차 보험 사기

이슈

탐색 및
검증

방안
제시

- 보험사기를 노린
의심환자로서
3년간 10회 이상
교통사고를 겪은
환자가 935명에 달함
- 이들이 찾은 의료기관도
특정한 장소에 몰려 있음
- 이로 인해 2015년
한해 보험사기로
적발된 금액이 6,459억 원에 달함

최근 3년 5회이상 사고 1만1460명... 환자 20%가 병원 6곳 몰려
정부 "보험사기 등 의심... 정보공유"

2013년 7월~2016년 6월 5회 이상
다발생 환자를 다수 진료한 병원

구분	소재지	환자수
A의원	대구	530명
B한방병원	서울	375명
C병원	인천	330명
D의원	대구	306명
E병원	인천	306명
F병원	대구	304명



자료: 건강보험심사평가원

〈출처: 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

● 교통사고 다발생 환자 및 의료기관 탐지 및 예측

▶ 사회적 이슈와 자동차 보험 사기

이슈

탐색 및
검증

방안
제시

- 교통사고 환자의 특성과 교통사고 환자를 진료한 의료기관의 특성을 분석함
- 사고다발생 지표, 병상 가동률 지표, 비용 과다와 장기 입원 지표 등을 이상패턴탐지를 위한 지표로 활용함

〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

● 교통사고 다발생 환자 및 의료기관 탐지 및 예측

▶ 사회적 이슈와 자동차 보험 사기

이슈

탐색 및
검증

방안
제시

- 자동차보험 진료비 심사 자료를 활용하여, 보험사기 의심환자와 병원을 찾아내는 “보험사기 빅데이터 정보공유 체계”를 구축하는 방안을 추진함

- 심평원은 교통사고 진료비 심사 업무를 위탁받아 수행하고 있음.
 - 민간보험사들은 자기고객의 사고기록만 보유하지만, 심평원의 모든 보험사들의 고객 사고기록을 보관하고 있음.
- 심평원의 데이터를 분석하면 일정기간 동안 교통사고가 다발생 환자 파악이 가능.

< 연도별 사고발생 횟수별 환자수 >

(단위 : 명)							
연도	계	1회	2회	3회 이상	5회 이상	10회 이상	20회 이상
'13.7.~'16.6.	4,360,492	3,781,572	471,014	96,446	10,525	857	78
2015	1,693,327	1,583,932	97,901	10,569	869	51	5
2014	1,650,723	1,545,998	93,949	10,014	701	55	6

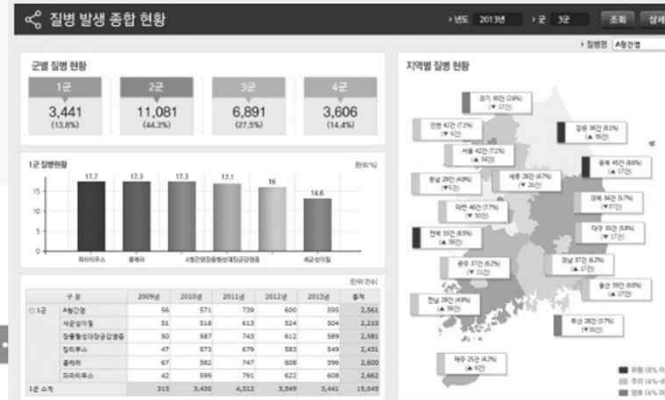
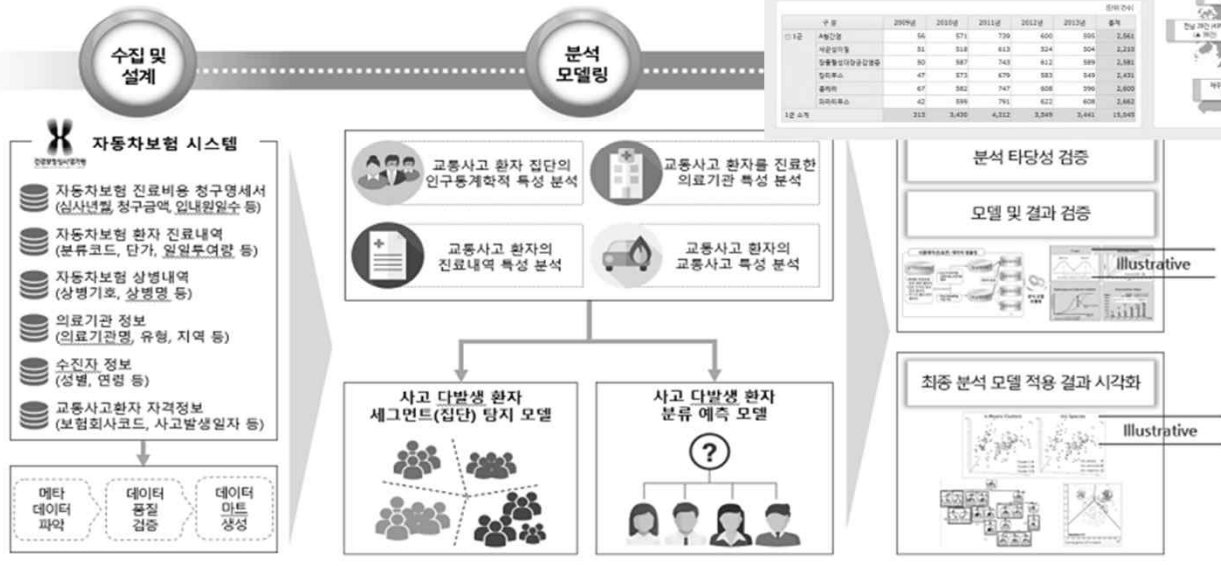
출처 : 건강보험심사평가원

<출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원>

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

● 이상패턴 탐지를 위한 분석 프로세스 및 시각화

○ 분석 프로세스



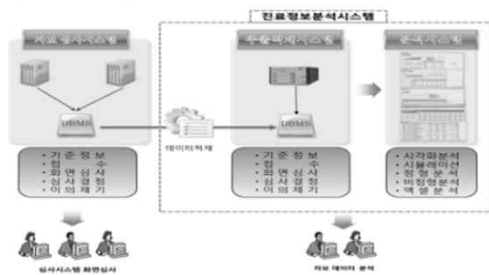
〈출처: 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

● 사고다발생 분석을 위한 진료정보분석시스템 구축

1차년도 : 진료정보분석시스템

① 목표시스템 구성



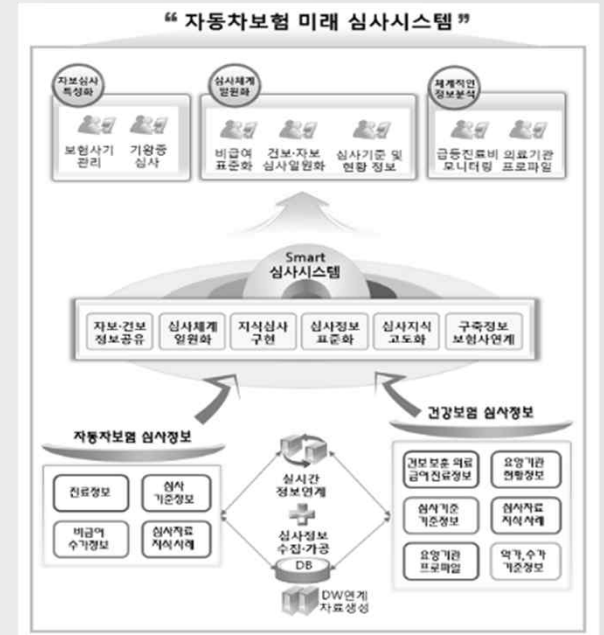
2차년도 : 사고다발생 분석(차세대)

- 지역별 사고별 유형을 분석한 데이터를 시각화 하기
- 모니터링을 통한 문제 인식
- 임원 및 현업 담당자도 동일한 화면으로 같은 시점에 정보 공유

- 주요지표를 통한 현황분석의 비교 및 검토
- 상위 랭킹을 통한 전체 비율을 직관적으로 인식

- 관련된 다른 주요지표를 통한 현황분석의 비교 및 검토
- 상위 랭킹의 변화를 통한 상대적 중요성 인식

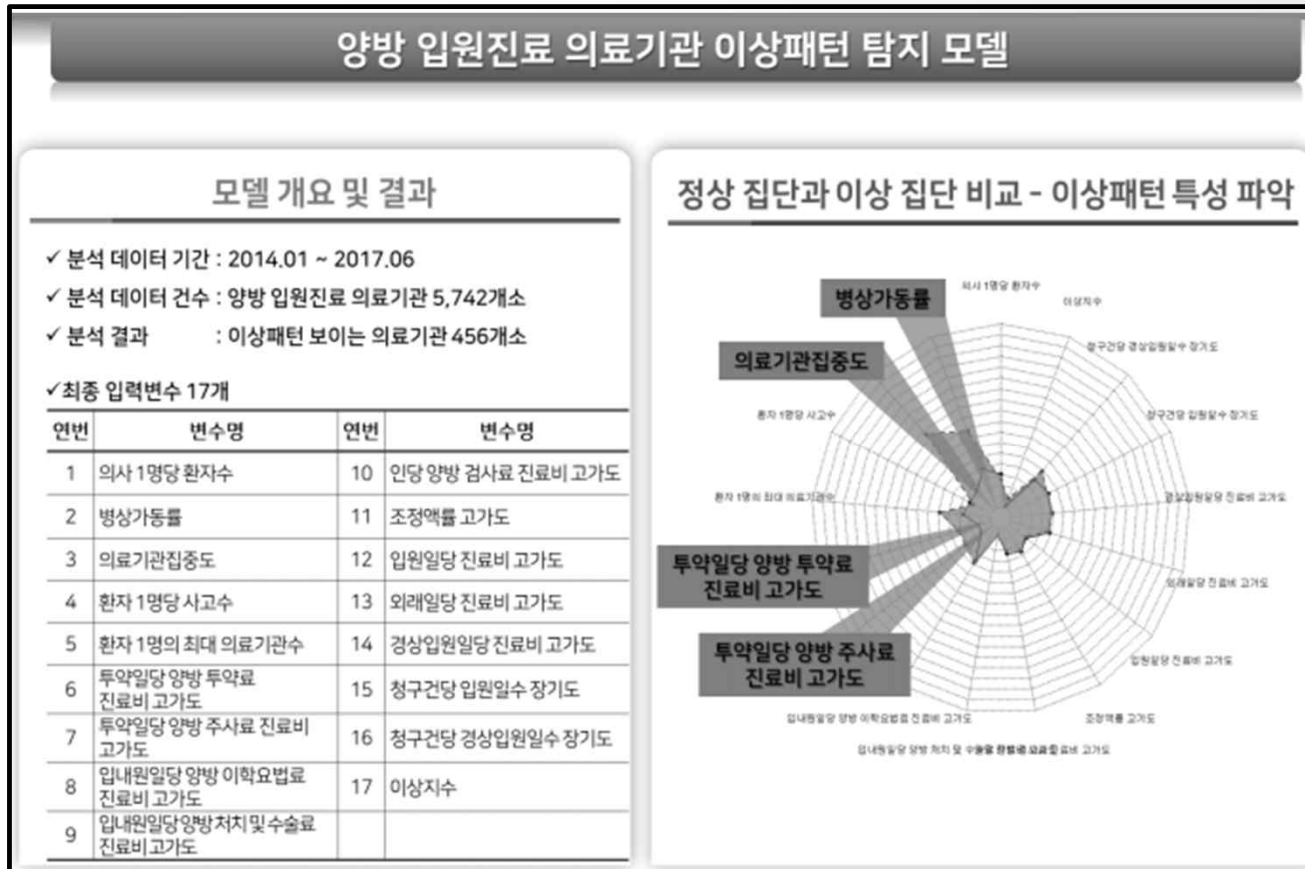
- 관련된 데이터의 화면 출력 및 레포팅
- OLAP 기능을 통한 필요한 데이터 생성



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

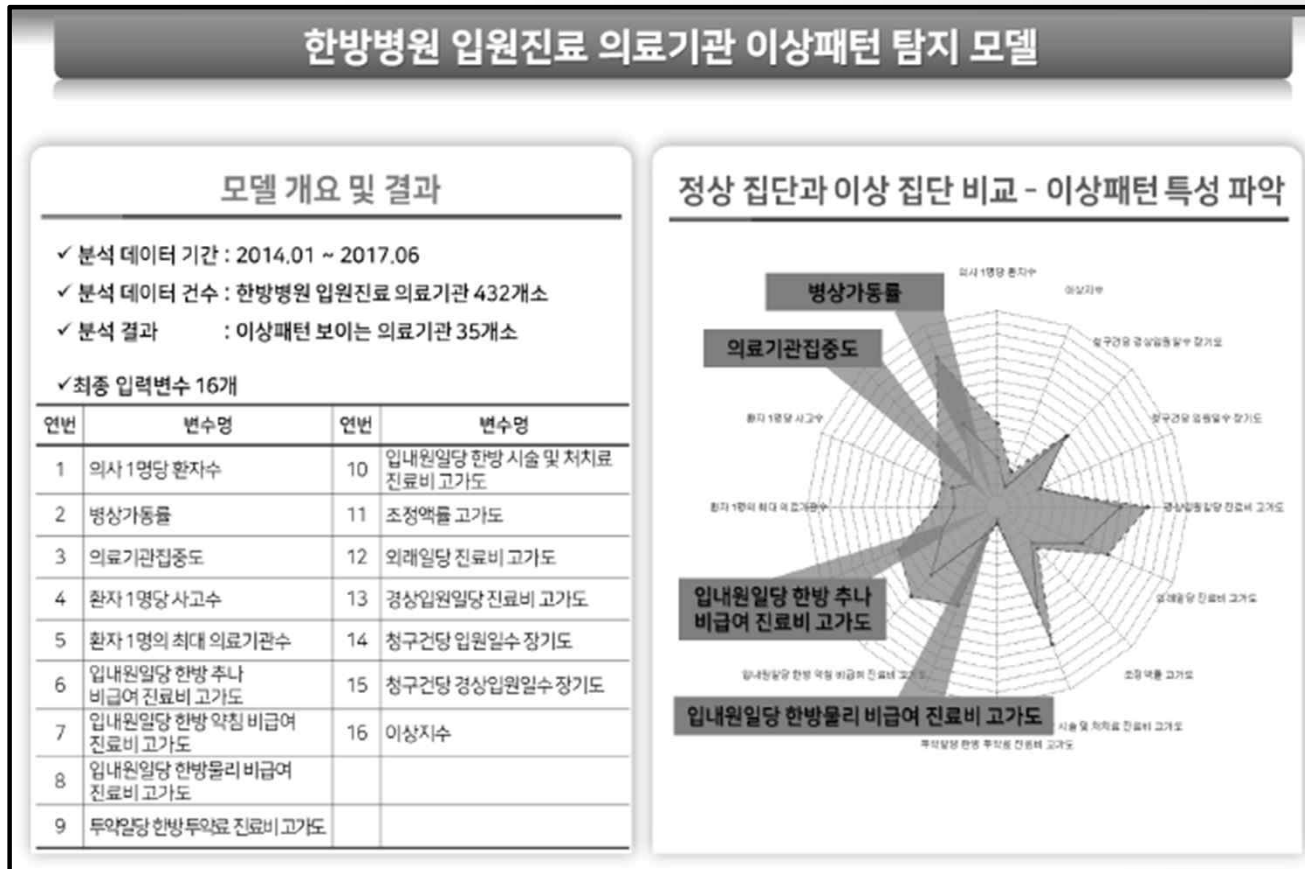
● 양방 입원진료 의료기관 이상패턴 탐지 모델



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

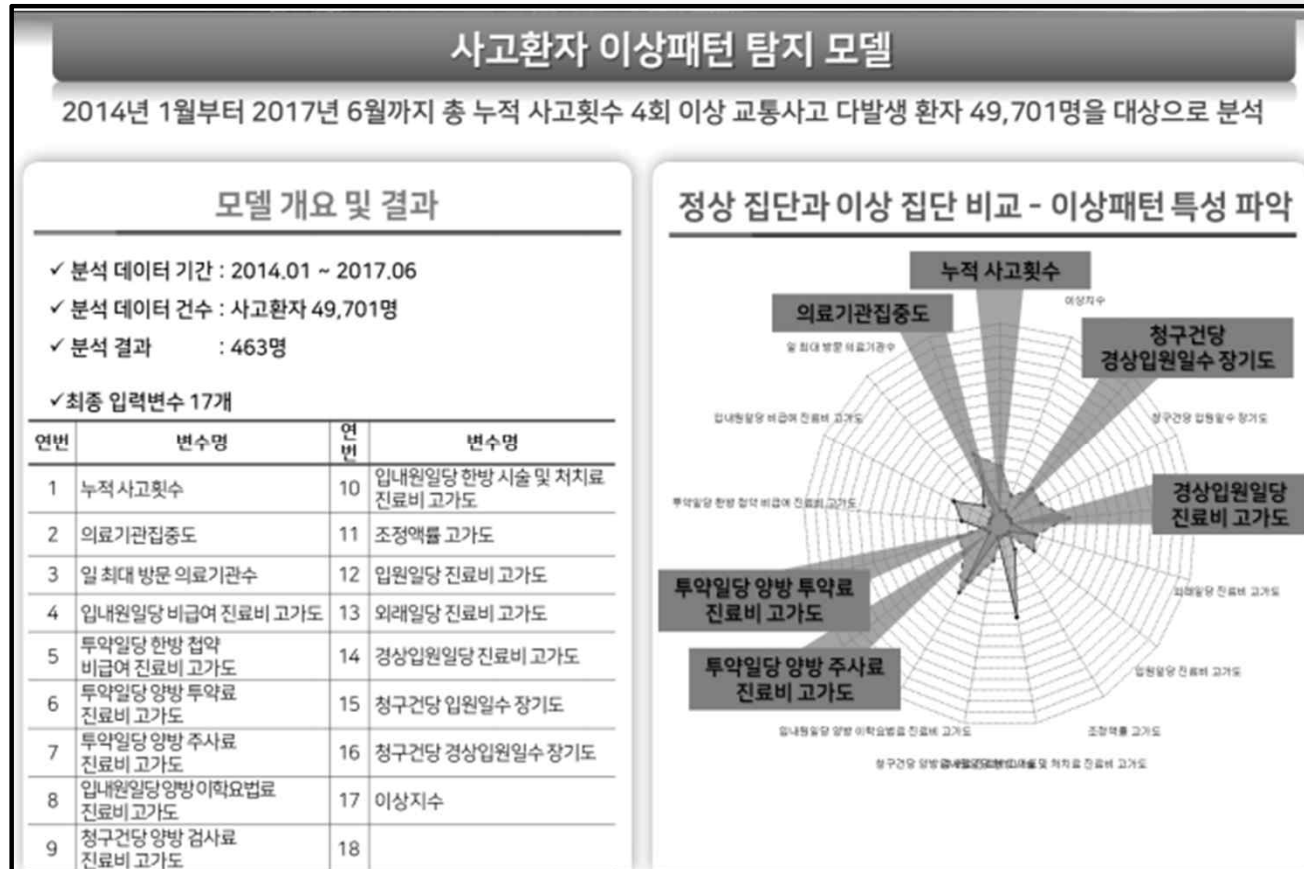
● 한방병원 입원진료 의료기관 이상패턴 탐지 모델



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

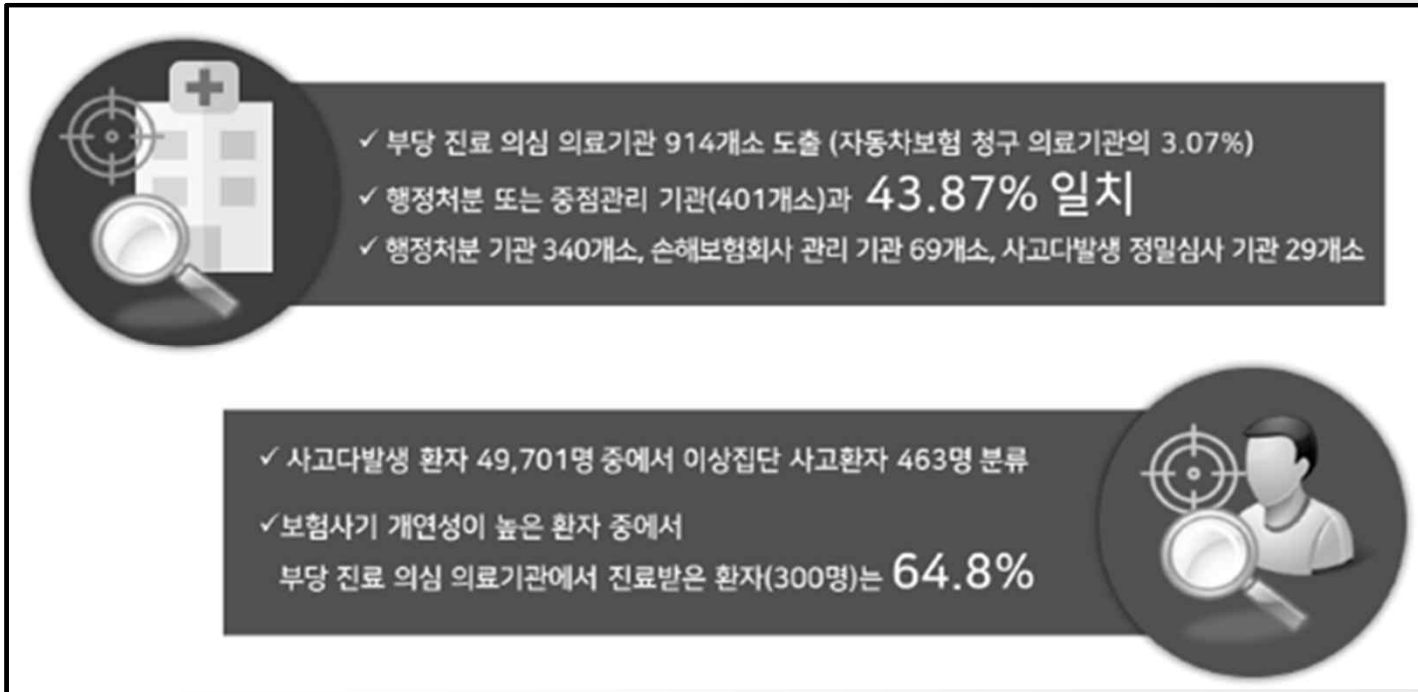
● 사고환자 이상패턴 탐지 모델



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

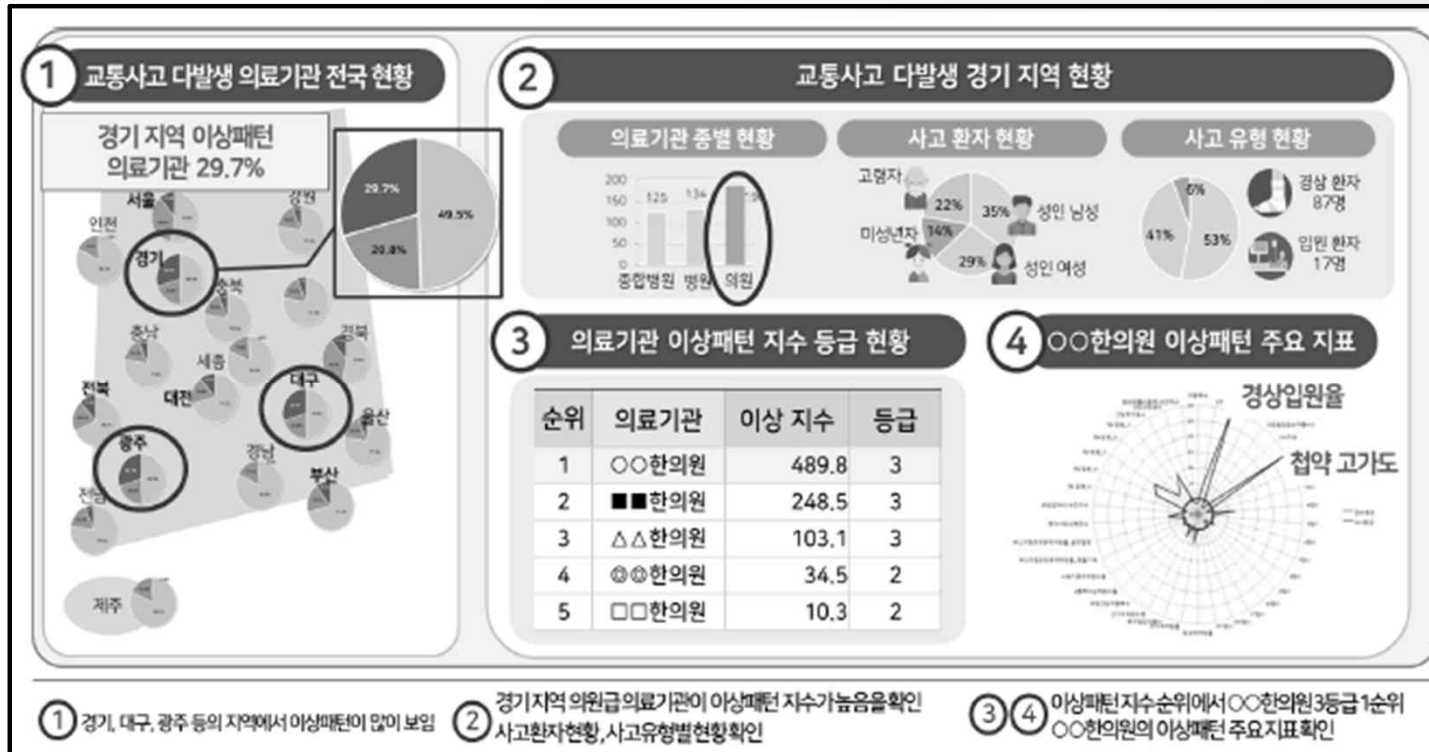
● 교통사고 다발생 의료기관 및 사고환자 분석 결과 검증



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

1) 자동차보험 교통사고 다발생 환자 탐지 및 예측

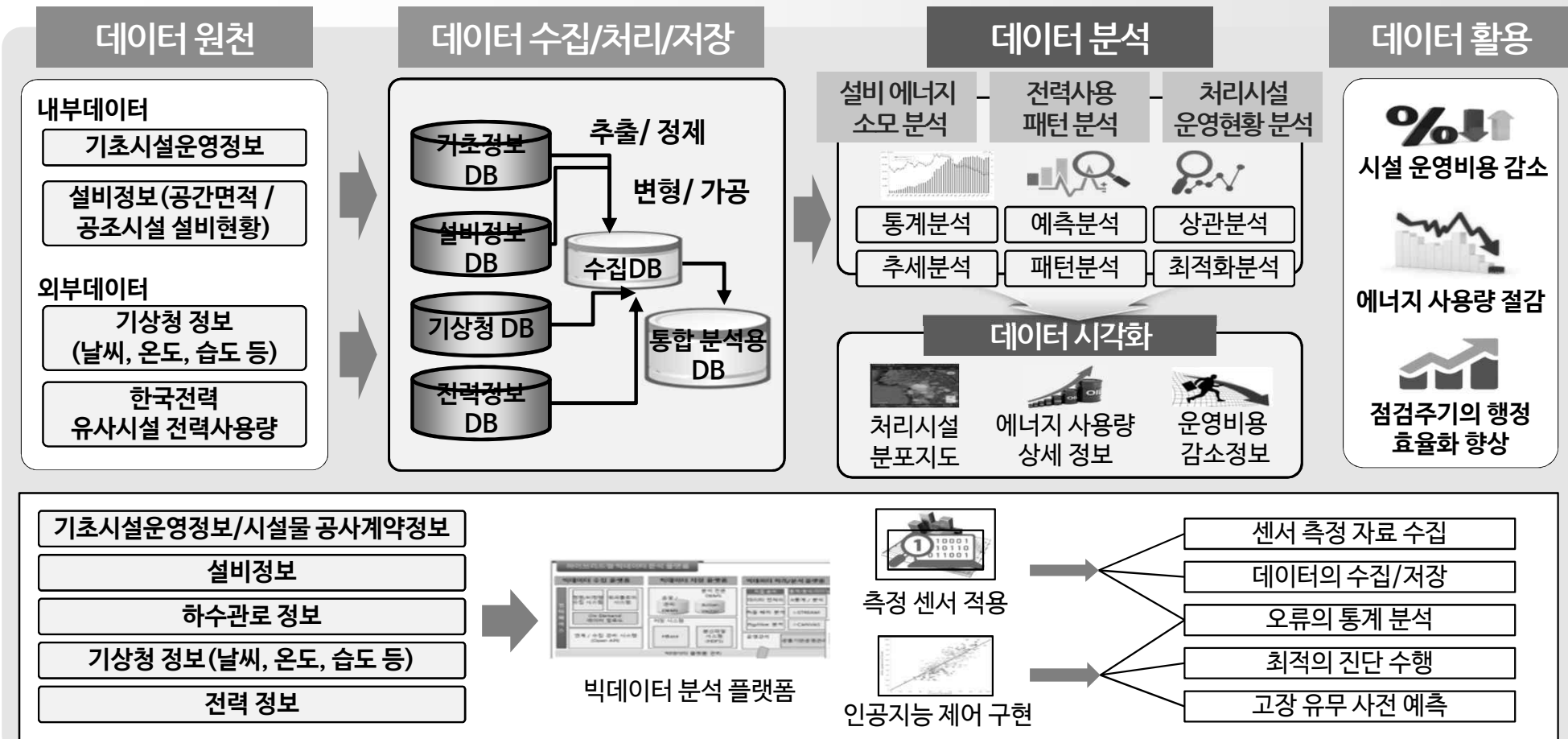
● 사용자 관점의 분석결과 시각화(진료정보분석시스템)



〈출처 : 2017년 빅데이터 성과공유대회, 행정안전부, 한국정보화진흥원〉

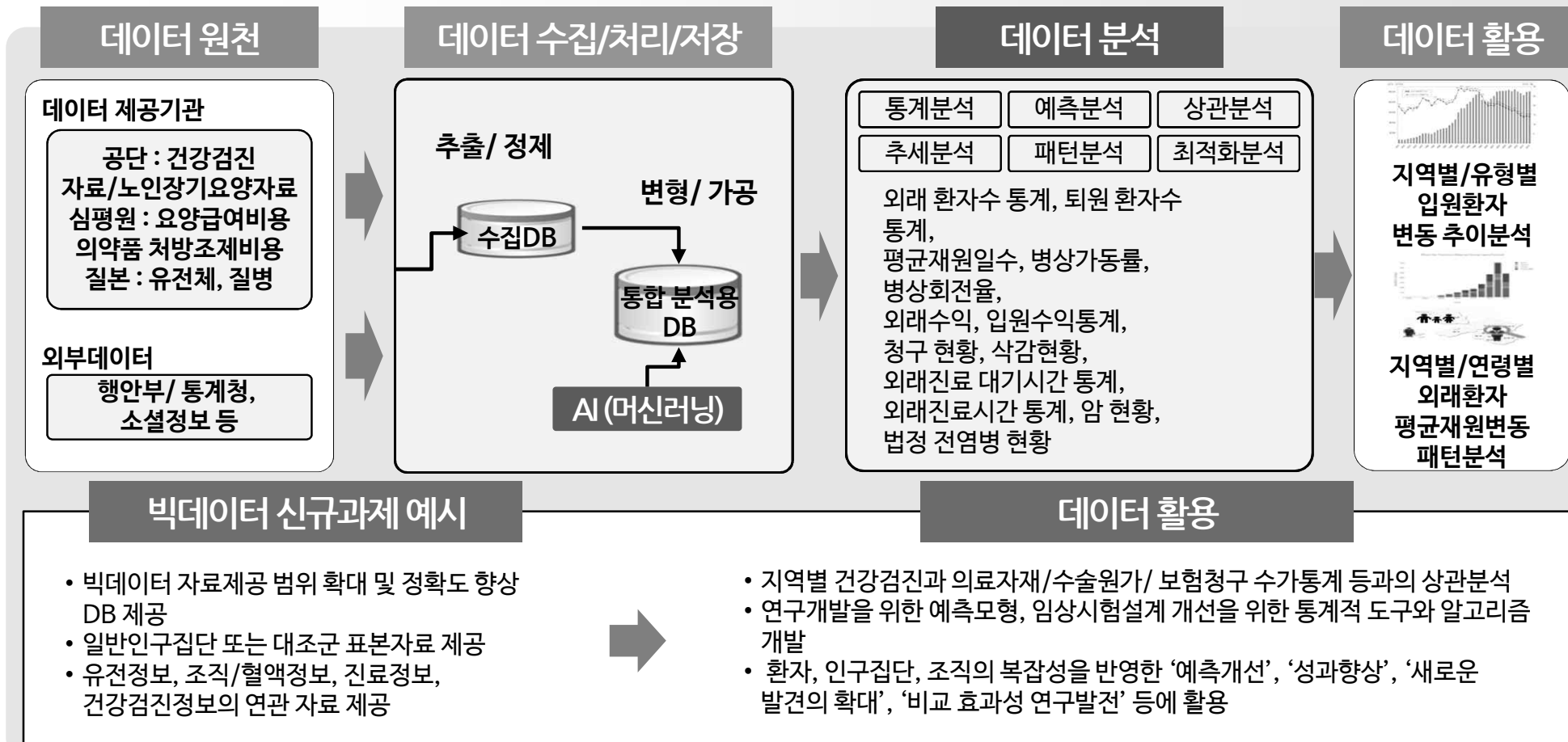
2) 시설물 운영 및 보건의료 빅데이터 분석시스템

● 빅데이터 및 IoT 기반 시설물 스마트 운영관리(예시)



2) 시설물 운영 및 보건의료 빅데이터 분석시스템

● 보건의료 빅데이터 제공 서비스(예시)



빅데이터(Big data)가 발전하는 주요 역사에
대해 논의해 봅시다

- 2011년까지 Hadoop이나 NoSQL 데이터베이스 등 기반 기술의 발전
- 2012년까지 클라우드 방식의 데이터 웨어하우스(Data Warehouse)
및 BI 도구의 보급
- 2013년까지 스트림 처리나 애드 혹(Ad-hoc) 분석 환경의 확대

학습 활동

A1

A1

빅데이터(Big data)가 발전하는 주요 역사에
대해 논의해 봅시다

2011년:

Hadoop: 다수의 컴퓨터에서
대량의 데이터를 처리하기 위한
시스템

NoSQL: 전통적인 RDB의 제약을
제거하는 것을 목표로 한
데이터베이스의 총칭

학습 활동

A1

A1

빅데이터(Big data)가 발전하는 주요 역사에
대해 논의해 봅시다

2012년:

Data Warehouse: 클라우드
서비스의 보급에 의해 빅데이터의
활용이 증가

셀프서비스용 BI: DW에 저장된
데이터를 시각화

빅데이터(Big data)가 발전하는 주요 역사에
대해 논의해 봅시다

2013년:

벌크형 수집: 우리가 전통적인 DW에서
사용하던 형태로 DBMS, 파일 서버
등에서 각각의 방식으로 데이터를
추출한다.

스트리밍형: 차례차례로 생성되는
데이터를 끊임없이 계속해서 보내는
방법으로 모바일 애플리케이션과
임베디드 장비 등에서 널리 데이터를
수집

논의 하기

데이터 거버넌스

- ✓ 데이터 거버넌스는 전사차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영 조직 및 책임 등의 표준화된 관리체계를 수립하고 운영을 위한 프레임워크 및 저장소를 구축하는 것을 말한다.
- ✓ 데이터 거버넌스의 대상은 마스터데이터, 메타 데이터, 데이터 사전이다.

데이터 거버넌스 구성요소:

- 1) 원칙 (Principle)
- 2) 조직 (Organization)
- 3) 프로세스 (Process)

데이터 거버넌스 체계:

- 1) 데이터 표준화
- 2) 표준화 활동
- 3) 데이터 관리 체계
- 4) 데이터 저장소
(레포지토리) 관리

정리 하기

빅데이터 처리 과정

- 빅데이터 수집:

수집한 데이터를 저장 · 분석하기 위해 데이터를
통합하거나 변환하는 작업을 일컫는다.

- 빅데이터 저장:

데이터를 사용할 목적으로 데이터를 안정하고 영구적인
방법으로 보관한다.

- 빅데이터 처리:

일괄처리 및 실시간 처리는 데이터를 빨리 처리하기 위한
시스템과 방법이 필요하다.

어떤 형태의 용도로 사용하도록 준비한다.

- 빅데이터 분석:

데이터로부터 의사결정을 위해서 모으고 변환하며,
분류하고, 특징을 찾아낸다.



- 빅데이터의 이해와 활용 -

수고하셨습니다.