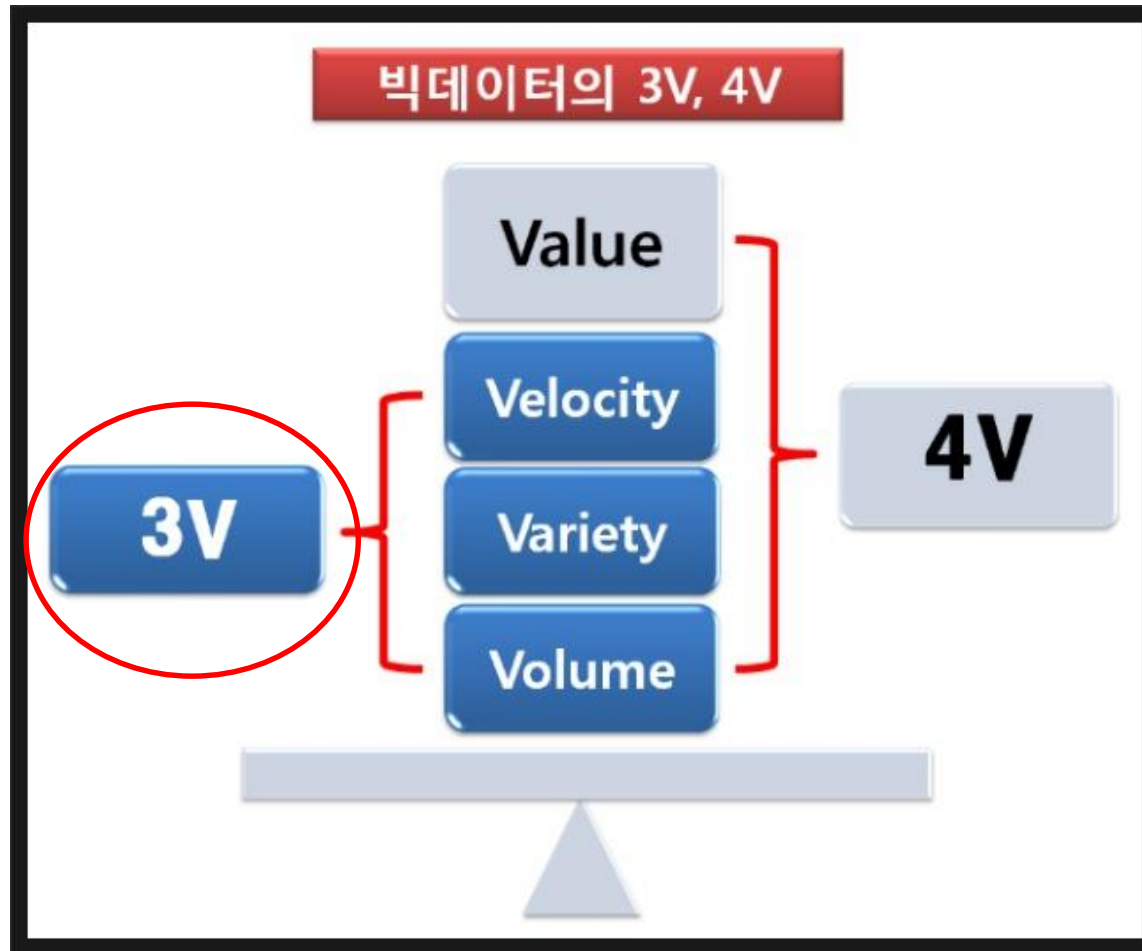


데이터 분석

학습 목표

- 데이터 분석 의미를 확인한다.
- 데이터 분석에 관련된 직무를 확인하고 직무에 관련된 준비사항을 계획할 수 있다.

3V

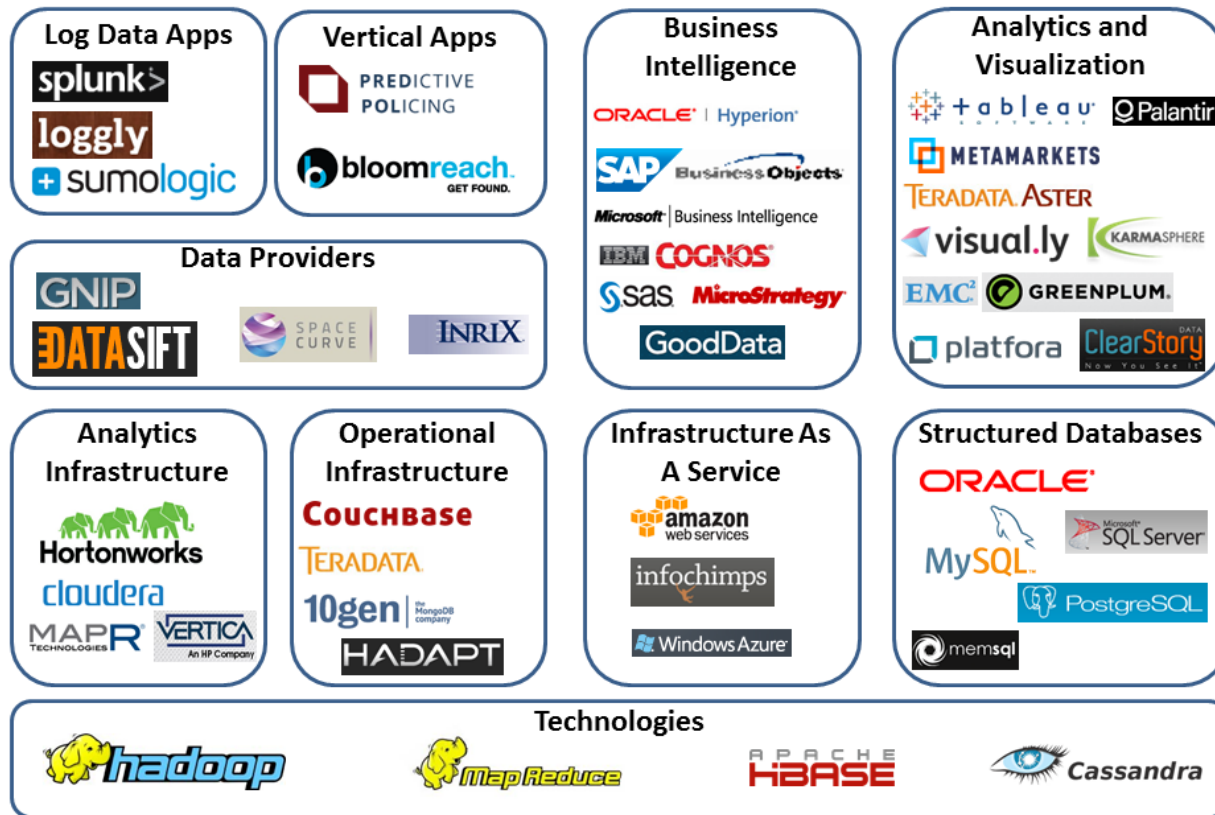


빅 데이터 소스

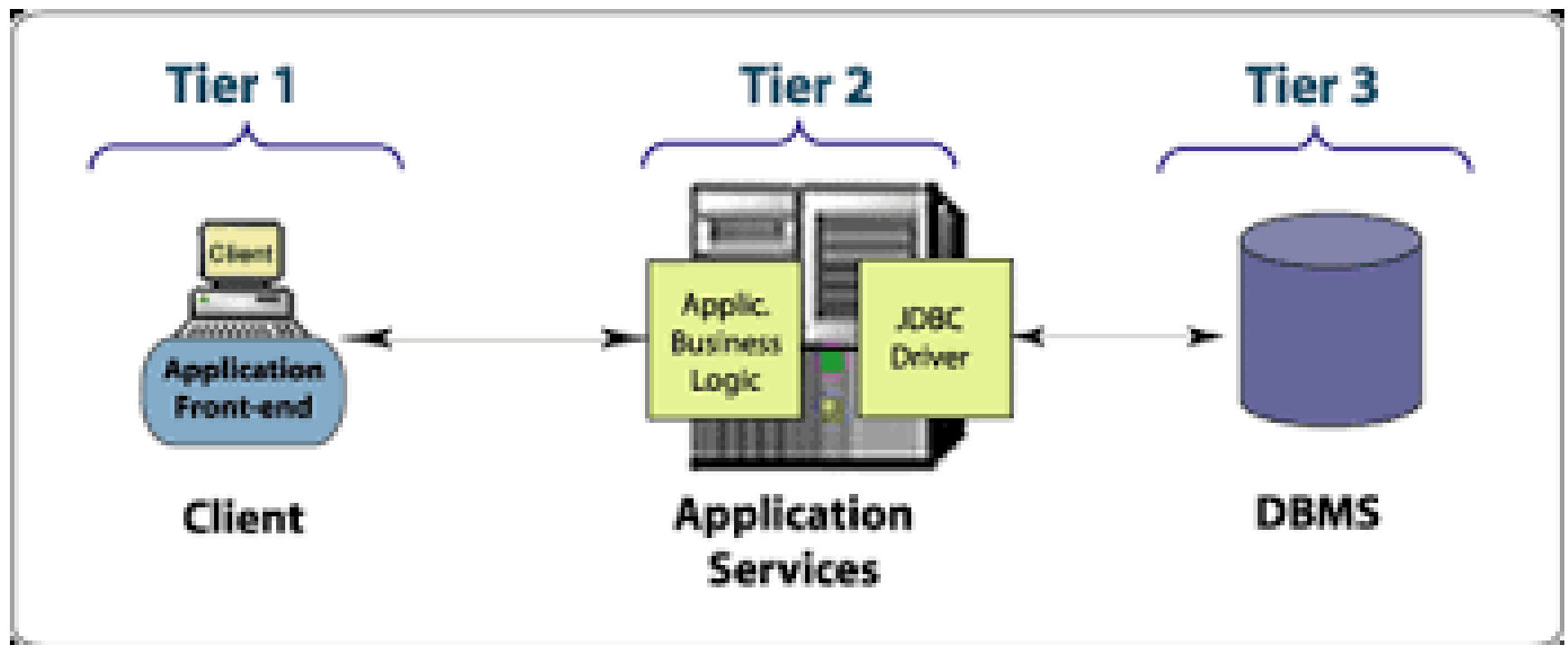
- 센서
- 유전자 정보
- 로그 데이터
- Social Network Service
- 디지털 활동

일반적 생각의 Big Data

Big Data Landscape



서비스 기본 구조



Big Data Processing Phase



데이터의 시각화



storytelling with data

a data
visualization
guide for
business
professionals

the Power of Infographics

Using pictures to
communicate and
connect with
your audiences

Foreword by Guy Kawasaki, author of *Enchantment*
and former chief evangelist of Apple

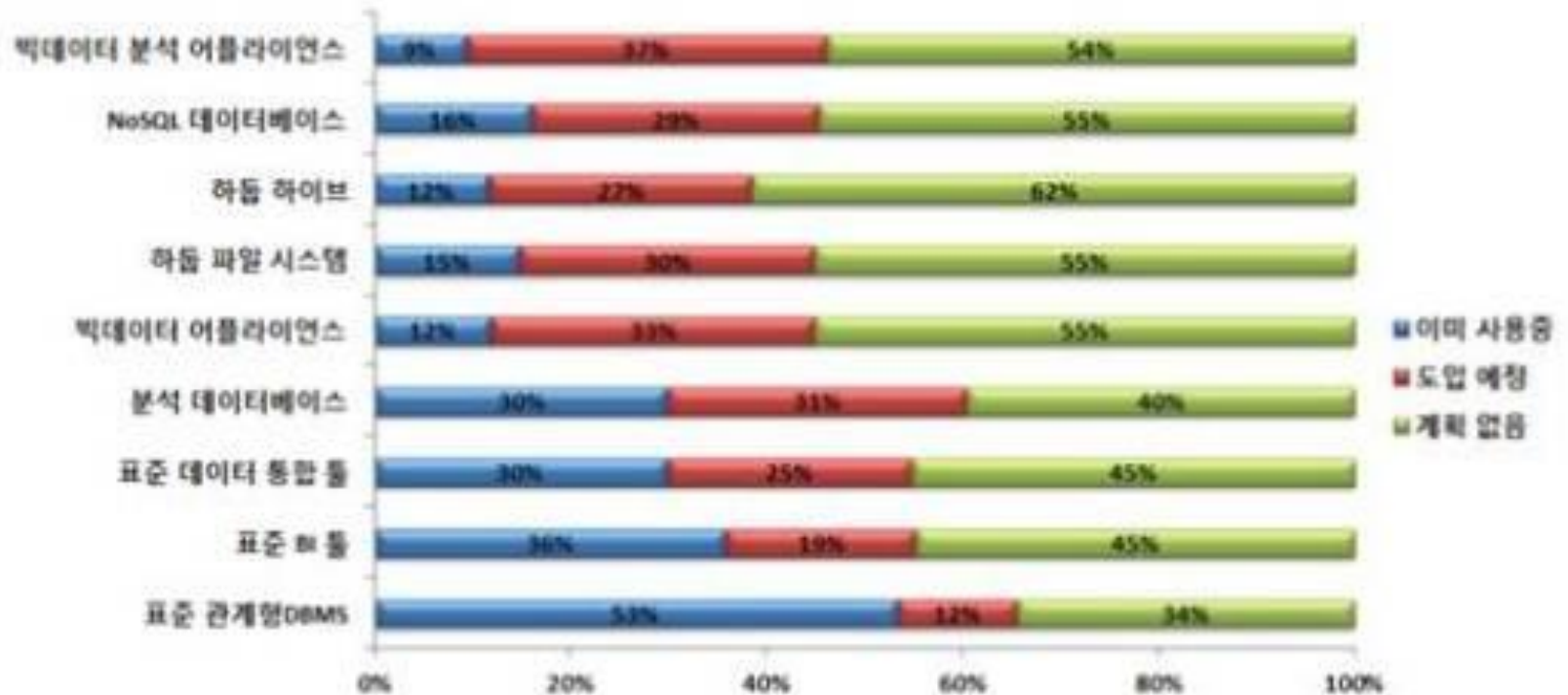
que

Mark Smiciklas






빅데이터 솔루션

그림 21 | 빅데이터 솔루션 도입 현황



결론은

IT ▾	국내 기업 빅 데이터 활용 4% 수준... 분석·서비스보다 하드웨어에 몰려 강영수 기자 ▾	
기사	100자평(0)	   + 크게 - 작게

입력 : 2016.09.26 03:06

오정연 정보화진흥원 ICT(정보통신기술) 융합본부 수석연구원은 "우리나라에서는 아직 빅 데이터를 활용한 성공 사례가 없어 기업들이 빅 데이터에 대한 투자를 꺼리고 있다"며 "초보적인 수준이더라도 기업과 사용자들이 고개를 끄덕일 만한 대표적인 성공 사례를 만들어야 한다"고 말했다.

3V focus

◆‘빅데이터 플랫폼’ 구축 프로젝트 러시 = 수년전부터 빅데이터 플랫폼 시범(파일럿) 프로젝트를 통해 빅데이터의 내부 접목을 타진하던 은행들은 올해 본사업 발주를 통해 전사 빅데이터 플랫폼 구축을 추진하고 있다.

금융권은 ‘빅데이터 플랫폼’의 구현 요건을 다음과 같이 정의하고 있다. ▲확장성 있는 대용량 처리 능력, ▲이기종 데이터 수집 및 통합 처리 능력, ▲빠른 데이터 접근 및 처리 능력, ▲대량의 데이터를 저장 관리할 수 있는 능력, ▲대량의 이기종 데이터를 원하는 수준으로 분석할 수 있는 능력이다.

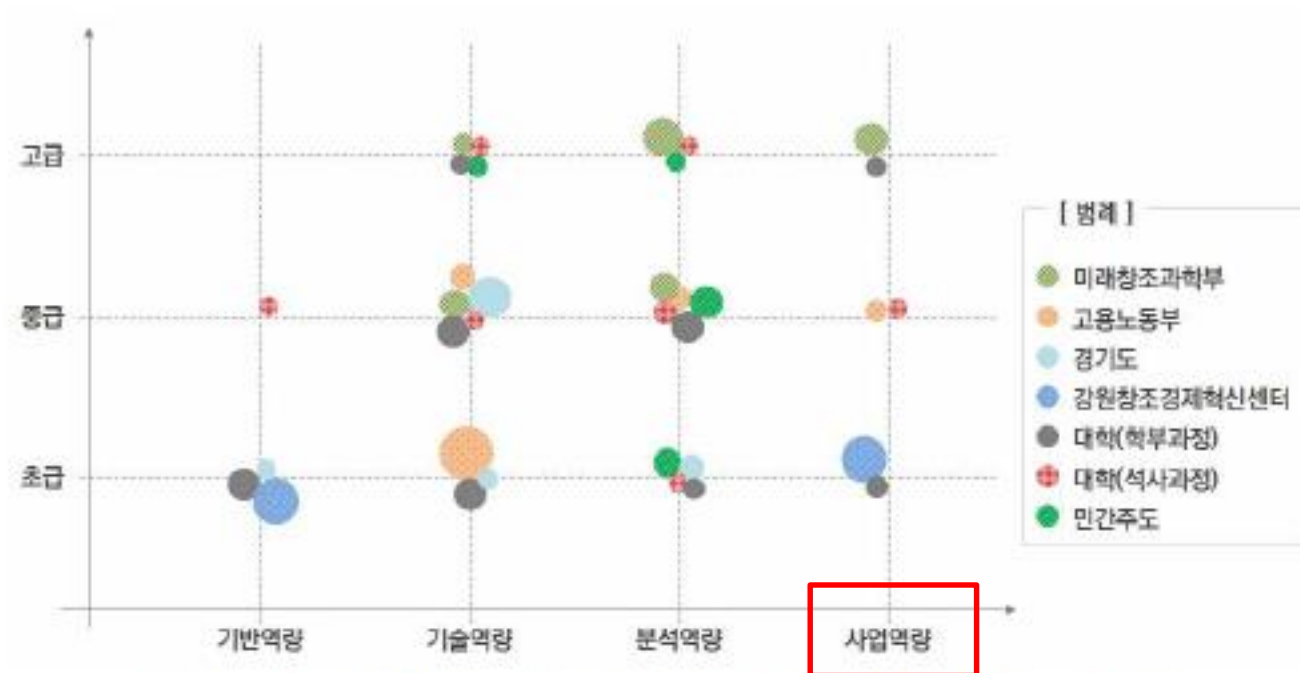
최근까지 금융권에서 진행된 빅데이터 사업은 은행권을 중심으로 빅데이터 기술의 금융분야 적용 가능성을 검증하거나, 비즈니스 ROI를 예측하기 위한 시범사업 차원에서 진행됐다. 올해는 은행을 중심으로 빅데이터 플랫폼 사업이 본격화되고 있다. 데이터 중심의 의사결정 체계 확보와 상품개발 프로세스를 마련하겠다는 게 목표다.

Activity



빅데이터 분석 목적?

부가 가치 창출



[그림 5-8] 국내 데이터 인력양성 커리큘럼의 역량별·수준별 분석 결과

분석의 의미?!

[표 3-7] 경기도 빅데이터 전문가 과정의 교육내용

주요역량	교육과목
기반역량	• 빅데이터 트렌드, 거버넌스, 데이터 사이언스, 미니 프로젝트
기술역량	• IT기초(DB, SQL, N/W, 클라우드), 빅데이터 플랫폼(Hadoop, Java)
분석역량	• (챌린지) 데이터 수집, 탐색(apache sqoop, apache Pig), 처리, 분석(aPache Hive), 시각화(R programming) • (어드밴스) 통계분석 도구 학습, 데이터 분석을 위한 기초 통계, 머신러닝

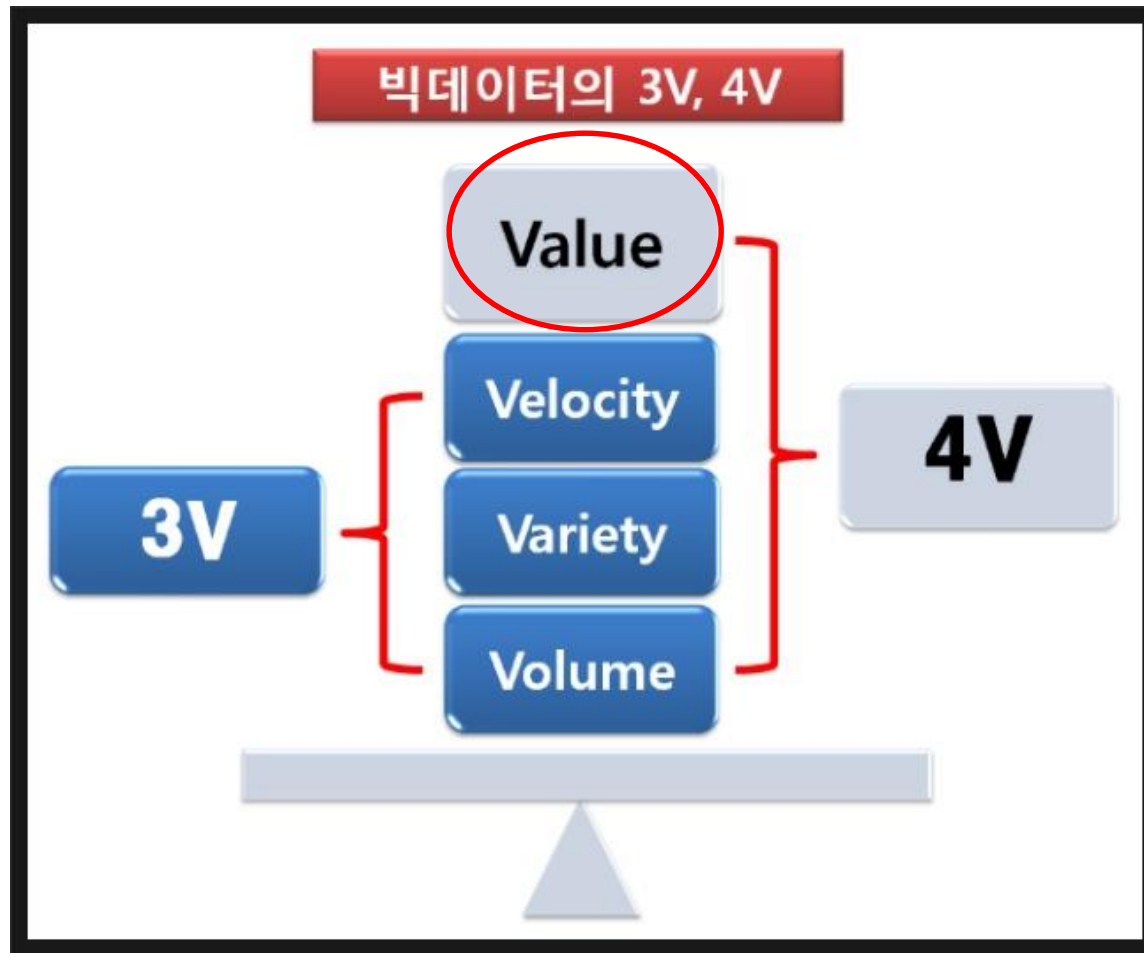
※ 자료 : 경기콘텐츠진흥원 홈페이지(www.gcon.or.kr)

[표 3-13] 삼성 SDS 멀티캠퍼스의 빅데이터 아카데미 과정

구분	입문	핵심	고급
기술 전문가	NoSQL 데이터 모델링	실시간 빅데이터 분석	-
		MongoDB! 개발부터 관리까지	
		클라우드 컴퓨팅을 위한 하둡(Hadoop)	
분석 전문가	빅데이터 분석과 시각화 입문	R을 이용한 데이터 분석	R을 이용한 예측분석 모델링
		디지털영상 데이터 분석(VA)	Spark를 활용한 빅데이터 분석
	이정훈의 빅데이터 분석 기본	이정훈의 Excel Based Machine Learning	오픈소스를 활용한 머신러닝과 딥러닝
		빅데이터를 위한 파이썬	
공통 역량	빅데이터 분석과 시각화 입문	이정훈의 Excel Based Machine Learning	-
	이정훈의 빅데이터 분석 기본	핵심! 빅데이터 분석 전문가	

※ 자료 : 삼성 SDS 멀티캠퍼스 홈페이지(www.multicampus.co.kr)

4th V



Resource is Data?



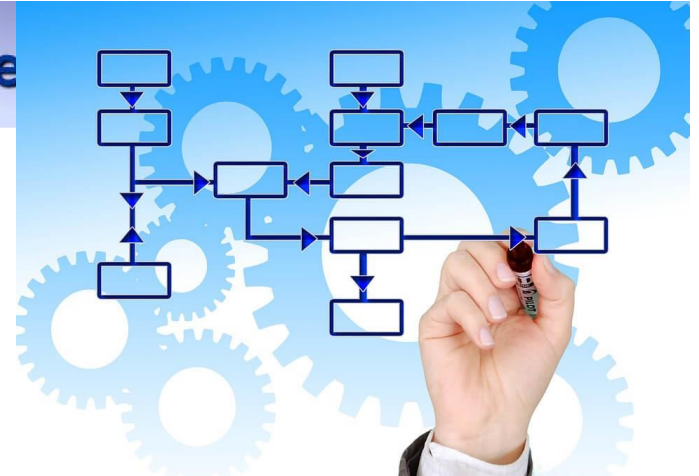
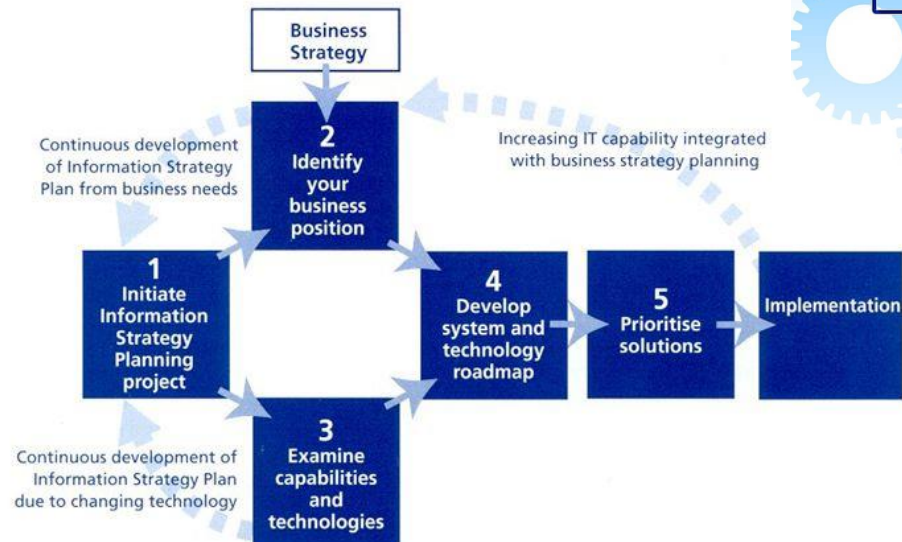
비즈니스 분석 과정1



비즈니스 분석과정2

Developing an Information Strategy

Five stages are linked:

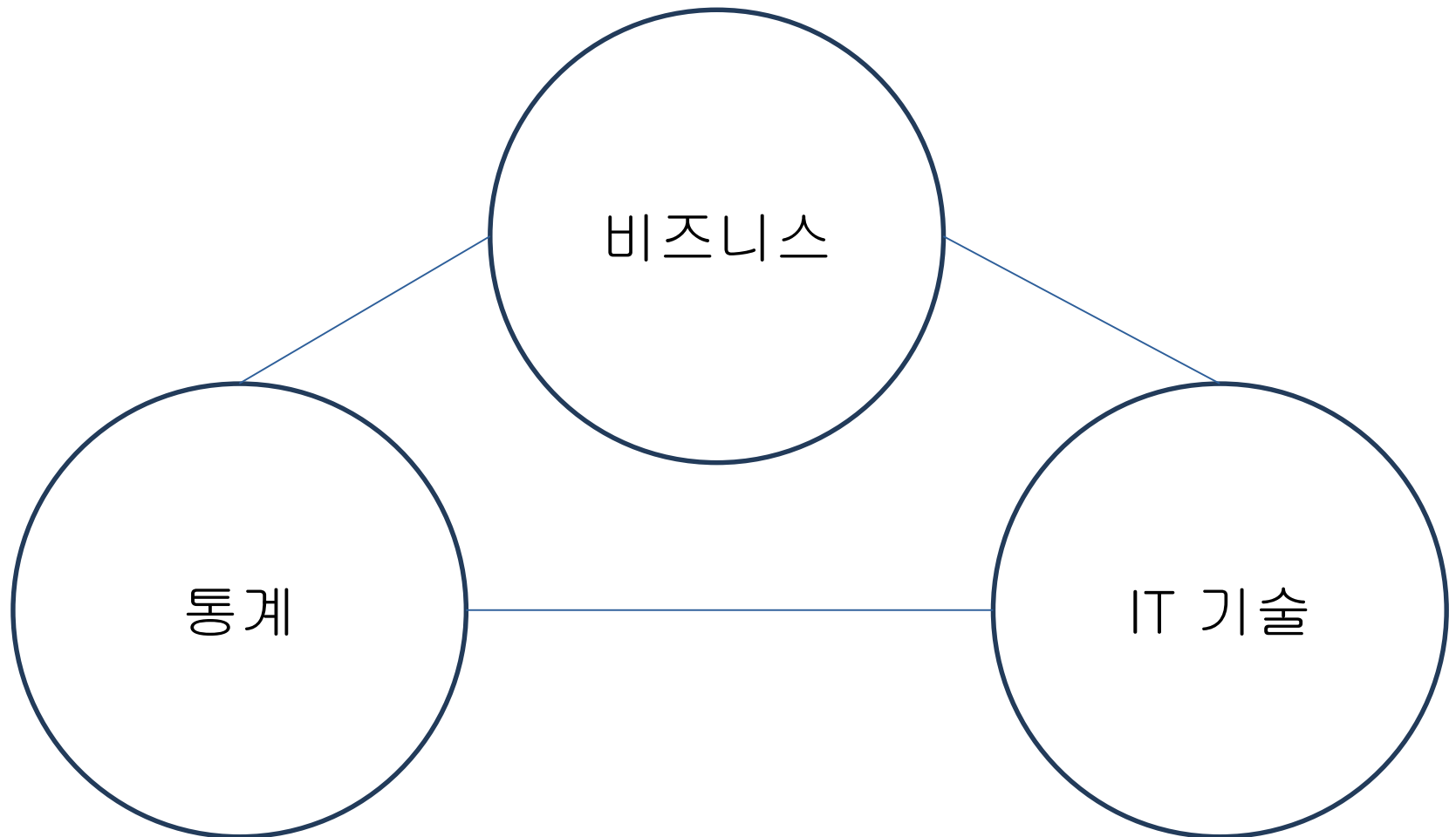


[illegible]

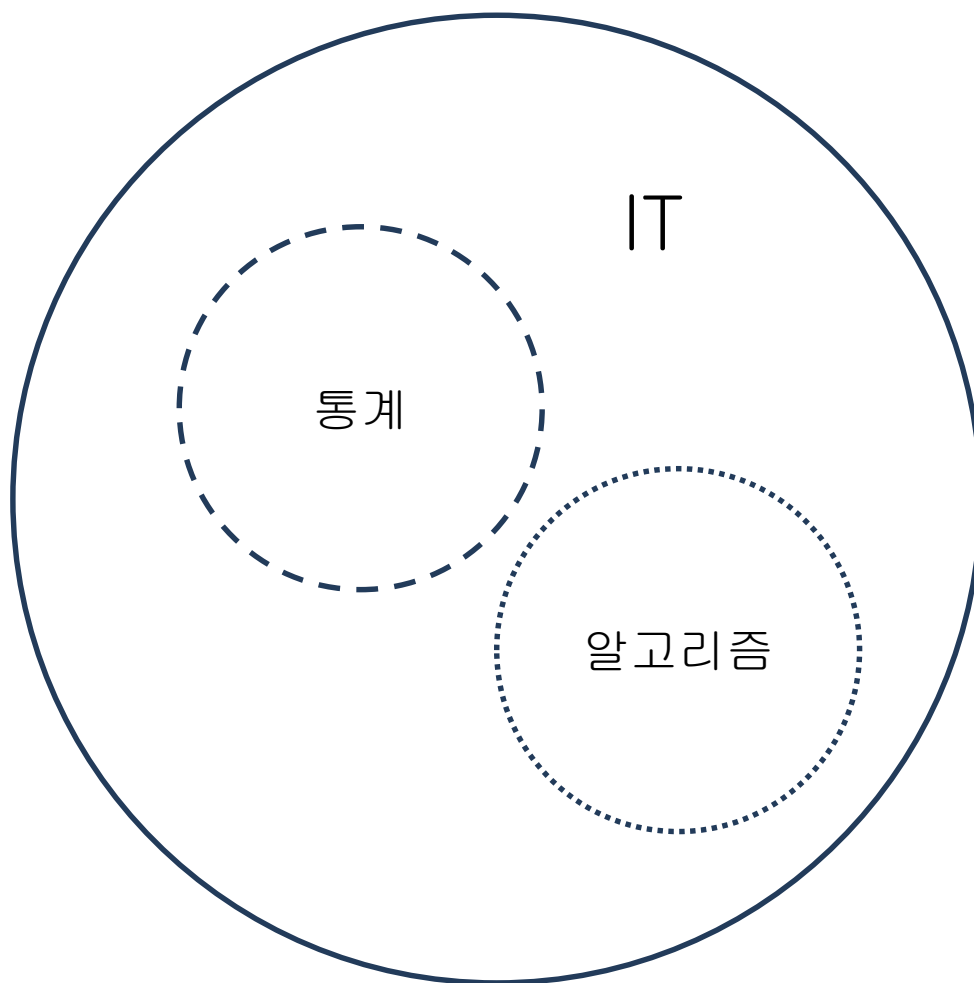
Data Mining

- data
- information
- knowledge
- wisdom

3가지 구성 요소



IT 기술에 숨겨진 것들



융합으로서의 빅데이터

하지만 이에 반해 우리나라는 '세분화된 전문역량을 갖춘 인재'들의 융합을 촉진하는 프로그램이 부족한 편이다. 특히, 대학은 기존의 학과-단과대학의 칸막이가 심해 데이터 교육의 특성인 융합교육 시스템을 갖추는데 걸림돌로 작용하고 있다. 데이터는 수리통계, IT, 비즈니스 등 다양한 학문이 융합된 분야로써 기존의 단일 학과에서 교육하기는 어려운 분야지만, 현재 대학에서 융합도 힘들고 학생들의 부전공이나 복수전공을 데이터 분야로 유도하는 것도 쉽지 않은 문제에 봉착해 있다.

이는 아직까지 학과간 융합교육이 일반화되지 않은 국내 대학시스템의 구조적인 문제로 단기간에 해결하기는 어렵다. 따라서 학과간, 부서간 장벽에서 자유로운 공공기관에서 학문간 융합을 바탕으로 한 융합데이터 인력양성 프로그램을 적극 개발하는 것이 필요하다.

ERP vs. Big Data

ERP	Big Data	비고
지금 상황	나아가고자 하는 방향	최종 목적지
과거의 데이터 정리 데이터베이스	과거의 데이터 실시간 데이터 광범위한 데이터	사용데이터 생성 시점
정리된 자료로 부터 정보획득	인과관계를 추출	A.I를 이용한 자동화
도메인 지식 논리적 사고	도메인 지식 논리적 사고 통계, 수리적 방법	머신러닝 이용

빅데이터와 관련된 직업

데이터 직무 구분

[표 2-7] 국내외 데이터 직무 구분 비교표

구분		한국 데이터진흥원	한국 정보화진흥원	고용 노동부	노규성 · 박성택 · 박경혜	DATA FLOO	SAS
데이터 설계		데이터 아키텍트	하둡/NoSQL/Map Reduce엔지니어	DB엔지니어 어링	빅데이터 관리 및 기술(지원)	빅데이터 솔루션 아키텍트	빅데이터 아키텍트
데이터 개발		데이터 개발자				빅데이터 엔지니어	빅데이터 개발자
데이터 운영관리		데이터 엔지니어					
데이터 활용	데이터 분석	데이터분석가/ 데이터 과학자	빅데이터 분석가	빅데이터 분석	빅데이터 분석	빅데이터 분석가/ 빅데이터 과학자	빅데이터 분석가/ 데이터 과학자
	데이터 시각화					빅데이터 비주얼라이저	빅데이터 디자이너
	데이터 연구					빅데이터 연구원	데이터 과학자
데이터 관리		데이터 과학자/ DBA	—	빅데이터 활용/ 빅데이터 관리 및 기술(지원)	빅데이터 활용/ 빅데이터 관리 및 기술(지원)	총괄책임자/ 빅데이터 매니저/ 빅데이터 과학자	프로젝트 매니저/ 데이터 과학자/ 빅데이터 관리자
데이터 판매		데이터 마케터	빅데이터 기술영업/마케터	—	—	—	—
데이터 컨설팅		데이터 컨설턴트	빅데이터 컨설턴트	—	빅데이터 기획	빅데이터 컨설턴트	—

데이터 직무 구분 및 필요역량

[표 2-8] 데이터 직무구분 및 필요 역량

직무 구분	기술적 역량	비즈니스 역량
데이터 아키텍트	<ul style="list-style-type: none"> • Oracle, Java, SQL, Hadoop, SQL Server에 대한 이해 • 데이터 모델링, ETL, EA, 오픈소스 애널리틱스 활용 역량 • Hadoop 애플리케이션 실제 활용 경험과 Hadoop 솔루션의 수명주기 관리 역량 	<ul style="list-style-type: none"> • 빅데이터 솔루션의 구조에 대한 이해를 바탕으로 Hadoop과 같은 기술을 이용하여 빅데이터 솔루션을 제공하는 역량 • 적절한 아키텍처 설계에 필요한 그 밖의 다양한 역량

직무 구분	기술적 역량	비즈니스 역량
DBA	<ul style="list-style-type: none"> • Linux, MySQL, Puppet, Hadoop, Oracle • Disaster Recovery, Configuration Management, Clustering, ETL 관련 역량 • Hadoop, MapReduce 등 빅데이터 솔루션에 대한 이해를 바탕으로 한 데이터 거버넌스 구축 및 데이터 품질관리 역량 	<ul style="list-style-type: none"> • 조직 내 데이터 전략 총괄 및 데이터 업무 추진 감독 • 강한 리더십을 바탕으로 한 고위경영진 및 임원과의 소통 능력 • 데이터의 우선순위 결정 및 핵심사업에서 정보 관리 리딩 경험 • 시장에서 통용되는 주요 데이터 솔루션과 제품에 대한 이해

직무 구분	기술적 역량	비즈니스 역량
데이터 엔지니어	<ul style="list-style-type: none"> DB를 이용하여 프로그래밍을 하는 직무로 JavaScript, SQL, Linux, Oracle, Hadoop, C++, Ruby, Phyton 등 다양한 프로그래밍 언어를 활용하여 빅데이터 처리, LOD 구축 등 데이터 개발 역량 일간 기준 페타바이트 또는 엑사바이트 크기 데이터 취급 역량 	<ul style="list-style-type: none"> 빅데이터 솔루션 아키텍트가 설계한 것을 구축하고 빅데이터 솔루션의 개발, 유지관리, 테스트, 평가 역량 큰 규모의 데이터를 수집, 저장, 관리, 분석하고 복수의 플랫폼을 사용하여 정보를 시사점(Insight)으로 전환시키며 이와 관련되는 H/W, S/W, 인프라를 지원하는 역량
데이터 분석가	<ul style="list-style-type: none"> Oracle, SQL, Java 프로그램에 대한 이해를 바탕으로 통계/머신러닝/텍스트마이닝 기반 데이터 분석 역량 실시간 분석 및 Tableau SW와 같은 비즈니스 인텔리전스 플랫폼 이해 SQL database, 프로그래밍 언어, R이나 SPSS와 같은 통계 소프트웨어 활용 역량 	<ul style="list-style-type: none"> 데이터 과학자를 보조하며 주어진 시스템 안에서 다양한 분석도구를 활용하여 데이터를 실제 분석하는 역량 하둡과 맵리듀스에 대한 최소한의 기초 지식을 가지고 가용 데이터로부터 새로운 시사점(Insight)을 개발하는 능력
데이터 비주얼라이저	<ul style="list-style-type: none"> 분석 결과 스토리텔링, 분석 정보 시각화, 분석 정보 시각표현 유저 인터페이스, 타이포그래피에 대한 이해 인터페이스 디자인, 사용자경험(UX) 디자인, 시각디자인에 대한 이해 	<ul style="list-style-type: none"> 데이터를 시각화하여 조직 또는 상사가 이해하기 쉽게 변환하는 역량 새로운 패턴과 시사점(Insight)을 찾기 위한 창조적 사고방식을 갖추고 데이터 과학자의 의도에 맞추기 위한 데이터 분석 과정 이해
데이터 과학자	<ul style="list-style-type: none"> Hadoop, Java, NoSQL, C++, 파이썬, R 등 다양한 프로그래밍언어 사용 능력 통계학, 분석학(Analytics), 수학 등 한 분야에 국한되지 않는 지식 알고리즘을 개발하기 위한 컴퓨터 사용 및 모델을 공유하기 위한 개념적 모델링 능력 Research 프로젝트를 통해 텍스트 마이닝, 군집분석, recommendation analysis, 극단값 확인, 예측 분석 또는 이와 유사한 기법 수행 능력 복수의 시스템과 데이터셋을 통합하는 방법에 대한 이해 	<ul style="list-style-type: none"> 다양한 내외부 소스로부터 데이터를 추출하고 가장 시급한 과제에 대한 데이터 기반 솔루션 고안 조직내부 및 외부 데이터를 관리/활용/분석하는 체계를 만들고, 프로세스 혁신 및 신제품 개발, 마케팅 전략 결정 등 의사결정 유도 효과적인 데이터 시각화 및 보고를 통해 예측과 결과를 경영진과 IT 부서에 통보 데이터 기술로 다양한 분야에서 새로운 비즈니스 또는 제품개발 기회 발굴 및 예산과 기획을 포함하는 프로젝트 관리 및 자문 능력
프로젝트 매니저	<ul style="list-style-type: none"> Oracle, Netezza, Business Objects and Hyperion ETL, Agile Software Development Python, R, Java, Ruby, SQL 등의 프로그래밍 언어에 대한 지식 Hadoop, Hive 또는 MapReduce 에 대한 기초지식 자연어처리, 기계학습, 개념 모델링, 통계 분석, 예측 모델링 경험 	<ul style="list-style-type: none"> 필요한 분석과 데이터 정의 및 분석체계의 구체화를 계획하는 직무 조직 내 기술부서와 전략 부서의 업무에 대해 이해와 소통능력 데이터 분석의 결과를 조직의 상급 관리자에게 설명하는 능력

데이터 관련 평균연봉

[표 2-9] 국가별 데이터 과학자 평균연봉

국가	1인당 GDP	데이터 과학자 평균연봉	1인당 GDP 대비
한국	25,990달러	43,636달러(약 4천8백만원)	1.7배
미국	57,220달러	109,324달러(약 1억2천만원)	1.9배
독일	41,895달러	53,771달러(약 6천만원)	1.3배
영국	42,106달러	51,986달러(약 5천8백만원)	1.2배
네덜란드	44,828달러	50,591달러(약 5천6백만원)	1.1배
일본	34,871달러	51,781달러(약 5천8백만원)	1.5배

※ 자료 : 한국(채용사이트 '사람인' 참고) / 미국·독일·영국·네덜란드·일본(급여조사 사이트 Payscale 참고)

Activity



기업이 요구하는 인재의 기본 소양?

금융권 관계자는 "현재 시세·오류·고객관리 데이터 정도로 분류·활용하는 게 현재 수준"이라며 "앞 다뤄 빅데이터 공모전 행사를 개최하는 것도 빅데이터 활용 아이디어를 찾기 위한 것"이라고 말했다.

아이디어 부족뿐 아니라 전문 인력과 인프라 부족, 각종 규제도 풀어야 할 문제다.

차상균 서울대 빅데이터연구원장(정보전기공학부 교수)은 "특히 은행과 카드사를 중심으로 실무단계 직원뿐 아니라 의사결정 단계에 있는 임원에게 빅데이터 관련 교육 의뢰가 밀려든다"며 "당장 빅데이터 고급인력을 구하기는 힘들다"고 말했다. 이어 "물리적 인프라 구축에 앞서 전문 인력에 투자해야 할 때"라고 조언했다.

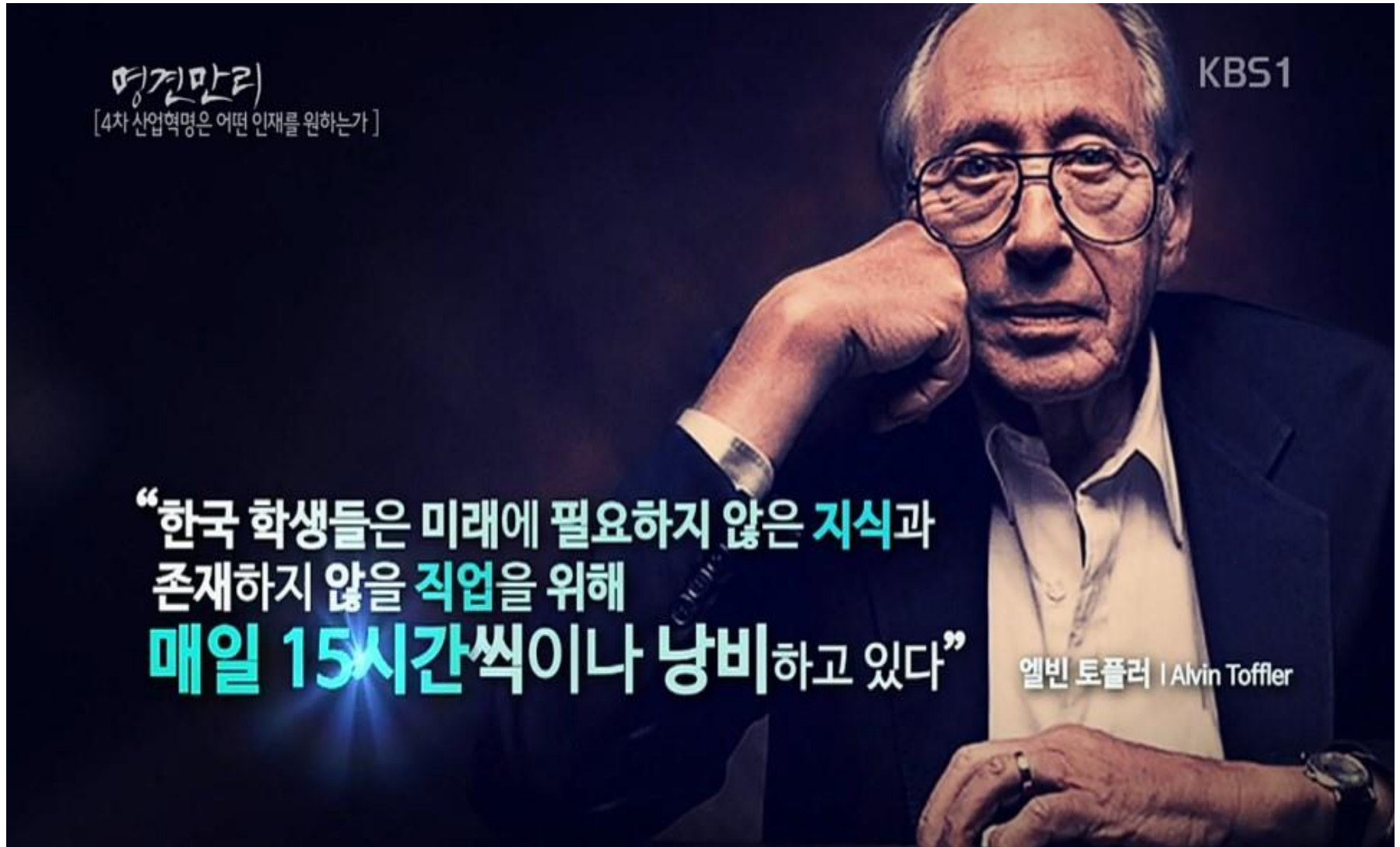
<http://www.ddaily.co.kr/news/article.html?no=157800>

◆빅데이터 전문인력 당분간 크게 부족 = 한편 빅데이터에 대한 금융권의 수요가 폭발적으로 늘고 있지만 정작 원하는 수준의 빅데이터 인력이 금융산업 현장에 크게 부족한 것은 심각한 문제점으로 지적되고 있다. 금융권에서 당분간 빅데이터와 관련한 고급 전문가들의 영입에 상당한 공을 들일 것으로 예상된다.

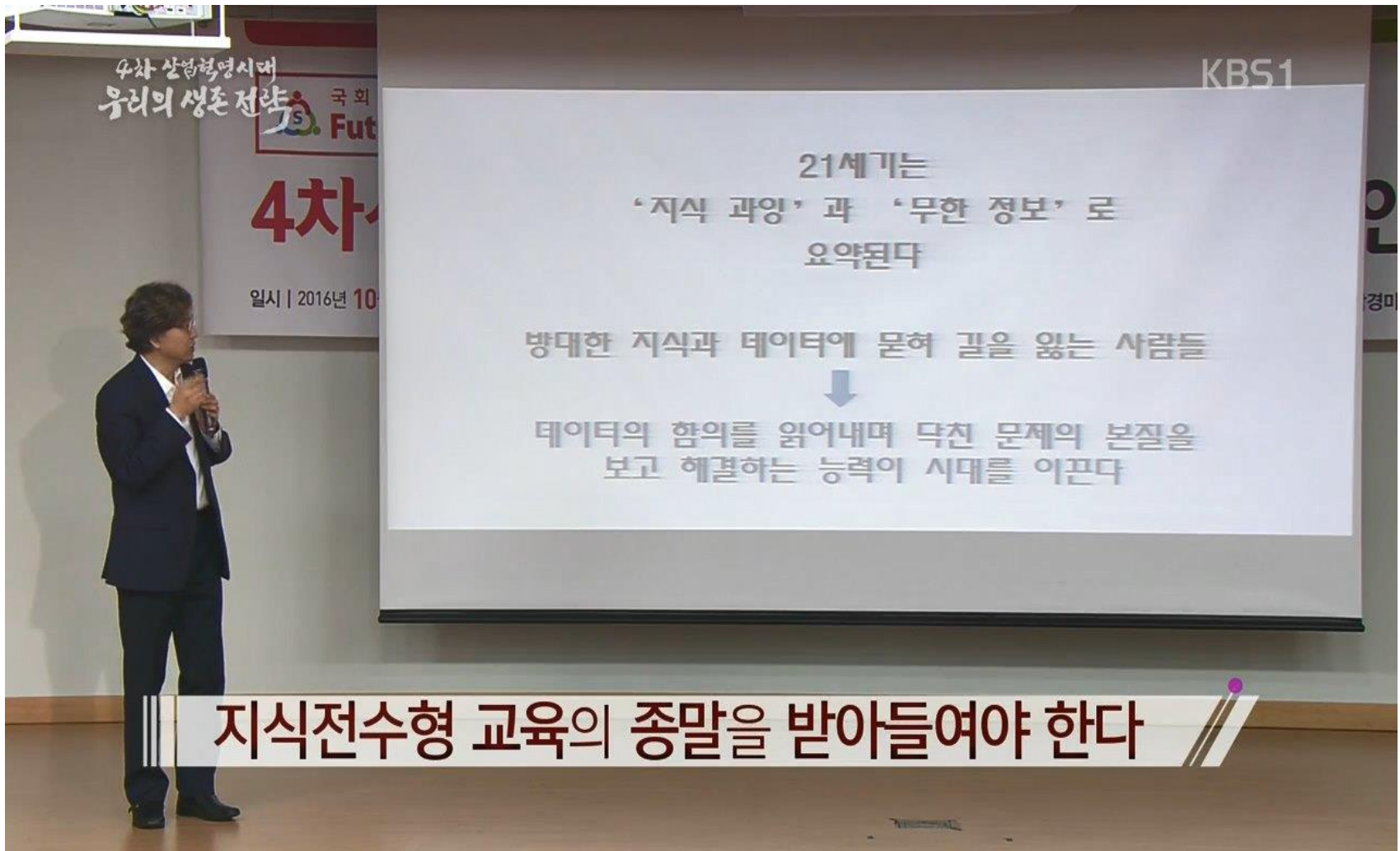
지난 4월, KB국민카드는 IT 직무 관련한 신입 사원을 공개 채용했다. 모집 인원은 비록 7명 내외에 불과했지만 학력, 성적, 성별, 나이 등에 관계없이 IT관련 자격증 보유자면 누구나 지원할 수 있도록 했다. KB국민카드가 수혈하고 싶은 IT인력은 빅데이터 분야의 전문가들이다. KB국민카드 관계자는 “외부 전문가들을 이런 방식으로 파격적으로 수혈하지만 IT부문은 처음”이라고 밝혔다.

국민카드 뿐만 아니라 금융권 전체가 디지털금융 전략 구현의 성패가 빅데이터에 달렸다고 보고, 다양한 형태로 외부 전문가 충원을 강화하고 있다. 내부 IT인력들도 빅데이터 분야 전문가들이 있지만 회사 입장에서선 더 차별화된 빅데이터 전략을 원하고 있는 것이다.

앨빈 토플러 2006년 한국 방문시 한마디...



현재 우리 학생들은...



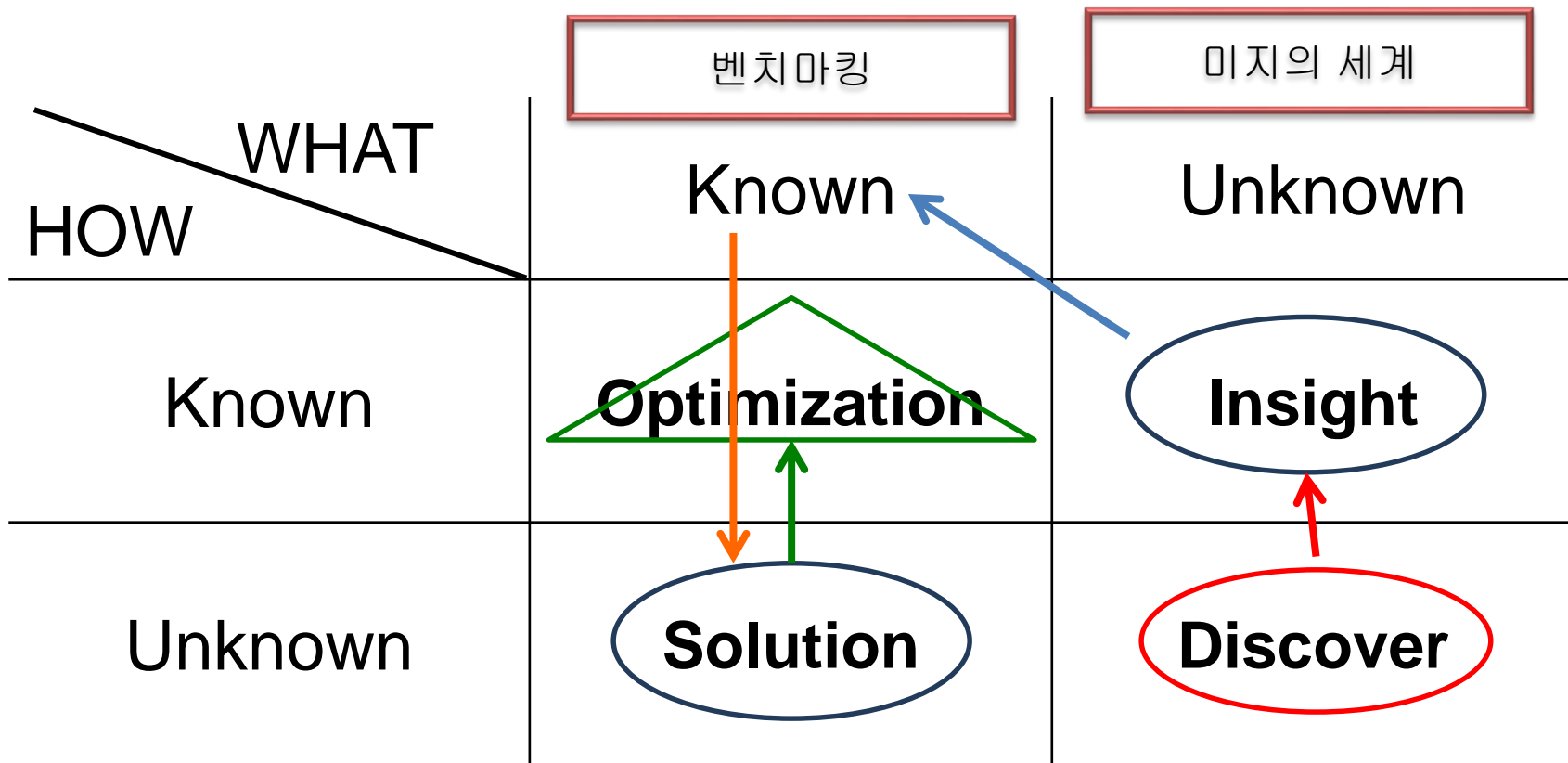
왜냐면? 지금 현재

유니
의 힘

필요한 건 지식의 양이 아니라

생각의 힘

시대의 흐름을 읽고
중요한 질문을 하며
필요할 때 답을 찾을 수 있는 능력



Computational thinking

decomposition

solve a problem
by **breaking** it
into smaller groups

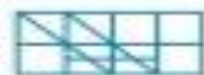


pattern
recognition

find the **order**



analyze the data



algorithmic
design

creating solutions
using a series
of ordered **STEPS**

