

# ADsP 데이터 분석 준전문가

# 데이터분석 준전문가란?

데이터 이해 및 처리 기술에 대한 기본지식을 바탕으로 데이터분석 기획, 데이터분석, 데이터 시각화 업무를 수행하고 이를 통해 프로세스 혁신 및 마케팅 전략 결정 등의 과학적 의사결정을 지원하는 직무를 수행하는 전문가를 말한다.

# 데이터분석 준전문가 직무

직무	세부내용
데이터 기획	비즈니스 목표 달성을 위해 내부 업무 프로세스를 기반으로 다양한 분석기회를 발굴하여 분석의 목표를 정의하고, 분석대상 도출 및 분석 결과 활용 시나리오를 정의하여 분석과제를 체계화 및 구체화하는 빅데이터분석과제 정의, 분석로드맵 수립, 성과 관리 등을 수행한다.
데이터분석	분석에 대한 요건을 구체적으로 도출하고, 분석과정을 설계하고, 요건을 실무담당자와 합의 하는 요건정의, 모델링, 검증 및 테스트, 적용 등을 수행한다.
데이터 시각화	다양한 데이터들을 대상으로 어떤 요소를 시각화 해야 하는지 정보 구조를 분석하며 어떤 형태의 시각화 모델이 적합한지 시각화에 대한 요건을 정의하고 시나리오를 개발하는 시각화 기획, 모델링, 디자인, 구축, 배포 및 유지보수 등을 수행한다.

# 시험 일정

구분	회차	접수기간	수험표발급	시험일	사전점수공개 및 재검토 접수	결과발표
데이터 분석 준전문가	제40회	1.22~26	2.8	2.24(토)	3.15~19	3.22
	제41회	4.8~12	4.26	5.11(토)	5.31~6.4	6.7
	제42회	7.1~5	7.26	8.10(토)	8.30~9.3	9.6
	제43회	9.30~10.4	10.18	11.3(일)	11.22~26	11.29

# 시험 개요

구분	과목명	문항수	배점	검정시험시간
		객관식	객관식	
필기	데이터 이해	10	100 (각 2점)	90분 (1시간30분)
	데이터분석 기획	10		
	데이터분석	30		
	계	50	100	

# 1과목

## 데이터 이해

# 데이터의 유형

❖ 정성적 데이터: 언어·문자로 표현 (ex. 회사 매출이 증가함)

↳ 저장·검색·분석에 당연히 많은 비용이 소모

❖ 정량적 데이터: 수치·도형·기호로 표현 (ex. 나이, 몸무게, 주가 등)

↳ 정형화된 데이터로 비용소모가 적다.

# 지식 경영의 핵심 이슈

❖ 암묵지: 학습과 경험을 통해 개인에게 체화되어 있어 겉으로 드러나지 않음

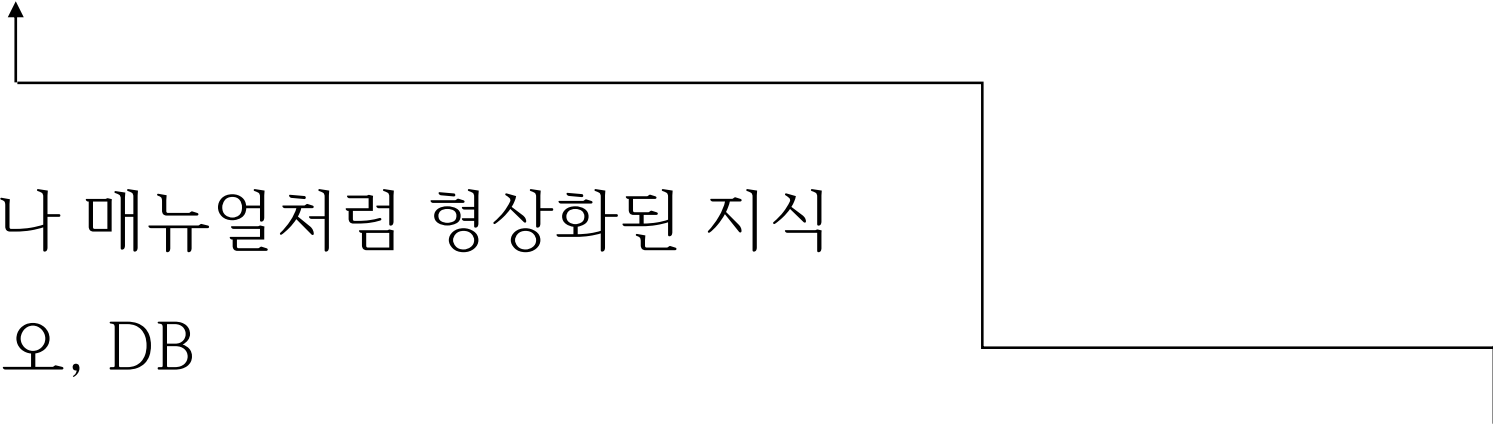
↳ ex) 김치 담그기, 자전거 타기

↳ 개인 체득(내면화) → 타인에게 대화를 통해 공유(공동화)

❖ 형식지: 문서나 매뉴얼처럼 형상화된 지식

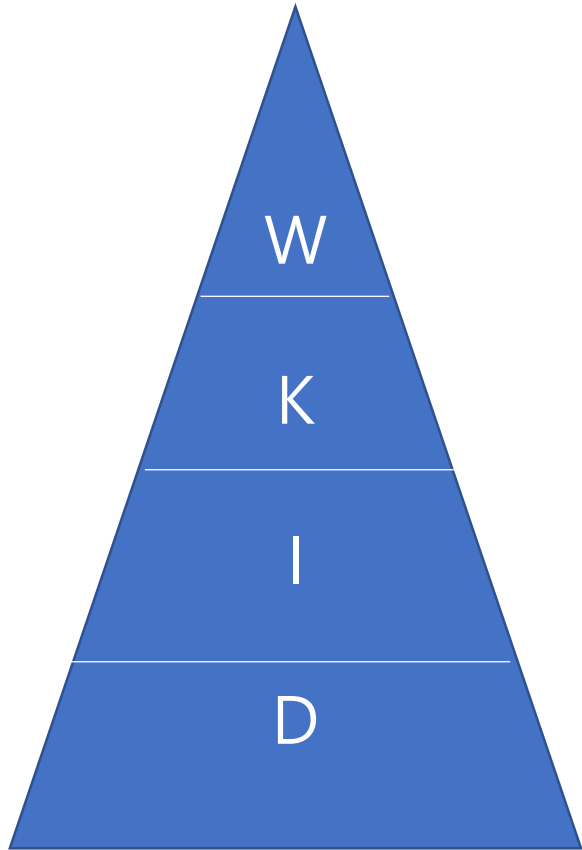
↳ 교과서, 비디오, DB

↳ 개인의 지식을 책 등으로(표출화) → 책에 내가 아는 지식 추가(연결화)





# DIKW 피라미드



- ❖ W(Wisdom) 지혜: 지식 + 아이디어, 창의력
- ❖ K(Knowledge) 지식: 정보를 바탕으로 예측 및 결과 도출
- ❖ I(Information) 정보: 데이터간 연관관계 속에서 의미 도출
- ❖ D(Data) 데이터: 객관적인 사실, 수치 & 기호

# 데이터베이스의 특징

- ❖ 통합된 데이터 : 데이터 중복 X
- ❖ 저장된 데이터 : 컴퓨터가 접근할 수 있는 저장 매체에 저장
- ❖ 공용 데이터 : 여러 사용자가 데이터를 공동으로 이용
- ❖ 변화되는 데이터 : 삽입, 삭제, 갱신이 자유롭게 가능, 항상 데이터베이스의 특징은 유지해야 함
- ❖ 데이터베이스 설계 절차
  - ↳ 요구사항 분석 → 개념적 설계 → 논리적 설계 → 물리적 설계

# 기업 내부 데이터베이스

## ❖ OLTP(On-Line Transaction Processing)

- ↳ 온라인으로 접속된 여러 단말 간의 처리 형태
- ↳ 실시간으로 호스트 컴퓨터가 데이터베이스에 접근할 수 있음

## ❖ OLAP(On-Line Analytical Processing)

- ↳ 다차원의 데이터에 접근하여 의사결정에 활용할 수 있는 정보를 얻음.
- ↳ '분석' 키워드

# 기업 내부 데이터베이스

❖ CRM(Customer Relationships Management: 고객 관계 관리)

↳ **고객관계관리**과 관련된 내·외부 자료를 분석해 고객 중심 자원을 극대화하고 이를 토대로 고객의 특성에 맞게 마케팅

❖ SCM(Supply Chain Management: 공급망 관리)

↳ **공급망** 단계를 **최적화**해 고객이 원하는 제품을 시간과 장소에 맞게 제공

# 제조분야 데이터베이스

## ❖ ERP(Enterprise Resource Planning)

↳ 기업 전체를 경영자원의 효과적 이용이라는 관점에서 통합적으로 관리하고 경영의 효율화를 기하기 위한 시스템

## ❖ BI(Business Intelligence)

↳ 기업이 보유하고 있는 데이터를 정리하고 분석해 의사결정에 활용

## ❖ RTE(Real-Time Enterprise)

↳ 실시간 기업의 새로운 기업경영시스템

# 금융분야 데이터베이스

## ❖ EAI(Enterprise Application Intergration)

↳ 기업 내 어플리케이션을 유기적으로 연동, 정보를 중앙 집중적으로 통합 관리 및 사용

## ❖ EDW(Enterprise Data Warehouse)

↳ 기존 DW를 전사적으로 확장한 모델

↳ 기업 리소스의 유기적 통합

# 유통분야 데이터베이스

❖ KMS(Knowledge Management System)

↳ 지식관리 시스템

❖ RFID(Radio Frequency ID)

↳ 주파수를 이용해 ID를 식별하는 시스템

# 사회기반 데이터베이스

## ❖ EDI(Electronic Data Interchange)

↳ 각종 서류를 표준화된 양식을 통해 전자적 신호로 바꿔 거래처에 전송하는 시스템

## ❖ VAN(Value Added Network)

↳ 부가가치 통신망, 독자적인 네트워크를 형성하는 것

## ❖ CALS(Commerce At Light Speed)

↳ 경영통합정보시스템, 전자상거래 구축을 위해 제품의 라이프사이클 관리.



# 빅데이터

❖ 빅데이터 -> '큰 데이터'

❖ 빅데이터를 나타내는 4V  가트너 그룹의 더그 래니의 3V

↳ Volume(양), Variety(다양성), Velocity(속도), Value(가치)

↳ 7V : 4V + Veracity(진실성), Validity(정확성), Volatility(휘발성)

❖ 클라우드 컴퓨팅 기술은 빅데이터 분석에 경제성을 제공

# 빅데이터에 거는 기대 비유

- ❖ 산업혁명의 석탄, 철 : 혁명적 변화
- ❖ 21세기 원유 : 생산성을 향상시키고 기존에 없던 새로운 범주의 산업 형성
- ❖ 렌즈 : 렌즈를 통해 현미경이 생물학 발전에 미친 영향
- ❖ 플랫폼 : 공동 활용의 목적으로 구축된 유무형의 구조물로서 역할

# 과거에서 현재로의 변화

- ❖ 사전처리 → 사후처리 : 일단 모든 데이터를 수집하고 분석
- ❖ 표본조사 → 전수조사 : 데이터 처리비용 감소, 모든 데이터 처리 가능
- ❖ 질 → 양 : 일단 데이터를 많이 수집하면, 오류정보보다 양질의 정보가 많아 유용할 것
- ❖ 인과관계 → 상관관계 : 데이터간 상관관계를 분석

# 빅데이터 가치선정의 어려움

- ❖ 데이터를 재사용하거나 재조합해 활용하면서 특정 데이터를 언제·어디서·누가 활용할지 알 수 없게 되어 가치 산정 어려움
- ❖ 기술이 발전하면서 기존에 없던 가치를 창출
- ❖ 현재는 가치가 없는 데이터라도 나중에 기술이 발전하면 가치가 있는 데이터가 될 수도 있음

# 빅데이터를 활용한 기본 테크닉

❖ 연관규칙학습 : 상관관계 찾기(장바구니 학습)

↳ 커피를 구매하는 사람이 탄산음료를 더 많이 사는가?

❖ 유형분석 : 문서, 조직을 특성에 따라 분류

↳ 이 사용자는 어떤 특성을 가진 집단에 속하는가?

❖ 유전자 알고리즘 : 자연선택, 돌연변이 등을 통해 진화시키는 방법

↳ 최대의 시청률(최적화, 최적해)을 얻기 위한 프로그램 배치

❖ 기계학습 : 훈련 데이터로부터 학습한 특성을 통해 예측

↳ 기존 시청 기록을 바탕으로 새로운 영상 추천

# 빅데이터를 활용한 기본 테크닉

- ❖ 회귀분석 : 독립변수와 종속변수 간의 관계 파악
  - ↳ 구매자의 나이가 구매 차량 타입에 어떤 영향을 끼치는가?
- ❖ 감정분석 : 특정 주제에 대한 글을 쓴 사람의 감정을 분석
  - ↳ 고객의 평가나 리뷰를 통한 분석
- ❖ 소셜 네트워크 분석 : 군집에서 영향력이 가장 큰 사람 찾기 등
  - ↳ 사람과 사람사이의 **관계망** 분석

# 빅데이터 위기 요인

❖ 사생활 침해: 개인정보 유출, 사생활 침해

↳ 개인정보 사용을 **제공자의 동의에서 사용자의 책임으로**

❖ 책임 원칙 훼손: 범죄를 저지르지 않은 사람을 체포하면 문제가 될 것

↳ 명확히 나타난 결과에 대해서만 책임을 물어야함 (**결과기반책임원칙**)

❖ 데이터 오용: 데이터 분석 결과가 항상 옳은 것은 아니다.

↳ 알고리즘 접근 하용, **전문가(알고리즘미스트)**가 필요해짐

# 빅데이터를 활용하기 위한 3요소

## ❖ 데이터

↳ 빅‘데이터’, 모든 것의 데이터화

## ❖ 기술

↳ 빅데이터 분석을 위한 기술의 출현

## ❖ 인력

↳ 데이터 사이언티스트, 알고리즘미스트



# 빅데이터 회의론

## ❖ 투자효과를 거두지 못했던 부정적 경험

- ↳ 도입만 하면 문제를 한번에 해결할 것처럼 강조
- ↳ 빅데이터 활용 및 인사이트 도출 실패

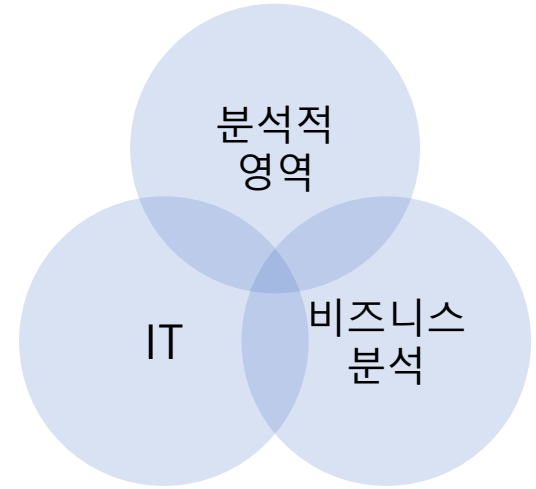
## ❖ 기존 분석 프로젝트를 빅데이터 분석으로 선전

- ↳ CRM 분석 성과를 빅데이터 분석으로 과대 포장

# 데이터 사이언스

❖ 데이터로부터 의미 있는 정보를 추출해 내는 학문

↳ 정형·비정형을 막론하고 데이터 분석 (총체적 접근법)



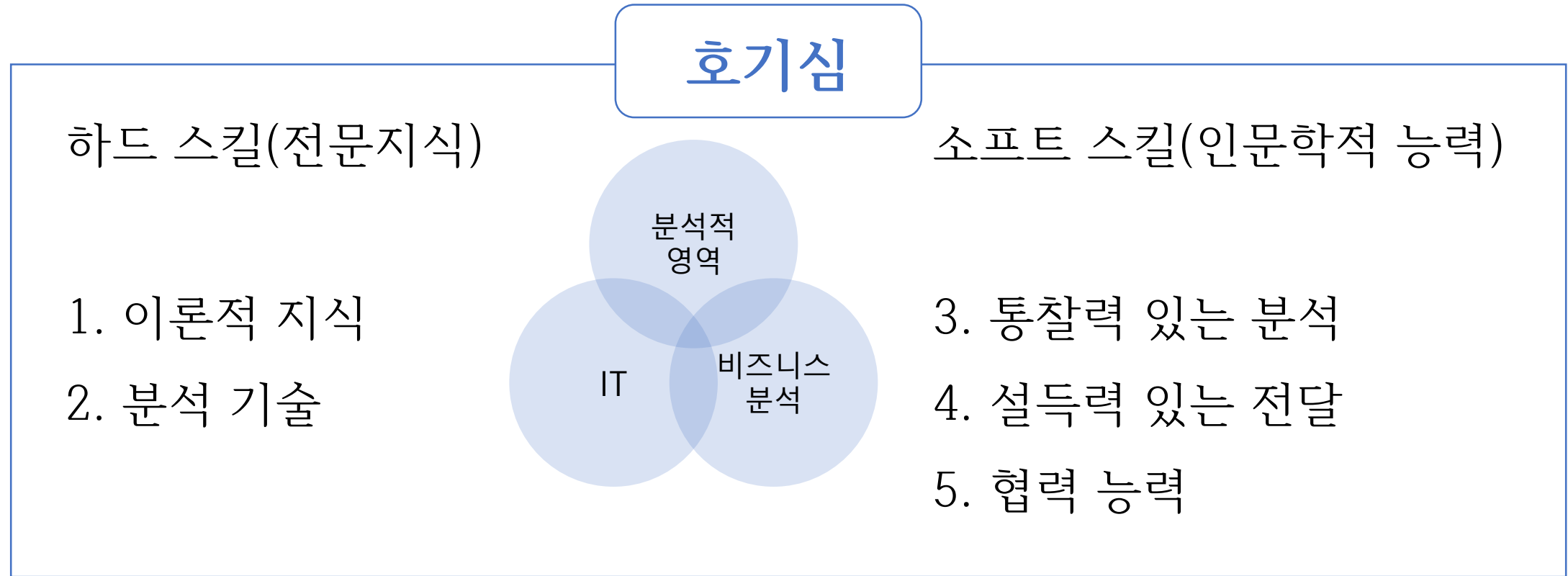
❖ 데이터 사이언스 영역

↳ 분석적 영역: 수학, 확률모델, 분석학 등등 과 같은 이론적인 지식

↳ IT: 프로그래밍, 데이터 엔지니어링, 고성능 컴퓨팅 과 같은 프로그래밍적 지식

↳ 비즈니스 분석: 커뮤니케이션, 시각화, 프레젠테이션 과 같은 비즈니스적 능력

# 데이터 사이언티스트 요구 역량



# DBMS(DataBase Management System)

- ❖ 데이터베이스 관리 시스템 : 오라클, 인포믹스, 액세스 등
- ❖ 관계형 DBMS : 우리가 자주 사용한 Table형식 DB
- ❖ 객체지향 DBMS : 데이터를 '객체'의 형식으로 표현되는 DB
- ❖ 네트워크 DBMS : 노드와 간선으로 표현되는 '그래프' 기반 DB
- ❖ 계층형 DBMS : 트리 구조를 기반으로 하는 DB

# SQL(Structured Query Language)

데이터 정의어 (DDL)	CREATE	Schema, Domain, Table, View, Index를 정의
	ALTER	Table에 대한 정의를 변경하는 데 사용
	DROP	Schema, Domain, Table, View, Index를 삭제
	TRUNCATE	테이블의 모든 행(row)의 데이터 및 행 삭제
데이터 조작어 (DML)	SELECT	테이블에서 조건에 맞는 튜플을 검색
	INSERT	테이블에 새로운 튜플을 삽입
	DELETE	테이블에서 조건에 맞는 튜플을 삭제
	UPDATE	테이블의 조건에 맞는 튜플의 내용을 변경함
데이터 제어어 (DCL)	COMMIT	데이터베이스 조작 작업이 정상적으로 완료되었음을 관리자에게 알려줌
	ROLLBACK	데이터베이스 조작 작업이 비정상적으로 종료되었을 때 원래의 상태로 복구
	GRANT	데이터베이스 사용자에게 사용권한을 부여
	REVOKE	데이터베이스 사용자의 사용권한을 취소

# 개인정보 비식별 기술

데이터 마스킹	데이터의 길이, 유형, 형식과 같은 속성을 유지한 채, 새롭고 읽기 쉬운 데이터를 익명으로 생성하는 기술 ex) 홍**, 이**, 010-4553-****
가명처리	개인 식별에 중요한 데이터를 식별할 수 있는 다른 값으로 변경하는 기술 ex) 홍길동, 35세, 한국대 재학 -> 임걱정 30대, 국내대 재학
총계처리	데이터 총계 합을 보냄으로써 개인 데이터의 값이 보이지 않도록 하는 기술 ex) 홍길동 180cm, 임걱정 154cm ... 평균키 160cm
데이터값 삭제	데이터 공유, 개방 목적에 따라 데이터셋에 구성된 값 중에 필요 없는 값 또는 개인식별에 중요한 값을 삭제 ex) 홍길동, 35세, 남, 서울 거주, 한국대 재학 -> 35세, 남, 한국대 재학
데이터 범주화	데이터 값을 범주의 값으로 변환하여 값을 숨김 ex) 홍길동, 35세 -> 홍씨 30~40세

문제풀이

# 1과목 문제풀이 정답

1	③	11	①	21	①	31	③	41	②
2	③	12	④	22	④	32	③	42	①
3	②	13	②	23	③	33	①	43	④
4	③	14	④	24	①	34	④	44	④
5	③	15	①	25	②	35	④	45	④
6	③	16	④	26	②	36	①	46	②
7	④	17	③	27	③	37	②	47	③
8	②	18	④	28	②	38	①	48	③
9	②	19	①	29	②	39	④	49	④
10	①	20	①	30	②	40	③	50	④