1과목 데이터 이해 10Q 2주관식



데이터

데이터: 있는 그대로의 객관적 사실, 가공되지 않은 상태 (주문수량)

정보: 데이터로부터 가공된 자료(베스트셀러)

데이터의 유형

(1)정성적, 정량적

정성적 데이터: 자료의 특징을 풀어 설명(기상특보, 주관식 설문응답)



정량적 데이터: 자료를 수치화(수치, 기호)

(2)정형, 반정형, 비정형

정형데이터: 정보 형태가 정해짐(엑셀,스프레드시트)



반정형데이터: 데이터를 설명하는 메타데이터를 포함(로그, html, xml)



비정형데이터: 형태가 정해지지 않음(sns, youtube, music)

암묵지, 형식시간 상호작용

암묵지: 개인에게 슥듭되고 겉으로 드러나지 않음

형식지: 문서, 메뉴얼 등의 형상화된 지식

공통화: 암묵지 지식을 다른 사람에게 알려줌

표출화: 암묵지 지식을 메뉴얼이나 문서로 전환

연결화: 교재, 메뉴얼에 새로운 지식 추가

내면화: 만들어진 교재, 메뉴얼에서 다른 사람의 암묵지를 터득

(공표연내)

DIKW 피라미드

데이터 Data : 있는 그래도릐 사실(a대리점 휴대폰 100만원 b대리점 휴대폰 200만원)

정보 Information: data를 통해 패턴 인식(a대리점의 휴대폰이 더 싸다)

지식 Knowledge: 패턴을 통해 예측 (a대리점에서 휴대폰을 사면 이득을 본다)

지혜 Wisdom: 창의적인 산물(a대리점에서 다른 휴대폰들도 b대리점보다 싸다)

데이터의 단위

KB < MB < GB < TB < PB < EB< ZB< YB (PETA < EXA < ZETTA < YOTTA)

(패지요 순서!)

데이터베이스 정의

1. DB: 일정 구조에 맞게 조직화된 데이터의 집합

스키마: DB의 구조와 제약조건에 관한 전반적 명세

인스턴스: 정의된 스키마에 따라 지정된 값

1. DBMS: DB를 관리, 접근 환경 제공하는 소프트웨어

관계형 DBMS 테이블 표로 정리

(오라클, mysql, maria db)

객체지향 DBMS: 정보를 객체형태로 정리

NoSQL DBMS 비정형 데이터를 저장하고 처리

(HBASE, Mongo, Dyamo DB)

1. SQL 데이터 베이스에 접근할 수 있는 하부언어

정의언어 (DDL): CREATE, ALTER, DROP

조직언어 (DML): SELECT, INSERT, DELETE, UPDATE

제어언어 (DCL): COMMIT, ROLLBACK, GRANT, REVOKE

데이터베이스의 특징

공용데이터: 여러 사용자가 다른 목적으로 데이터 공동 이용

통합된 데이터: 동일한 데이터 중복되어 있지 않음

저장된 데이터: 저장매체에 저장

변화되는 데이터: 새로운 데이터 추가, 수정, 삭제에도 현재의 정확한 데이터 유지

(공동저변)

데이터베이스 설계 절차

요구조건분석

개념적 설계

논리적 설계

물리적 설계

(개논물)

기업활용 데이터베이스

OTL 데이터를 수시로 갱신

OLAP 디자인 데이터를 대화식으로 분석

CRM 고객과 관련 자료 분석, 마케팅 활용

SCM 공급망 연결 최적화

ERP 기업 경영 자원을 효율화

BI 기업 보유 데이터 정리, 분석하여 리포트 중심 도구

BA 통계 기반 비즈니스 통찰력

Block Chain 네트워크 참여한 모든 사용자가 정보를 분산, 저장

KMS 기업의 모든 지식을 포함

Data Ware House(DW)

주제 지향성: 분석목적 설정이 중요

데이터 통합: 일괄화 된 형식으로 저장

시계열성: 히스토리를 가진 데이터

비방향성: 읽기전용, 수시로 변하지않음

ETL DW와 DM 등의 DB시스템에 데이터 적제

Data Lake

비정형 데이터를 저장하며 하둡과 연계하여 처리

하둡: 병렬처리 오픈소스 프레임워크

빅데이터의 이해

출현 배경: 인터넷 확산, 경제성 확보, 하둡을 이용한 분산 컴퓨팅, 비정형 데이터 확산

빅데이터의 3V(가트너 정의)

Volume 규모: 데이터 양 증가( 구글 번역 서비스)

Vairety 다양성: 데이터 유형증가

Velocity 속도: 데이터 생성 처리 속도증가

그외

Value 가치, Veracity 신뢰성

빅데이터에 대한 비유

산업혁명의 석탄, 철

원유

렌즈: 현미경이 생물학 발전 영향,

플랫폼: 공동 활용 목적으로 구축된 구조물

빅데이터가 만들어낸 변화

표본조사 -> 전수조사

사전처리 -> 사후처리

질 -> 양

인과관계 -> 상관관계

(전후양상)

비즈니스 모델

빅데이터 활용 위한 3대요소

인력, 자원(데이터), 기술

(인자기)

데이터 사이언스

정형/비정형 데이터를 막론하고 데이터를 분석 총체적 접근법

데이서 사이언스 핵심 구성요소

Analyitcs 이론적 지식

IT 프로그래밍적 지식

비즈니스 분석: 비즈니스적 능력

(AI비)

데이터 사이언티스트의 필요역량

하드스킬: 이론적 지식(수학,통계 등) 가트너 제시 역량이 미포함

소프트 스킬: 스토리텔링, 리더십, 창의력, 분석 등

(하드는 이과적, 소프트는 문과적 느낌)

빅데이터 가치 패러다임 변화

1) Digitalization 아날로그 세상을 디지털화

2) Conenction 디지털화된 정보들의 연결

3) Agency 연결을 효과적으로 관리

순서 (DigitaCA메라)