

1장. 데이터베이스



SQL

컴퓨터와의 대화

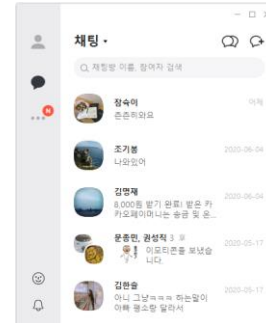
- 초고도 정보화 사회이다.
- 패스트푸드점에서 햄버거를 주문하면 POS에서 모든 것이 처리된다.



- 스마트폰으로 모든 정보를 얻을 수 있다.

한국거래소 (KOSX) - 실시간 시장 정보

종목	종가	전일 대비	거래량	시가총액
0055 삼성전자	41,100	↑	1,200,000	15,000,000,000,000
0054 SK하이닉스	12,500	↓	800,000	10,000,000,000,000
0052 현대자동차	18,000	↑	500,000	8,000,000,000,000
0051 LG전자	15,000	↑	400,000	6,000,000,000,000
0053 삼성전자	41,100	↑	1,200,000	15,000,000,000,000
0054 SK하이닉스	12,500	↓	800,000	10,000,000,000,000
0052 현대자동차	18,000	↑	500,000	8,000,000,000,000
0051 LG전자	15,000	↑	400,000	6,000,000,000,000



게시판

검색엔진

주식거래

메신저

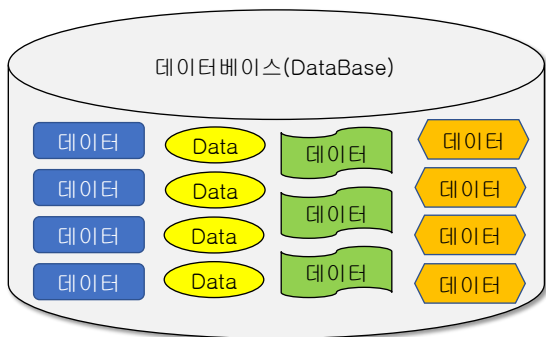
인터넷 신문

- 이런 모든 것이 가능해진 기술적 배경이 바로 데이터베이스이다.



SQL 데이터베이스의 요건

- 데이터베이스(DataBase)는 컴퓨터의 기억 능력을 십분 활용하여 자료를 가공 및 저장, 활용하는 일체의 기술이다.



- 자료(Data)를 활용하려면 합산, 집계 등의 알고리즘을 적용하여 정보(Information)로 가공해야 한다.

자료(Data)					정보(Information)			
학생	국어	영어	수학		학생	총점	평균	석차
1번	85	92	76	→ 집계, 가공 →	1번	253	84.3	2
2번	100	80	86		2번	266	88.6	1
3번	88	75	82		3번	245	81.6	3
⋮								

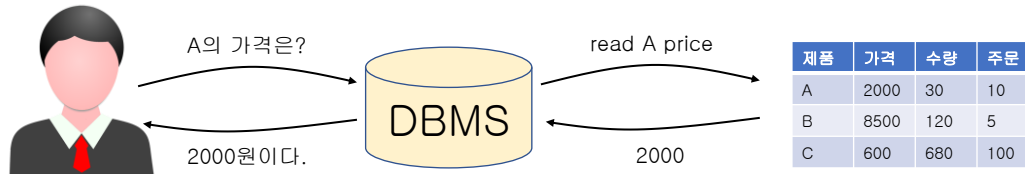


SQL 데이터베이스의 요건

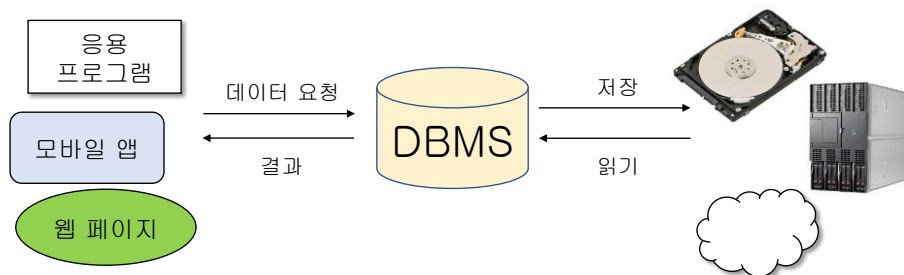
- 현대적인 데이터베이스가 갖추어야 할 요건
 - 대용량
 - 효율성
 - 무결성
 - 활용성
 - 공유성
 - 보안성
- 거대한 데이터를 저렴하고 완벽하게 저장하며 언제든지 누구나 활용할 수 있으면서도 안전해야 한다.
- 고도의 기술을 요하며 수십년에 걸친 연구, 개발이 필요하다.
- 데이터베이스 시스템은 비싸고 그 기술을 다 배우고 제대로 활용하는 것도 만만치 않다.

SQL DBMS

- DBMS(DataBase Management System : 데이터베이스 관리 시스템)는 정보의 저장과 관리를 전담하는 특수한 소프트웨어



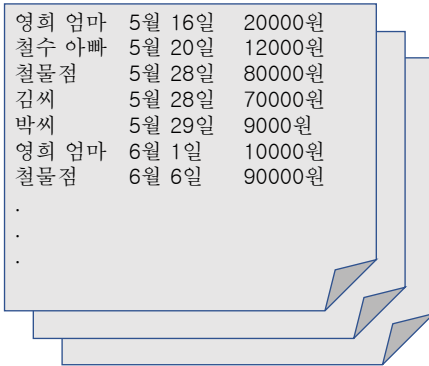
- 데이터 관리를 DBMS가 전담하면 프로그램과 데이터가 분리되고 종속성이 제거된다.
- 약간의 성능 저하가 발생하는 대신 유연성이 극적으로 증가한다.
- DBMS에 의해 응용 프로그램과 데이터가 다:다 양쪽으로 연결된다.



- 요구가 더 복잡해져 응용 프로그램이 해야 할 잡스러운 작업까지 대신한다.

SQL 역사

- 가장 고전적인 방법은 종이에 적어 놓는 것이다. 검색이나 수정은 어려워 활용성이 떨어진다.

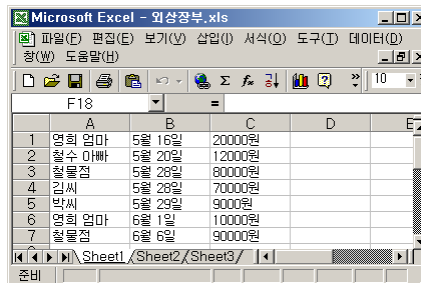


영희 엄마	5월 16일	20000원
철수 아빠	5월 20일	12000원
철물점	5월 28일	80000원
김씨	5월 28일	70000원
박씨	5월 29일	9000원
영희 엄마	6월 1일	10000원
철물점	6월 6일	90000원
.		
.		

- 파일 시스템 : 메모장에 텍스트 형태로 기록하거나 엑셀을 사용한다. 저장할 뿐 관리 능력이 미흡하고 여러 사람이 동시에 사용할 수 없다.



영희 엄마	5월 16일	20000원
철수 아빠	5월 20일	12000원
철물점	5월 28일	80000원
김씨	5월 28일	70000원
박씨	5월 29일	9000원
영희 엄마	6월 1일	10000원
철물점	6월 6일	90000원



	A	B	C	D	E
1	영희 엄마	5월 16일	20000원		
2	철수 아빠	5월 20일	12000원		
3	철물점	5월 28일	80000원		
4	김씨	5월 28일	70000원		
5	박씨	5월 29일	9000원		
6	영희 엄마	6월 1일	10000원		
7	철물점	6월 6일	90000원		

SQL 역사

- SAM은 정보를 파일에 순서대로 쟁여 놓는 방법이며 순차 검색만 가능하다.
- ISAM은 정보의 순서를 기억하는 인덱스가 있어 이분 검색이 가능하다.

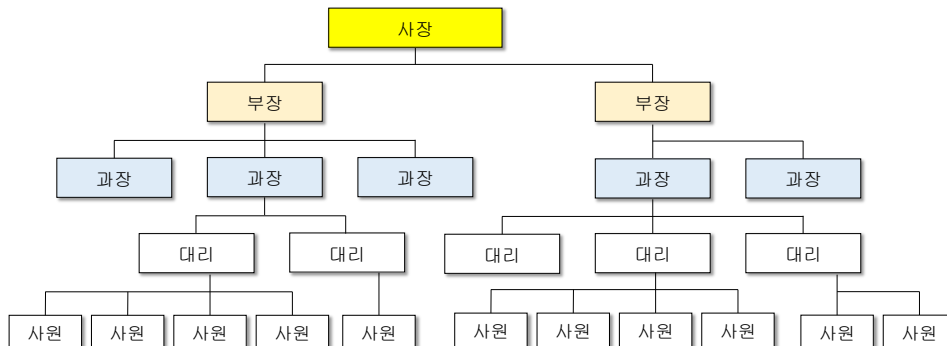
이름	나이	주소
이승면	58	미국
박쟁희	52	서울
전두한	36	부산
노때우	24	광주
김공삼	40	대전
김태중	45	춘천
노무현	48	인천
이명박	36	전주
박그네	32	제주
문대인	56	서울

SAM

인덱스	이름	나이	주소
1	이승면	58	미국
2	박쟁희	52	서울
3	전두한	36	부산
4	노때우	24	광주
5	김공삼	40	대전
6	김태중	45	춘천
7	노무현	48	인천
8	이명박	36	전주
9	박그네	32	제주
10	문대인	56	서울

ISAM

- 계층형, 네트워크형 : 복잡도에 비해 실용성이 떨어져 특수한 분야에만 가끔 사용한다.





SQL 역사

- 관계형 데이터베이스(Relational Database)는 1969년 Edgar.F.Codd 박사의 논문을 기반으로 탄생
- 모든 데이터를 표 형태의 테이블에 저장하며 관계를 정의한다.

이름	부서	입사일	월급	부서명	책임자	사무실
김상영	영업부	2015.6.25	620	영업부	구홍녀	A동 3층
김영주	영업부	2017.4.13	540	인사과	김정수	B동 2층
문병대	인사과	2012.3.1	550	오락부	김경선	A동 1층
김규민	오락부	2019.8.1	340			
문여울	인사과	2018.9.12	490			
문한울	오락부	2020.10.25	430			

- 유연하여 로직 변화에 신속히 대처할 수 있다. 자원 소모가 많으며 고성능 하드웨어와 고급 튜닝 인력이 필요하다.
- 객체지향형 : 모든 것을 객체로 저장한다. 성능상의 열세를 극복하지 못했고 실용적으로 활용할 분야가 드물어 아직 연구 단계이다.
- 빅데이터 : 현대의 데이터는 대용량(Volume)인데다 형태도 다양하고(Variety) 생성 주기가 빨라(Velocity) RDB로는 수집, 저장, 분석이 어려운 지경에 이르렀다.
- 일관성을 약간 희생하더라도 성능과 용량을 극적으로 향상시킨 빅데이터 기법이 대두되었다. Hadoop 솔루션, 비정형 데이터를 다루는 NoSQL(Not Only SQL) 문법, 데이터마이닝이나 인공지능, 딥러닝을 활용한 분석



SQL 종류

- DBMS의 종류는 우리가 늘상 먹는 라면만큼이나 다양하다.
- Oracle : 기능적 완성도가 높고 시장 점유율도 높다. 지존
- SQL Server : 윈도우 환경에서 탁월한 성능을 발휘하며 닷넷과의 통합성도 우수하다.
- MySQL : 오픈소스, 다양한 운영체제를 지원하며 표준 SQL을 준수하여 학습용으로 적합하다. PHP와 궁합이 잘 맞아 웹 게시판 제작 용도로 많이 활용한다.
- DB2 : IBM 제품. 오라클 다음으로 시장 점유율 2위를 기록하고 있다.
- PostgreSQL : 오픈소스이며 객체 관계형 DBMS이다. 일본에서 인기가 많다.

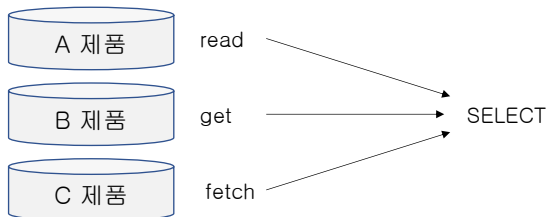
순위	DBMS	점수
1	오라클	1350
2	MySQL	1270
3	MS SQL Server	1070
4	PostgreSQL	522
5	MongoDB	437
6	IBM DB2	161

- 부동의 1위는 오라클이고 MSSQL, MySQL, DB2가 각축을 벌이고 있다.



SQL 표준어

- 초창기의 데이터베이스 제품은 제조사마다 구조가 독특해 관리 방법이 제각각이었다. 데이터를 읽는 기본적인 명령조차 제품마다 달랐다.



- 제품을 바꿀 때마다 다시 배워야 한다. 사회적 낭비를 초래한다.
- 공통적인 표준 언어로 탄생한 것이 SQL이다. 제품마다 구조나 관리 방식이 달라도 SQL을 통하면 똑같은 방법으로 다룰 수 있다.
 - 키가 170 이상인 회원의 목록을 조사하라.
 - 이번달 매출 합계와 평균을 조사하라.
 - 모든 상품의 가격을 20% 인하한 값으로 변경하라.
- 자연어로 되어 있지는 않으며 고유의 문법 체계를 가지고 있다.
 - `SELECT * FROM tMember WHERE height >= 170;`
- SQL 문도 의사 소통을 위한 언어라는 면에서 자연어와 동질성이 있다.



SQL의 역사

- SQL의 시초는 1970년에 IBM의 시스템 R에서 도입한 SEQUEL(Structured English Query Language)
- 상표권 분쟁으로 인해 SQL(Structured Query Language)로 이름 변경
- 쿼리(Query)는 DBMS에게 요청한다는 뜻이며 한국말로 "질의"로 번역한다.



Structured English Query Language

구조적

질의

언어

- 일시적인 통일을 이룬 후 경쟁에 의해 방언(Dialect)이 생기기 시작했다.
- 공신력 있는 국제 표준 단체에서 SQL의 표준을 만들고 관리한다.

SQL 표준	특징
SQL86	ANSI에서 제정한 최초의 표준
SQL92	대규모 개정 및 정리. 실질적인 첫 표준
SQL99	정규 표현식, 트리거, 절차적 흐름.
SQL2003	XML 관련 기능 추가. 시퀀스 생성기. MERGE 구문 추가
SQL2008	INSTEAD OF 트리거 추가. TRUNCATE 구문 추가
SQL2011	임시 데이터베이스 지원
SQL2016	JSON 지원. 행 패턴 인식. DECFLOAT 타입 추가

- SQL99가 실질적인 표준이다. 제조사마다 확장 문법을 제공한다.



SQL의 특징

- **SQL은 대화식 언어이다. 단편적인 질문 위주여서 명령문이 짧고 간결하다.**
 - 가장 많은 월급은 얼마인가? → 520만원이다.
 - 520만원을 받는 직원의 이름은 무엇인가? → 홍길동이다.
 - 홍길동은 어느 부서 소속인가? → 영업부이다.
- **단순한 명령을 조합하여 복잡한 명령을 처리한다. 가급적 한번에 처리하는 것이 좋다.**
 - 월급을 가장 많이 받는 직원이 속한 부서는?
- **제어문이 빈약해 C#, 자바 같은 고수준 언어와 함께 사용한다.**
- **둘 째로, SQL은 선언적인 언어이다. 구체적인 절차를 일일이 명령으로 기술하는 것이 아니라 원하는 것만 밝힌다.**
 - 부서별 월급 평균을 구하고 평균에 미달하는 월급을 받는 직원이 가장 많은 부서를 구하라.
- **절차적 언어는 앞쪽부터 순서대로 처리한다. SQL은 과정은 기술하지 않고 요구 사항만 전달한다.**
 - 절차적 언어 : 찬장의 흰 사발을 꺼내 우물가에 가서 물을 가득 담아 안방으로 냉큼 가져 와라.
 - 선언적 언어 : 돌쇠야, 물 떠와라.
- **지시가 복잡한 건 상관없지만 모호해서는 안되며 원하는게 뭔지 정확히 밝혀야 한다.**



SQL 테이블

- 관계형 데이터베이스는 정보를 표 형태로 정리한다. 예 : 주소록

이름	주소	전화번호	키	성별
김상형	경기도 오산시	111-2222	180	남자
김한슬	경기도 용인시	333-4444	178	여자
권성직	경기도 화성시	555-6666	175	남자
최상미	경기도 화성시	777-8888	166	여자
문종민	서울시 송파구	999-0000	179	남자

- 익숙하고 직관적이다. 정보를 표 형태로 정리해 놓은 것을 테이블(table)이라고 부른다.
- 엔터티(Entity) : 테이블이 표현하는 대상. 세상의 모든 것.
- 레코드(Record) : 테이블에 저장된 엔터티 하나. 도표의 가로줄에 해당한다.
- 필드(Field) : 레코드의 세부 속성. 테이블의 세로줄에 해당한다. 엔터티에 따라 필드의 목록은 달라진다.
 - 컴퓨터 : CPU, 메모리 용량, 그래픽 카드, 사운드 카드, 하드 디스크
 - 가계부 : 날짜, 금액, 사용처, 수입, 지출, 잔액
 - 재고 : 분류, 품목, 제품명, 입고날짜, 재고, 유통기간



SQL 테이블

- 필드 여러 개가 모여 레코드 하나가 되며 레코드 여러 개가 모여 테이블이 된다.

	이름	주소	전화번호	키	성별
레코드 →	김상형	경기도 오산시	111-2222	180	남자
레코드 →	김한솔	경기도 용인시	333-4444	178	여자
	권성직	경기도 화성시	555-6666	175	남자
	최상미	경기도 화성시	777-8888	166	여자
	문종민	서울시 송파구	999-0000	179	남자

테이블

- 레코드와 필드는 문맥에 따라 여러 가지 동의어가 정의되어 있다.
- 보는 관점에 따라 용어가 조금씩 다를 뿐이며 칭하는 대상은 같다.
- 실무에서는 별 구분없이 섞어서 사용한다.

정식 명칭	도표 관련 용어	한국말로	모델링 용어
레코드	로(Row)	행	튜플(Tuple)
필드	컬럼(Column)	열	어트리뷰트(Attribute)



SQL DB 오브젝트

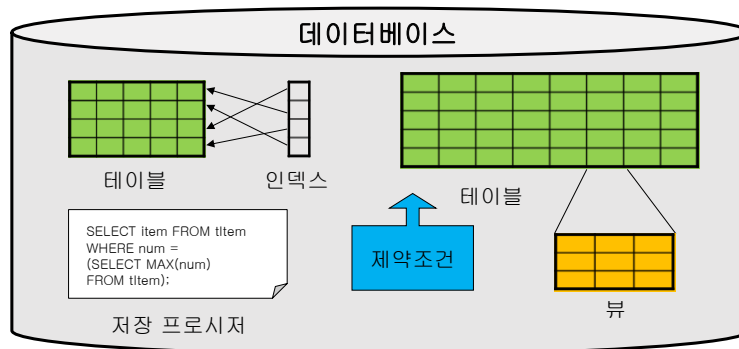
- 여러 개의 테이블이 모여야 정보로서의 가치를 발휘한다. 회원, 상품, 구매라는 세 가지 정보가 모여야 온전한 쇼핑몰을 표현할 수 있다.

번호	이름	주소
1	구홍녀	무거동
2	김상영	신북 로터리
3	김영주	운산면

번호	상품	가격
1	노트북	120
2	원피스	32
3	옥장판	16

날짜	고객	상품
2021-5-8	3	2
2021-5-12	2	1
2021-5-13	1	3

- 관련 있는 테이블과 이를 지원하는 장치가 모두 구비되어야 완전한 데이터베이스가 되어 현실 세계를 제대로 반영할 수 있다.
- 데이터베이스에 저장되는 모든 것을 통칭하여 DB 오브젝트라고 하며 간단하게 개체 (Object)라고 부른다.





SQL 명칭 규칙

- **DB 오브젝트는 서로 구분하기 위해 고유한 이름이 있어야 한다.**
 - 같은 범위 내에서 이름이 중복되어서는 안 된다.
 - 대소문자는 구분하지 않으므로 마음대로 붙여도 된다. 일관성을 지키는 것이 좋다.
 - 길이는 최대 128자까지 가능하다.
 - SQL 예약어는 쓸 수 없다.
 - 유니코드에 포함된 모든 문자를 다 사용할 수 있다.
 - #이나 @ 같은 기호로 시작하는 명칭은 DBMS가 정의한 특별한 의미를 가진다. 언더바(_)외의 기호는 가급적 쓰지 않는다.
- **명시적인 규칙보다 의미를 명확히 알 수 있는 간결한 이름을 붙이는 것이 더 중요하다.**
- **오브젝트의 종류나 타입을 구분하는 접두어를 붙이는 방식을 권장한다.**
 - DB : 첫 자만 대문자. Study
 - 키워드 : 모두 대문자. SELECT, CREATE
 - 테이블 : t 접두로 시작하며 첫 자는 대문자로 쓴다. tCity, tMember
 - 필드 : 모두 소문자. 여러 단어일 경우 어근만 대문자. name, orderDate
- **한 번 정한 규칙을 일관되게 지키는 것이 바람직하다.**



SQL 데이터 타입

- 정보의 크기나 형태를 규정하는 것을 데이터 타입(Data Type)이라고 한다.
 - 정보의 크기에 꼭 맞게 메모리를 알뜰하게 사용한다. 수십억을 넘어가는 화폐 액수는 큰 메모리가 필요하지만 사람의 나이는 고작 200을 넘지 않으니 약간의 메모리면 충분하다.
 - 실수나 에러로 인한 사고를 방지할 수 있다. 정수로 정의한 나이 정보에는 15, 39 같은 숫자만 들어가야지 "아직 어림", "비밀임" 따위의 문자열을 저장해서는 안된다.
 - DBMS가 타입을 미리 알고 있으면 최적화된 방법으로 정보를 읽고 빠른 속도로 연산할 수 있어 성능이 향상된다.
- 5개의 타입만 해도 현실 세계의 데이터 98% 이상을 표현할 수 있다.

타입	설명
INT	정수
DECIMAL	실수
CHAR	고정 길이 문자열
VARCHAR	가변 길이 문자열
DATE	날짜



SQL 데이터 타입

- 고정 길이인 CHAR 타입은 길이보다 짧은 문자열의 뒷부분을 공백으로 채우는데 비해 가변 길이인 VARCHAR 타입은 문자열 길이만큼만 저장한다.

abc
한글
SQL Server

CHAR(10)

3	abc
4	한글
10	SQL Server

VARCHAR(10)

- VARCHAR형은 용량면에서 유리하고 CHAR형은 길이가 일정해 속도가 빠르다.
- 길이가 일정한 필드에는 CHAR형을 쓰는 것이 좋고 길이가 들쭉날쭉한 문자열은 VARCHAR 타입이 적합하다.

21	좋은 정보 감사합니다.
3	^ _ ^
37	덕분에 골치아픈 문제 잘 해결했습니다.
23	금전 대출. 015-396-1523

VARCHAR(100)

좋은 정보 감사합니다.
^ _ ^
덕분에 골치아픈 문제 잘 해결했습니다.
금전 대출. 015-396-1523

CHAR(100)