**01 3 데이터베이스기본개념 데이터 유형**

2023.09.23 토 오후 4:23 ・ 11분 23초  
임소정

참석자 1 00:00  
데이터 유형에 대해서 살펴보겠습니다.   
데이터베이스에 있어서 가장 중요한 목적은 데이터를 모아두는 것이다라고 그랬죠 또 우리가 빅데이터 분석을 한다의 출발은 데이터를 수집하는 것에부터 시작을 하죠.

참석자 1 00:20  
데이터에서 정보를 정보를 찾아내서 정보 처리 시스템을 구축을 해서 의사결정을 한다.   
출발은 뭐부터 시작하나요? 데이터 수집부터 시작을 하게 되는 거죠.   
여러분이 어떤 분석을 하는 그런 목적에 맞게 데이터를 수집을 해야 되는데

참석자 1 00:42  
그러면 데이터 유형에는 어떤 것들이 있을까 살펴보도록 하겠습니다.   
형태에 따라서 먼저 이제 무엇에 따라서 분류를 할 거냐 조건이 있는데 형태에 따라서 어떤 모양으로 구성되어 있느냐에 따라서 정형이다 반정형이다 비정형이다 이렇게 구분을 해요.

참석자 1 01:09  
정형이다. 그러면 정형되어 있어 그 말은 고정된 구조나 형식을 가져요.   
관계형 데이터베이스에 있어서 테이블과 같습니다.   
행과 열로 이루어지는 데이터예요. 그래서 예를 보면 이런 거죠.   
네 반듯반듯하죠 그래서 엑셀을 생각해도 되겠습니다.

참석자 1 01:33  
지역 학과 성명 성적 학적 문자 숫자 이렇게 잘 이루어져 있죠 바둑판 모양으로 그래서 고객에 대한 정보나 또는 재고 데이터나 매출 내역이나 이렇게 고정된 구조와 형식을 가진 데이터를 우리는 정형 데이터다라고 합니다.   
반 정형 데이터가 있는데 이거는 구조는 있어요.   
구조는 있는데 관계형 데이터베이스처럼 엄격하게

참석자 1 02:06  
고정되어 있지 않은 데이터예요. 그래서 xml이라든지 제이슨이라든지 이런 형식으로 표현이 돼요.   
그래서 키 값상이라든지 배열이라든지 중첩된 요소 등으로 데이터를 구조하는 이런 부분이에요.   
그래서 분명히 내용은 있죠 이름은 존이고요.   
나이는 30이고요. 도시에 있는 지역은 뉴욕이에요라는 이런 건 내용은 있지만 반듯하게 구조화되어 있지는 않아요.

참석자 1 02:37  
웹 페이지에서 메타 데이터 제이슨 형식에서 이제 api 응답 또 설정 파일 이러한 것들이 이제 이렇게 이루어져 있어요.   
비정형 데이터는 전혀 전혀 구조 고정된 구조나 형식이 없는 데이터예요.   
이런 거죠.

참석자 1 02:59  
텍스트라든지 이미지 동영상 오디오 파일 이런 것들이 돼요.   
그래서 소셜 미디어의 게시물이라든지 사진 동영상 음성 녹음 파일 이메일 내용 또 이제 우리가 댓글 달 때 나오는 그런 문장들 이런 것들이 이제 비정용 데이터예요.   
그래서 이제

참석자 1 03:22  
텍스트 분석을 할 때 우리가 어떤 유용한 정보를 뽑아내는 텍스트 분석을 할 때는 비정형 데이터를 많이 수집을 해가지고 분석을 합니다.   
형태에 따라서 정형 반정형 비정형이 되는데 우리가 이제 데이터베이스 쪽에서는 이렇게 정형 데이터를 주로 이용을 하게 됩니다.   
이번에는 특성에 따라서 데이터를 분류를 해보겠습니다.   
아주 다양한 데이터의 형식이 있는 것을 볼 수가 있죠.

참석자 1 03:57  
그래서 이러한 것들은 이제 서로 여러분이 책을 읽는다든지 또는 의사소통을 할 때 필요한 부분인데요.   
먼저 범주형이다 수치형이다 이렇게 나누어 볼 수가 있어요.   
범주형이다. 그러면 이제 종류를 나타내는 값이에요.   
연산을 할 수가 없죠

참석자 1 04:18  
예를 들어서 이제 성별이다. 남도하기 열을 한다든지 물론 이제 여러분이 생각할 때는 남자 인원수 여자 인원수를 해서 더하기를 하겠다.   
그런 의미가 아니라 우리가 1 더하기 2는 3이잖아요.   
그럼 남 더하기 1은 뭐가 나오지? 그렇게 의미가 없는 걸 말을 해요.   
음식 짜장면 더하기 짬뽕은 뭐지? 이런 것들에 연산의 의미가 없는 것을 우리는 이제 범주형이다라고 그래요.

참석자 1 04:50  
수치형은 이제 숫자로 표현되는 것들이에요.   
많이 있죠 우리가 이제 볼 때 키 몸무게 이런 것들이 다 숫자로 표현되는 거죠.   
그래서 특성에 따라서 데이터는 크게 범주형이다 숫자형이다 이렇게 구분을 해요.   
이 범주형은 다시 명목형과 순서형으로 나눠지는데

참석자 1 05:15  
순서가 있냐 없냐라는 거예요. 즉 서열이 의미가 있느냐 없느냐에 따라서 명목형과 순서형으로 나누어지게 되는데 명목형은 서열이 없어요.   
즉 거주 지역 서울 부산 대구 대전 이런 식으로 음식 짜장면 짬뽕 탕수육 이렇게 순서가 없고 성별 남녀 이런 식으로

참석자 1 05:42  
근데 순서형은 서열의 의미가 있어요.   
그래서 학년 1학년 2학년 3학년 4학년 이런 식으로 나타내는 것이 우리가 순서형이에요.   
그래서 범주형은 어떤 이제 종류를 나타내는 것이고 명목형은 서열이 없는 거 의미 없는 거 순서형은 서열이 의미가 있는 것을 말을 해요.   
즉

참석자 1 06:05  
성별에서 남녀나 여남이나 의미가 없죠 근데 학년은 1학년 2학년 3학년 4학년 의미가 있죠 이런 것들이에요.   
그러니까 여러분이 이제 이런 거죠. 혈액형을 따질 때 a형이다 b형이다 이런 거 하고 성적에서 a 학점 b학점 이건 또 의미가 다르죠 그래서

참석자 1 06:30  
a형이다 b형이다 이건 이제 명목형이 되지만 학점에서 a 학점 b 학점은 이제 순서형이 된다 이렇게 구분해 주면 될 것 같아요.   
숫자로 표현되는 데이터를 우리는 수치형이다라고 했어요.   
그래서 키다 몸무게가 있는데 이상형이다 연속형이다라고 세부적으로 구분을 해요.

참석자 1 06:55  
이산형이다라고 하면 이산 새일 수가 있다라는 거예요.   
고객의 수 판매량 이런 식으로 정수로 1개 2개 3개 이렇게 있어요.   
고객의 수가 몇 명 있어요? 지금 몇 명 왔어요? 1.2명 왔어요 그런 거 없죠 한 명 왔어요 2명 왔어요 이렇게 되는 거죠.

참석자 1 07:18  
물티슈 몇 개 팔았어요? 한 개 팔았어요 두 개 팔았어요 1.2개 팔았어요? 그런 건 없죠 정수로 딱딱 떨어지는 것 그걸 우리는 이산형이다라고 그래요.   
연속형은 말 그대로 딱딱 끊어지지 않고 연속적으로 이어지는 숫자 값이에요.   
당신의 키가 얼마입니까? 170이요

참석자 1 07:43  
170.125 170.25 이런 식으로 이어져 있는 거죠.   
실수 형태로 나온다라고 생각해 주면 돼요.   
당신의 몸무게가 얼마입니까? 45인데요.   
44.5인데요. 이런 식으로 온도 지금 온도는요 36.2도요.   
단 몇 점 맞았나요? 네 80.1 맞았어요.

참석자 1 08:08  
이런 식으로 정수로 딱딱 떨어지지 않고 실수처럼 이렇게 아주 세밀한 값들이 이어지는 것을 우리는 연속형이다라고 그래요.   
그럼 여러 개 구분할 수가 있겠죠 특성에 따라서 범주형과 수치형 범주 종류 나타내고 수치는 연산의 의미가 있는 거죠.   
크기 비교 산술 연산이 가능한 것 범주형은 다시 목록형 순서형

참석자 1 08:38  
순서가 서열이 의미가 없다. 명목형 의미가 있다.   
순서형 새일 수 있다 이산영 딱 떨어지지 않고 실수처럼 이렇게 이어진다.   
그러면 연속형 이렇게 구분할 수 있어요.   
그러면 이제 여러분이 이제 학습을 잘했으면 한번 정리를 한번 해보세요.   
이것에 대한 답변을 줄 수 있는지 보세요.

참석자 1 09:06  
데이터베이스의 정의에 대해서 설명을 해보세요.   
그리고 데이터베이스에 네 가지 특징이 있었어요.   
네 그것에 대해서 설명을 해보세요. 그리고 데이터를 유형별로 분류해서 설명을 해보세요.   
네

참석자 1 09:24  
그리고 이 문제에 대한 답을 한번 맞춰볼까요? 여러분 데이터베이스의 정의로 옳지 않은 것은 무엇입니까? 조직의 존재 목적이나 유용성 면에서 존재 가치가 확실한 필수적인 데이터다라고 그랬고요.   
정보 소유 응용에 있어서 지역적으로 유지가 돼야 된다.   
지역적이다. 컴퓨터가 접근할 수 있는 저장 매체에 저장된 자료이다.

참석자 1 09:52  
동일한 데이터의 중복성을 최소화해야 된다.   
중복성 컴퓨터 네 존재 가치가 확실해야 된다.   
지역적이다. 옳지 않은 것은요? 네 바로 2번이죠.   
지역지역이 아니라 전업적으로 유지가 되어야 되는 부분이죠.   
여러분 맞춰보세요. 다음 중 이상형 데이터에 속하는 것은? 그랬어요.

참석자 1 10:18  
그럼 여러분 머릿속에 이상형은 어디에 속하지? 특성에 따라서 수치형 중에서 이상형이 있어 이상형은 셀 수 있어 딱딱 정수로 떨어지게 셀 수가 있어라는 것을 이제 골라주면 되겠습니다.   
실수가 나올 수 없는 정수만 나오는 것을 따져보면 될 것 같아요.   
고객에 대한 성별 숫자가 아니군요.

참석자 1 10:43  
가입 기간 근데 1년 단위가 뭐에 따라 다르겠지만 1년 가입했어요 1.2년 가입했어요.   
가입 고객 수 몇 명인가요? 5명이에요 다 5.1명이에요 그런 건 없죠 고객의 거주 지역 수치가 아니군요.   
네 그래서 여러분이

참석자 1 11:04  
이게 머릿속에 딱 있다라고 하면 이산형 세 수 있다 점수로 나타낸다라고 한다면 정답은 3번이다라고 볼 수가 있겠죠 네 여기까지 데이터베이스에 대한 기본 개념에 대한 내용이었어요.

clovanote.naver.com