初級シスアド講座　テキスト
第８回　　２月１８日 データベース
データベースの基礎、関係演算と集合演算、ＳＱＬ、正規化
大阪府立母子保健総合医療センター　宮野 章

1. データベースとは

データを主体(ベース)とした情報システムです。データを統合して保存し，それを複数のプログラムが共有する形式をとります。そのためデータの重複が減り，データ間を関連づけることで検索も行えます。
・DBMS(DataBase Management System)
DBMS はデータベース専用の管理システムです。データベースを運用する機能のほか、重要なデータのバックアップをとる機能や，障害発生時に残された記録からデータベースを復旧する機能，あるいはアクセスできる者を登録してセキュリティを高める機能，複数の処理で同時に同じデータを更新しようとすることを防ぐ，排他制御機能等が必要です。
・関係データベース
データベースはデータの構造により数種類ありますが，データを列×行の表形式で記憶する関係(リレーショナル)データベースが主流です。この表を関係表といい，データの種類に応じて複数の関係表に分けて作成します。

2. SQL の文法

関係データベースを作成したり利用するための言語が，SQL(Structured Query Language)です。SQL では，関係表を定義する CREATE TABLE 文,関係表にデータの値を入れる INSERT 文,値を更新する UPDATE 文,行を削除する DELETE 文などがあります。

●SELECT 文
・文字項目の検索
WHERE 句で，列名と値とを比較する検索式を入れます。文字列は`（単一引用符）で囲みます(全角文字の場合，DBMS により ’ の左側に N を付ける場合があります)。
SELECT［列名］FROM［表名］
WHERE［文字列型の列名］= ‘［値］’

・数値の検索
数値や数値項目を比較する場合，比較には比較演算子(=，<，<=，>，>=)を用います。文字列と異なり、数値は ’ で囲みません。
SELECT［列名］FROM［表名］
WHERE［数値型の列名］［比較演算子］［値］

比較条件を論理演算子(NOT,AND,OR)でつなぐこともできます。複数の論理演算子を組み合わせた比較条件の場合，その優先順位は NOT，AND，OR の順となります。また、WHERE 句での検索条件では，数値や数値項目の代わりに計算式を記述することもできます。

・AND 条件の指定
“～AND～"で記述します。

例:「検査表」から，中性脂肪の値が 200 以上 500 以下の氏名を検索する。
SELECT 氏名 FROM 検査表
WHERE 中性脂肪 >= 200 AND 中性脂肪 <= 500

・OR 条件の指定
同じ列を対象にしており，比較条件が"～もしくは～"(or 条件)の場合，IN 述語を用いることもできます。WHERE 文節の後に，
IN([比較値 1]，[比較値 2])の形式で記述します。
例:地域が「大阪府」もしくは「京都府」である行の，全列を検索する。
SELECT*FROM 住所一覧
WHERE 地域 IN(N ’ 大阪府 ’ ，N ’ 京都府 ’)

・特定の文字の検索
文字列型の項目で，文字列中に特定の文字を含むものを検索する場合，LIKE 述語を用います。特定の文字の指定では，エスケープ文字を使用できます。

SELECT［列名］FROM［表名］
WHERE［文字列型の列名］LIKE［比較文字・エスケープ文字］

エスケープ文字として，一般に%と_（アンダースコア）があります。%は任意の文字列(文字なしも含む)を表し，_ は任意の 1 文字を表します。

例:氏名が「川」で始まる行の氏名を検索する(「川」以降の文字数は不問)。
SELECT 氏名 FROM 住所一覧
WHERE 氏名 LIKE ‘ 川% ’

・関数
SELECT 文では，各種の関数が用意されています。それを用いると，合計，平均，最大値，最小値，件数の算出が自動的に行えます。

SUM 関数
　合計値を算出します。
AVG 関数
　平均値を算出します。
MAX 関数，MIN 関数
　MAX 関数では最大値を，MIN 関数では最小値を算出します。
COUNT 関数
　行数を数えます。通常は列名は関係ないので，()内には全項目を表す*を入れます。

・空白値の検索
IS NULL 文を用いることにより，空白値(NULL:ナル)が入っている項目を検索することができます。この場合，表の定義で空白値が許されていることが前提条件となります。

例:地域で空白値が入っている行を数える。
SELECT COUNT(*) FROM 住所一覧
WHERE 地域 IS NULL

・整列(並べ替え)
整列とは並べ替えることで，指定した項目(キー項目といいます)の値の大小順に検索結果を並べかえます。ORDER BY 句を用い，キー項目と大小順を指定します。大小順では，昇(小→大)順は ASC，降(大→小)順では DESC を用い，省略した場合は昇順と見なされます。

SELECT [グループ分け項目，関数] FROM [表名]
ORDER BY [整列項目] [ASC/DESC]

・数値のグループ分け
GROUP BY 文節により，関係表中の行を，指定した項目の値別にグループに分けて表示できます。一般に SUM などの関数と併用して，指定した項目の値ごとに合計値や平均値などを集計するのに用います。

SELECT [グループ分け項目，関数] FROM [表名]
GROUP BY [グループ分け項目]

・GROUP BY での検索
特定のグループだけを選ぶ場合，HAVING 文節を用います。
HAVING 文節は，GROUP BY 文節と併せて，もしくは集合関数と併せて用います。

例：地域ごとに検査値をグループ分けを行い，件数が 2 件以上のものから最大値と最小値を出力する。なお最大値の前には`MAX=`を，最小値の前には`MIN=`を，それぞれ付ける。
SELECT 地域，’MAX=’，MAX(検査値)，’MIN=’，MIN(検査値) FROM 住所一覧
GROUP BY 検査値 HAVING COUNT(*)>=2

・重複を除外する
DISTINCT 文節を用います。

●データ型
データの形式です。大きく文字項目と数値項目，及びその他の項目に分けられます。JIS の規定を中心に，主要な DBMS での対応も記述します。

・文字項目
アルファベットや記号，あるいは全角文字が入る項目です。数字を入力することはできますが，計算には使えません。CHAR， NCHAR があります。

・数値項目
計算に用いることができるデータ型です。そのため数値と記号(+， -)しか入りません。INT(整数)，SMALLINT(桁数の少ない整数)，DEC(10 進数)，FLOAT(浮動小数点数)，REAL(実数)などがあります。

・その他の項目
DATE，TIME があります。メーカによって，金額表示(数字の先頭に自動的に$や\を付けたり，カンマを入れる)できるデータ型などが追加されています。

データ型	内容
CHAR	・半角文字(DBMS によっては全角文字も含む)に用いる ・文字数を，かっこをつけて指定する 例:CHAR(12)←半角文字 12 文字
NCHAR 注：N は National の略	・全角文字に用いる ・全角文字数を，かっこをつけて指定する 例 NCHAR(8)←全角文字 8 文字
INT	・整数値に用いる
DATE	・日付に用いる ・西暦年/月/日で表記する。 例 1999/01/01←1999 年 1 月 1 日
TIME	・時刻に用いる ・時/分/秒/100 分の 1 秒で表記する 注 DBMS によって DATETIME などの指定がある

3. 正規化とは

●正規化の目的
データベースでのデータ設計における手法で，関連するデータ項目ごとに表として独立させる作業です。正規化の目的として，下の 2 点があります。

・データ重複の排除
同じデータが複数の表に存在することをなくすことで，データの記憶領域を小さくできます。ただし後述するキー項目だけは，表と表の間を結ぶ項目として，重複して存在します。

・データの更新・追加・削除箇所の局所化
同じデータが複数存在すれば，削除や更新で漏れが生じる可能性があります。またデータを追加する場合でも，一カ所の追加で済みます。

●正規化の段階
正規化はいくつかの段階に分けて進めます。第 1 正規化～第 5 正規化までの段階がありますが，初級シスアドでは第 3 正規化までが問われています。

・第 1 正規化
データ項目を固定部分と繰り返し部分に分け，固定部分を繰り返し部分に重複させる作業です。これは関係表の各行のデータ要素を，原子的なそれ以上分割できない状態にすることです。第 1 正規化が行われる前の表を非正規形といい，例えば以下のような表です。

品名コード	品名	問屋コード	問屋名	単価	数量
000001	品名 A	0010 0011	○○会社 △△会社	2，000	40 40
000002	品名 B	0020 0021	□□会社 ××会社	3，000	50 50

上の表では，1 つの行に，2 つの問屋情報が入っています。これを別々の行として独立させます。従って，以下のようになり，このような構成を第 1 正規形といいます。

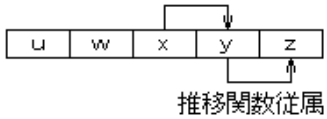
品名コード	品名	問屋コード	問屋名	単価	数量
000001	品名 A	0010	○○会社	2，000	40
000001	品名 A	0011	△△会社	2，000	40
000002	品名 B	0020	□□会社	3，000	50
000002	品名 B	0021	××会社	3，000	50

・第 2 正規化
第 1 正規形の中からキー項目を選び出すとともに，それの値を指定することで，データが特定できる項目を探します。この関係を関数従属といいます。キー項目に完全関数従属している項目と，部分的に関数従属している項目に分割し，それらを別の関係表として独立させます。

・関数従属とは
あるデータ項目の値を決めると，別のデータ項目の値が一意に定まる関係です。完全関数従属とは，その中でも主キーの値が決まると別のデータ項目の値が定まる，1 対 1 の関係です。関数従属関係の書き方ですが，ある x に対して y が一意に定まる場合，x→y と書きます。それに対し，主キーが二つ以上の項目で構成されている場合，その中の一つの項目と 1 対 1 の関係であるものを部分関数従属といいます。

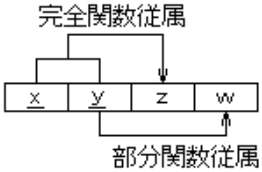
関数従属

属性（項目）xの値が決まるとyの値が自動的に決まる場合、「yはxに関数従属」といい、「x→y」で表す。
属性x、y、zにおいて、yはxに関数従属（x→y）し、zはyに関数従属（y→z）し、xがyに関数従属でないとき、「zはxに推移関数従属」という。



推移関数従属

複数の属性を連結した結合キーにおいて、結合キーのすべての値が決まれば他の属性の値が自動的に決まるとき、完全関数従属という。また、結合キーのいずれかの値が決まれば他の項目の値が自動的に決まるとき、部分関数従属という。次のように、主キー(x＋y)のうち、yだけでwが決まるのであれば、「wは主キーに対して部分関数従属」であるという。

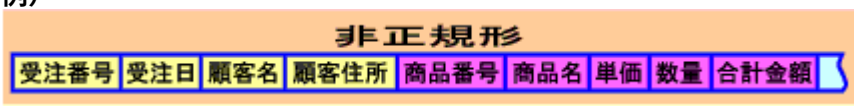


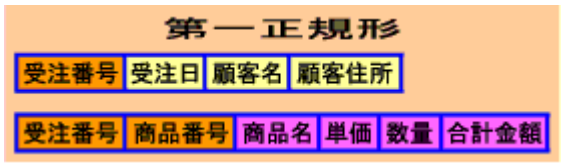
部分関数従属

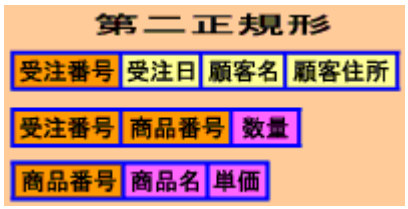
・主キーと外部キー
主キー(PRIMARY KEY)とは，それぞれの表の中で他の項目が関数従属する項目で，主キーとなる項目の値を定めると，関数従属する項目の値も一意に定まります(患者 ID などがこれに当たります)。外部キー(FOREIGN KEY)とは，別の関係表で基本キーとして使用されている項目です。外部キーにより表と表が関係付けられ，データの更新や削除などで整合性を保つことができます。

・第 3 正規化
キー項目の値を定めると関数従属している項目の値が決まり，その値が決まることでさらに別の項目の値が決まる関係を推移的関数従属といいます。第 3 正規化は，第 2 正規形から推移的関数従属の関係にある項目の組み合わせを，別の表として独立させる作業です。第 3 正規化を行なった表を，第 3 正規形といいます。

例)







4. DBMS の機能

・DBMS とは
データベースの作成や更新を管理するソフトウェアです。ミドルウェアに分類されます。排他制御機能，障害復旧機能，不正アクセス対策機能などがあります。

・排他制御機能
同じデータに複数の更新処理が同時に加わった場合，データがおかしなものとなります。このため最初の更新処理が終わるまで次の更新処理を停止(更新ロック)して，データの整合性を保つ機能が排他制御です。

排他制御はデータベースに限らず，同じデータに複数の更新が同時にかかる場合に必要です。

・バックアップファイルとジャーナルファイル
データベースで障害が発生しても復元できるように，ある時点のデータベースを記録する項目には，更新したデータの更新前ジャーナル(更新前の状態)と更新後ジャーナル(更新後の状態)，更新時刻，トランザクション識別子(どの処理が更新したかを示すもの)などがあります。ソフトウェアの異常終了など，比較的軽度の障害で用います。

・障害復旧機能
データベースでのデータ更新処理は 2 系統で行い，両方の処理が正常に終了してから書き込み動作を行うようになっています。この方式を二相コミットといいます。また定期的にデータベース全体のデータをバックアップして，バックアップファイルを作成する機能やデータの更新を記録したジャーナルファイルを作成する機能もあります。それらを利用した障害復旧方法として，ロールバックとロールフォワードがあります。

・ロールバック
二相コミットを利用して，障害発生時にデータを更新前の状態に戻す方法です。ハードディスクに障害が発生していない，軽度の障害で利用される復旧手法です。

・ロールフォワード
ハードディスクが壊れるなどの重度の障害が発生した場合，バックアップファイルと，それを作成してから障害発生までが記録されたジャーナルファイルのデータとで，障害発生前のデータベースを再現できます。このようにバックアップファイルとジャーナルファイルとでデータベースを再構築する方法を，ロールフォワードといいます。

●不正アクセス対策機能
利用者ごとにアクセスできるデータの範囲や更新処理の可否などを設定できます。このアクセス制限の設定表をアクセスリストといい，SQL では GRANT 文で設定できます。

5. 関係演算

関係データベースの演算のうち，最も基本的な操作は選択・射影・結合で，初級シスアド試験ではその 3 操作が繰り返し出題されています。

・選択(selection)
一つの表から，条件に合った行を取り出す(導出する)操作です。下の例では，「定価 10 万円以上」を条件としています。

品名コード	品名	定価
001	IL6 キット	120，000
002	IgG キット	45，000
003	RF キット	15，000
004	IL12 キット	160，000

→

品名コード	品名	定価
001	IL6 キット	120，000
004	IL12 キット	160，000

・射影(proiection)
一つの表から，条件に合った列を取り出す操作です。下の例では，「品名」を条件としています。

品名コード	品名	定価
001	IL6 キット	120，000
002	IgG キット	45，000
003	RF キット	15，000
004	IL12 キット	160，000

→

品名
IL6 キット
IgG キット
RF キット
IL12 キット

・結合(join)
二つの表から，特定の列を「キー」にして一つの表を作成する操作です。下の例では，「発注コード」がキーになってし・ます。

発注コード	品名
010511	IL6 キット
010601	IgG キット
011025	RF キット
011118	IL12 キット

発注コード	問屋名	個数
010511	A 商事	125
010601	B 株式会社	83
011025	C 流通センター	250
011118	D 商会	48

↓

発注コード	問屋名	品名	個数
010511	A 商事	IL6 キット	125
010601	B 株式会社	IgG キット	83
011025	C 流通センター	RF キット	250
011118	D 商会	IL12 キット	48

6. 商用データベース

利用者に有償でデータを情報を提供するデータベースです。データベースの構築や運営を有償で行うサービスとは別のものです。

●商用データベースでの検索
・フリーターム
商用データベースでは，キーワードを指定して検索することが多いです。たとえば過去一年間の新聞記事を収録したデータベースで，初級システムアドミニストレータ試験関係の記事を検索する場合，キーワードとして「初級システムアドミニストレータ試験」を指定するように任意の単語(文字列)で検索することを，フリーターム検索といいます。

・シソーラス(thesaurus)
検索用に，同義語や類似語，広義語，反意語等を体系的に整理した辞書があり，これをシソーラスといいます。

・検索条件の組み合わせ
複数のキーワードを，AND 条件，OR 条件で組み合わせて検索することができます。例えば「初級システムアドミニストレータ試験」と「初級シスアド試験」のいずれかが該当する場合には OR 条件で，「初級システムアドミニストレータ試験」の「試験日」を検索する場合には AND 条件で組み合わせます。

・検索の絞り込み
キーワードにより複数のデータが検索された場合，さらにそれからキーワードの追加などで検索条件を増やし，目的のデータを絞り込んで検索します。