

5. 階層型データ・ベース

加藤文夫 斎藤正男
(東京大学)

5.1 階層型データ構造

プログラムとファイル装置のインターフェースを制御するロジカルI/OCSをアクセス・メソッドと呼ぶ。アクセス・メソッドは単に入出力作業を行なうだけではなく、データがどんな形でファイルされているかによって働き方を変える。論理構造としては木構造を前提とし、階層編成を標準のデータ構成とするデータ・ベース・システムが、IBM社のIMS/V(S (Information Management System Virtual Storage)である。ここでは、そのデータ構造、機能などにつき概説する。

5.2 歷史

IMS は、1960 年代のはじめに、米国の航空会社が航空宇宙産業における多年の経験をもとに、アポロ計画のために開発したシステムである。宇宙船開発、製造の部品表管理を計算機の助けを借りて行なうため、200 万点にもものぼる部品を収容するデータ・ベース・システムを作ること为目标としていた。部品表管理は非常に複雑な作業であり、質的にも量的にも高度な技術が要求される。第3世代の計算機が発表されると、OS/360 を土台としたソフトウェア IMS/360 が、IBM の協力を得て 1969 年に完成した。その後、IBM が OS/VS を土台にして、IMS/360 を機能的にいくぶん拡張したシステムとして、IMS/VS が登場した。

5.3 使用計算機と必要記憶容量

IBM システム / 370 モデル 145, 155 II, 158, 165 II, 168
主記憶容量 384 Kbyte (OS/VS1), 512 Kbyte (OS/VS2)
(バッチ処理のみの場合) (256) (348)

5.4 特徴

①汎用システムであり，拡張性に富む。②データ・ベース（DB）機能とデータ通信（DC）機能をもつ。③ユーザ・プログラムはデータの物理的構造と独立である。④同時に複数のプログラムが働作できる。⑤データの安全確保がはかられている。⑥システム・ログの利用により，チェック・ポイント，再開始，データ・リカバリの機能がすぐれている。⑦システムの稼動状況の評価のための統計情報が，システムにより供給される。

5.5 データ構造

[a] 基本構造とセグメント

前にも述べたごとく、データの基本構造は図1のような木構造である。この図で、リーグ、チーム、監督、選手はそれぞれ1つのセグメント・タイプを表わす。各セグメント・タイプに具体値を与えた例が、図2.である。セリーグ、巨人、長

島、ジョンソンなど4角で囲まれたデータが、取扱いの単位であり、セグメントと呼ばれる。図2のセリーグ、パリーグのように、最上位を占めるセグメントを根源セグメント、それに続くすべてのセグメントを従属セグメントという。各根源セグメントは1つの木構造を定義する。そのそれぞれをデータ・ベース・レコード(DBR)と呼ぶ。1つのDBRの中で、上位から下位へ、左から右へという規則によって、各セグメントを1意的に順序づけることができる。この順序を階層順序といい、これに従ってあるセグメントから次のセグメントを指すポイントを階層ポイントという。各セグメントは、どのセグメント・タイプに属するかを示すセグメント・コード、ポイント、データなどから成る。データ部には、固定長のものと可変長のものとがあり、場合に応じて使い分けることができる。

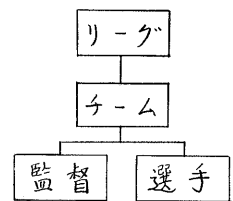


図1. 基本構造

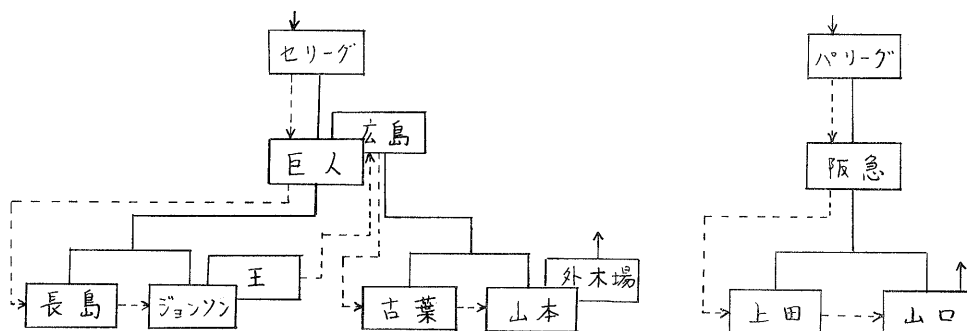


図2. 図1に対応するデータ・ベース・レコード

[b] フィジカル・データ・ベース(PDB)

データ・ベースは記憶装置への格納法(あるいはアクセス・メソッド)によって、(b-1)階層順次編成、(b-2)階層索引順次編成、(b-3)階層直接編成、(b-4)階層索引直接編成、の4種に分類される。

(b-1) 階層順次編成 [HSAM]

記憶装置として、ダイレクト・アクセス、順次アクセスの両者を使用できる。1つのデータ・ベースをDBRごとに分け、それぞれに含まれるセグメントを階層順序に従って1列に並べて記憶するものである。根源セグメントにキーが定義されていれば、DBRはその根源セグメントのキーの値の小さい順に並べられる(図3)。

ブロック #1

セリーグ	巨人	長島	ジョンソン
------	----	----	-------

#3

外木場	パリーグ	阪急	上田
-----	------	----	----

#2

王	広島	古葉	山本
---	----	----	----

#4

山口	次のDBR
----	-------

図3. HSAM

(b-2) 階層索引順次編成 [HISAM]

HISAMに索引つけたものである。すなわち、各DBRの根源セグメントに互いに異なる値をもつキーを定義し、それを索引として用いる(図4)。根源セグメントに対し、索引を通じてランダムにアクセスできる利点をもつ。

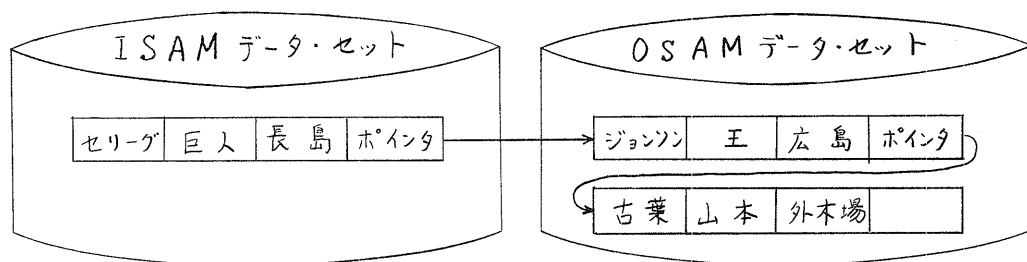


図4. HISAM

(b-3) 階層直接編成 [HDAM] (図5)

ユーザの用意するランダムマイジング・モジュール(ハッシング)により、根源セグメントへのアクセスを行なう。各セグメントは、階層ポインタ、子ポインタ、双子ポインタなどのダイレクト・アクセス・ポインタで結ばれる。これらのポインタによる直接アクセス、データの消去された記憶領域の再利用などの点ですぐれている。ダイレクト・アクセス記憶装置が用いられる。

(b-4) 階層索引直接編成 [HIDAM] (図6)

HDAMがランダムマイジング・モジュールを介して根源セグメントへのアクセスを行なうのに対して、HIDAMでは索引を用いる。図6に示すように、一方は各根源セグメントのキーとそのアドレスを集めた索引データ・ベースであり、他方は根源セグメント以下全セグメントがダイレクト・アドレス・ポインタで結合されている。

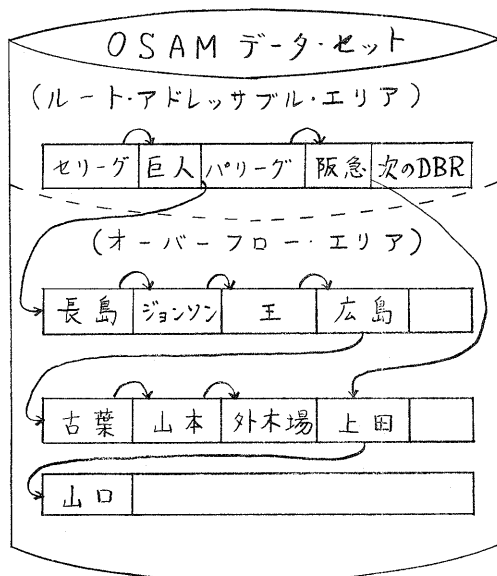


図5. HDAM

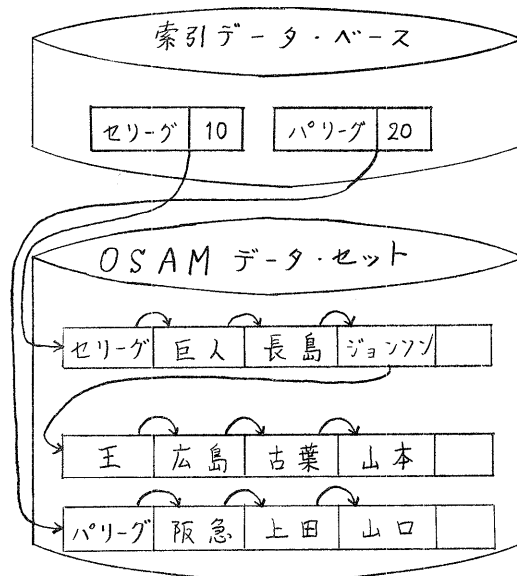
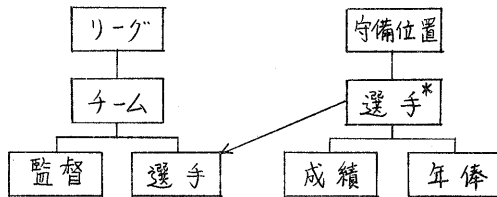


図6. HIDAM

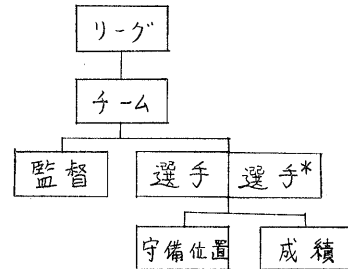
〔c〕 ロジカル・データ・ベース

2つ以上のPDBにまたがって、同一のセグメント・タイプが要求されるとき、データは1箇所にのみ格納され、他のPDBではそのセグメント・タイプを指すポインタが代用され、データの重複を避ける。また、あるPDBから他のPDBへのアクセスも可能となる(図7)



(a)「選手」を共有する2つのPDB

図7. ロジカル・データ・ベース



(b) (a) から得られるロジカル・データ・ベースの1例

〔d〕 2次索引

索引は、根源セグメントに対してのみ定義されたものであった。これに対して、2次索引は従属セグメントのための索引であり、木構造の下位レベルのセグメントから、上位へのアクセスが能率化される。

5.6 データの取扱い

〔a〕 データ・ベースの定義

IMS/VISにおけるデータ・ベースは、DDLであるData Base Description (DBD)により定義される。DBDは、ユーティリティ・プログラム(DBDG)に入力されるユーザの命令で生成され、データ・ベースの記憶構造、論理構造、索引構造などが記述される。DBDは処理プログラムから見たデータの論理構造と記憶構造の間のマッピングとしての役割を果たし、これによってデータとプログラムの独立性が保障される。

〔b〕 データの検索・更新・処理

プログラム言語として、COBOL, PL/I, アセンブラが許されているが、処理プログラムには原則として入出力部がない。データの検索、更新、消去、挿入などデータ・ベースと直接接触する部分は、すべてサブシステムData Language I (DL/I)をCALL文で呼ぶことによって実行される。使用できる主な命令には、GET UNIQUE, GET NEXT, REPLACE, DELETE, INSERT などがあり、セグメントの指定などと共にCALL文の引数として与えられる。統計処理、ソーティングなどその他の処理は、処理プログラム中で行なわれる。

5.7 医療への導入について

IMS/VISは、特殊な機能を除きほぼ万能ともいえるシステムである。それだけに規模も大きく、医療に導入するとしたら国あるいは都道府県レベルでの共同利用システムとなろう。その際、強力なDC機能が重要な役割を果たすと思われる。また医療用に新たなシステムを開発する場合、木構造を基本とし、クロスリファレンスの機能を加えたデータ構造、2次索引、可変長セグメントなどの考え方が参考になるように思われる。