データベース・システム 51－4
（１９８６・１・２０）

ローレンス●バークレー研究所のＤＢ研究と科学技術ＤＢについて

川越　恭二
日本電気（株）　Ｃ＆Ｃシステム研究所

１　はじめに

本報告は、筆者が１９８４年７月より１年間滞在したカリフォルニア大学ローレンス・バークレイ研究所でのデータベース研究とサイエンティフィック・データベース（科学技術データベース）の状況と筆者が滞在中に行なった科学技術データベース技術の課題の一つである時間概念を扱うようにしたデータベースのデータモデルと格納構造に関するものである。

ローレンス・バークレイ研究所（以降、ＬＢＬと略す。）でのデータベース研究は、統計データベースから科学技術データベースへと発展させており、A. Shoshani, H.K.T. Wong, F. Olken, D. Rotem, J. MaCarthy の研究者を中心に物理、化学等の研究者との密接な交流により実用レベルへの適用を意識したデータベースの基本的技術の育成を行なっている。

科学技術データベースは、データベース分野の拡大に向けての一つの研究分野であり、科学・技術者へのデータベース支援を目指している。このため、統計データベースとともに研究が活発化しており、本報告ではこのデータベースの状況と、筆者が滞在中に行なった時間に関係したデータの統一した見方への接近法について説明する。なお、後者については、サイエンティフィック・データベースでの時系列データやイベント履歴データにみられる時間に関係したデータの保管方法の明確化の要請という動機によっている。

２。　ＬＢＬとデータベース研究

２－１　ＬＢＬ概要

ＬＢＬは、バークレーのキャンパスの裏山（高さ３００－４００ｍ）の中腹にあり、バスで約１０分のところにある。このバスはＬＢＬ　シャトルバスと呼ばれ、ｏｆｆシャトル，ｈｉｌｌシャトル，ｓｔｒａｗｂｅｒｒｙシャトルの３ルートがありいずれも無料である。この内ｏｆｆシャトルは、バークレーのキャンパスとＬＢＬとを結び、研究者の通勤用として利用されている。車による通勤は、駐車場の確保の点からかなり困難であるが、バスが１０分間隔で運行されているため特に不自由は感じられない。建物が高い位置にあるため、窓からはサンフランシスコやバークレーが一望できる他、ＦＯＧのないときはゴールデン・ゲート橋やベイ橋がはっきりと見える。

ＬＢＬは、Ernest Orland Lawrence によって作られた。１９２９年に彼がサイクロトロンを発明したのち、その実証のためにより大きな実験装置が必要となり１９３１年にアクセルレイタをこの近所に作ったのがはじまりであった。この装置は、原子力　分子物理　核化学　放射線利用生物学　などの発展に大きく貢献したとのことで、６０年代には８人の研究者がノーベル賞をもらっている。

現在、Sherley というディレクタの下で９つの研究部門と管理サービス部門があるが、研究部門構成は柔軟で、たびたび変更される。例えば、物理研究部門と計算機研究部門は同一部門であったが、２－３年前に独立し計算機科学と数学部門となり、１年前に計算機センタと一緒になり計算部門（Computing Division）となった。

キャンパスに近いためにカリフォルニア大学バークレー校（ＵＣＢ）とＬＢＬは密接な関係にあるが、予算的にはＬＢＬは国家予算で運営されているので独立した機関であるといえる。研究者の交流は盛んであり、兼務者やＰｈＤ取得者のＬＢＬ職員も多いようであるし、セミナー参加、学生実習など、が盛んにおこなわれている。

計算機環境としては、ＶＡＸ１０数台に以前はＣＤＣで、最近はＶＡＸ８６００が設置されている。端末は、１人１台が確保されると同時に、各部屋に既にＮｅｔｗｏｒｋ端子が設置されており、部屋（１人に１部屋）を移動しても即座に端末が利用できる。

計算機科学研究部門には以前は５つのグループがあったが、予算削減のために３グループに縮小された。消滅したグループはコンピュターグラフィックスと計算機アーキテクチャであり、残りの３つは、ネットワーク、データベース、アプリケーションシステムである。

１）ネットワーク：　分散型システム研究でＬＡＮの測定とモデルの研究、分散システム設計の研究をしていて、ＵＮＩＸ，ＶＡＸの環境を十分に活用し、実用化と同時に理論的な研究を両立させて進めている。
２）データベース：　私のいたグループで、統計データベースの効率的利用、格納方式、および科学技術データベースの操作、格納方式、モデルの研究を行なっている。
３）応用システム：　エネルギー省、厚生省などからの資用で、エネルギー／公害／人口などに関する統計データ管理、処理、及び複数データの相関に関しての研究を行なっている。

２－２　データベースＧｒの研究内容

ＬＢＬのデータベースＧｒは、５名の研究者と数人のＰｈＤ候補学生から構成されており、以下の研究を行なっている。

１）　論理データベース：論理データモデル、メタデータ概念、科学技術データモデリングなど
２）　物理データベース：　転置ファイル、多次元ファイルなど
３）　ユーザインターフェース：　ＧＵＩＤＥ，ＣＡＢＬＥ，ＳＵＢＪＥＣＴなどのエンドユーザ言語

グループのリーダは、Arie Shoshani であり、彼が私の研究指導者であった。彼は、イスラエル人で非常に温和であり、データベースの研究専門誌であるTransaction On Database Systemsのエディタの一人である。
他には、 H.K.T. Wong, F. Olken, D. Rotem, J. MaCarthy がいる。まず、Wongは数人のＰｈＤ候補者を指導していると同時に幾つかの研究開発プロジェクトを管理している。彼の主な研究成果は、Ｅ－Ｒモデルベースの図式による対話的エンドユーザ・データベースインターフェース（ＧＵＩＤＥ）を研究開発したこと［１５］、彼が大学に在学中にMylopoulosと共同でデータベースとプログラミング言語を統一したＴＡＸＩＳを提案した［１６］。現在の彼の興味はデータベース物理構造にあり、科学技術データベース要の新しい格納構造をもつ関係データベースシステムを研究開発している。この新しい格納構造はbit-transposed-file と呼び関係データベースの転置ファイルをさらに推し進め、タップルをビット単位に切り出し各ビットを縦方向（即ち、カラム方向）にファイルに保存し、必要ならばビット列を圧縮化するものである。従来の転置ファイルと比べて、データ圧縮率の向上が可能な他、関係演算を復号化や逆転置することなく実行することができる。また、ビット操作であるため容易にハードウエア化が可能である。しかし、格納構造一般にいえることであるが、状況によってはアクセス効率が悪化する。
一方、Olken は、キャンパスのStonebraker 研究室の出身であり、統計操作や格納構造に関心を持っている。特に、多次元ファイル（Multi-dimensional file）の構造や、関係データベースにおけるサンプリング操作の研究を行なっている。以前は、主記憶レジデントなデータベースにおける効率的ジョインアルゴリズム（Hash Equi-Join Algorithm）を開発している［９］。McCarthyは、Shoshaniのグループではないがデータベース関連の研究を行なっているためグループの活動、討論に参加している。彼のテーマは、メタデータであり科学技術、統計などのファイル交換に不可欠なデータ記述形式を明確にしようとしている。［１］
Shoshaniはグループリーダであり、以前はＣＡＢＬＥ（Ｅ－Ｒモデルのデータ操作言語）やＳＵＢＪＥＣＴ（統計データベース用のデータモデルとデータベース操作言語）　を開発したりin［６］、ヘッダ圧縮と呼ぶデータ圧縮法を提案し統計データベースシステム（ＳＥＥＤＳ）に実装したが、最近は、科学技術データベースに関心をもち、科学技術データベースの特性を種々の科学技術応用分野の研究者との討論から明らかにしている［３］。彼は、現在研究よりは研究管理に時間を使っており手腕を発揮していると思われる。
以下に、上で記したbit-transposed-file 、メタデータそして科学技術／統計データベースようの多次元データ構造　について、その概要を説明する。
１）　［bit-transposed-file ］　［１４］
　　従来の科学技術データベースでは最もシンプルなシーケンシャルファイル構成が最もよくすかわれている。それは、逆ファイル、Ｂトリー、ハッシングなどの索引法が大容量のデータベースの保存方式にしては高価であるためである。すなわち、属性の属性値集合が小さい科学技術

データベースではほとんどの索引法は単に少数のデータセットへ分割するだけにすぎないこと、データベースがＳＴＡＴＩＣなとき索引法の動的な操作能力は不要であること、科学技術データベースへのアクセスには個々のレコードへの順操作が多いがこれには従来の索引法は有用でないことがあげられる。この問題にたいして、科学技術データベースの特性に合った方式として提案したものがbit-transposed-file であり、本方式により従来の格納構造を持つＤＢＭＳに比べて時間、空間の点で１桁向上している。本方式は、１）インデックスエンコダー、２）転置ビットベクトルロケイタ、３）ビットベクトルオペレータ　からなり、各々、各レコードフィールドのビット列への変換、ビット列の転置形式での保存、３）ビット列への操作　をおこなう。

２）　［メタデータ］　［１］

メタデータは、いかにデータが形式化されて格納されているか、データはどういう意味を持っているかを記述するものであり、計算機側でのデータ操作のための非手続き記述情報とすることができる。メタデータのメタデータも可能であり、メタデータの内容と構成に関する記述である。何れも関係表として格納可能である。メタデータに関する問題には、データとメタデータとの同期化である。データとの記述矛盾を避けるために、SELF-DESCRIBING-FILEとそれを用いたツール群が必要である。SELF-DESCRIBING-FILEとは、データだけでなくメタデータをも同一格納形式で記述可能としたファイルであり、上記ツールの結果もまたSELF-DESCRIBING-FILEとする方法である。実現例として、ＣＯＤＡＴＡファイル形式とデータベース操作ツールを開発している。

３）　［多次元データ構造］　［２］

２－３で示す様に、科学技術データベースの格納構造として多次元データ構造が適している。しかし、多くの多次元データ構造方式には適・不適があるためデータ特性と効率との関係を検討している。また、データベースを多次元的に格納するには、　１）１次元化のための順序付け　２）セル分割　３）圧縮を検討しなければならず多次元データ構造はこのうち１）　と２）とに対応するものである。検討の結果、ＱＵＡＤ／ＯＣＴトリーは次元に対してシンメトリックアクセスが可能であり、一様でないデータに対しても良好であるが、相関のあるデータに対しては格納空間アクセス時間の増加をまねく。また、ＫＤトリーは、相関のあるデータに対しては良好であるものの、一様でないデータに対してはトリーをバランスさせることができない。多次元Ｂトリーはあきらかにシンメトリックでない。最後にＧＲＩＤファイルのような直交分割法については、相関のあるデータに対しては効率上の問題がある。

２－３。　科学技術データベース　［４］

ここでは、ＬＢＬで行なわれている科学技術データベースの活動を中心に科学技術データベースの状況を説明する。

（１）　データモデリング

科学技術データベースのデータモデリングからの特質は以下の４点である。

－　コンプレックスデータタイプ
－　意味データモデル
－　時制データ
－　メタデータ

以下にメタデータを除く３点に関して説明する。

［コンプレックスデータタイプ］

科学技術データベースでは、ベクトル、マトリクス、時系列データのようなコンプレックスデータタイプの操作が必要である例は多い。最も基本となる文字、数値からその集合体としてのベクトル、マトリクス、さらに多次元空間内の表、グラフへ複合化し、そしてヒストグラムは統計分析には必要であり、実験データ記述のためのテキストデータも不可欠である。単一のシステム、データベースではこれらすべての要求を満たすことはできないと予想される。

［意味データモデル］

より高次の意味を表現しなければならないことがある。例えば、多次元多階層空間のモデリン

グや回帰分析結果の係数、分散行列、残差ベクトルなどは元のデータとの関係を保つ必要がある。科学技術データベースは複数のデータベースの集合となるため相互を関係づける意味情報が必要である。例えば、観測データ、時間とともに変化する温度、圧力の計測データ、や検知器の校正データからなるデータベースで、各データ間の関係を保存しないと、例えば、特定の観測データに対する、その観測に使用された検知器の特性を考慮した分析をすることができなくなる。

［時制データ］

時制データは、科学技術データベースでは本質的である。従来の商用ＤＢＭＳでは、更新された情報のみを対象とすればよく、履歴情報などの時制データは重要でない。このため、この種の情報をＤＢＭＳ内に保存することはできるが、時間軸にそったアクセスはサポートされず、単に、特定時刻で切り出されたデータ部分集合へのアクセスを許すにすぎない。観測データの時間要素を多次元空間の次元とすることはできるが、規則的に観測されるデータでなく離散的情報（イベント履歴情報）や非規則的なデータ（ランダムなノイズ）の操作は困難である。また、データがグループごとに異なる割合で発生するときにも問題がある。たとえば、磁気場の測定が１分ごとに行なわれるのに対し、観測が１秒おきのとき、分析にあたってこれらのデータ間の相関を考慮する必要があり、効率的な格納方法を要する。このため、静的データと動的データとの結合を考慮したモデリング、ユーザにデータの意味を明確に掴められるように様々な種類のデータのモデリング、ユーザの操作インターフェースなどが必要である。

（２）　物理構造

以下に、科学技術データベースの特質を物理構造で考慮するには、以下の３点が必要である。

－　多次元データ構造
－　データ圧縮
－　科学技術データベース操作

［多次元データ構造］

科学技術データベースの多次元性は、大きな特徴である。例えば、物理実験にみられる、条件を様々に変化させて計測データを取得する場合、その計測データを得るための条件は多次元空間を構成し、計測データはその空間内の１点に相当する。また、時間や空間も当然、多次元空間の要素である。このため、論理レベルだけでなく物理構造までもこの多次元空間を意識することが重要である。単に、従来のＤＢＭＳのとる複合キーの構成という水準ではない。

［データ圧縮］

科学技術データベースのデータの大量さとスパース性は、データ圧縮の必要性と効果を示すものであり、圧縮によるアクセス効率の問題とのトレードオフでも優位にたつものである。従って、科学技術データベースの格納にあったっては、データ圧縮を必ず考える必要がある。

［科学技術データベース操作］

科学技術データベース特有あるいは従来のデータベース応用分野にないアクセスは、１）サンプリング　と　２）アグリゲート　と３）転置である。たとえば、観測データから統計操作を要求しようとする際まず問題となるのは、従来の操作にサンプリング操作が欠如していること、あるいは、あったとしても時間がかかることである。このため、たとえば、ＪＯＩＮ後にサンプリングしようとすると内部ではＪＯＩＮすることなくサンプリングするようにするとか、物理構造のレベルでのサンプリング法を考えることが重要である。また、アグリゲートに関しては、ＭＩＮ，ＭＡＸ，ＡＶなどのほか、分散や中心数も必要であろう。転置についてはデータベースの多次元性と関連しており必要なデータ（表）の行と列を入れ替える操作である。

以上のべた科学技術データベースの研究項目について、ＬＢＬでは個々に研究をすすめており、データモデリングについては、１）時制データのモデリング（Shoshani and Kawagoe），２）Ｅ－Ｒモデル中心の意味表現（Shoshani他）を進めている。一方、物理構造に関しては、１）各種データ圧縮技法のシステム的な評価（Olken）　２）転置とアグリゲートのためのアルゴリズムの開発（Rotem and Olken）（データ量に依存するが１－２　パスアルゴリズムであり、従来のマルチパスソート法にくらべて計算量が少ない。）　３）関係演算ごとのサンプリング技法の開発（Olken）　４）多次元データ構造におけるセルのオーバラップ化とその最適化方式　などの研究を行なっている。

３．　時制データモデリングと格納構造

本章では、科学技術データベースで用いられる時制データに関してその特性とデータモデリング及び格納構造について記述する。

３－１　時制データ

時制データをデータモデリングの視点から眺めたとき、以下の特徴が挙げられる。

－　データの意味：　時制データは、時間データとその軸にそった時制データ列から構成される。
－　時制データ列：　時制データ列は、そのデータタイプから値集合と実体群集合に分けられる。
－　時間データ　：　時刻に相当し、規則的／不規則的、点／区間、連続／離散、デンス／スパース、に各々分けることができる

以下に上記の項目を詳しく説明する。

まず、時制データとは通常時系列データにみられるようにＦ＝Ｘ（ｔ）　ｔ＝１、２、・・・の形式をもつと考えられるが、より一般的な表現をすれば、時制データＦは　時間データ集合ｔと時制データ列　Ｔｔから構成される。すなわち　Ｆ＝（ｔ、Ｔｔ）と考えることができる。例えば、時系列データ　ＦはＦ＝（ｔ、Ｘｔ）であり、検知器Ｄの観測Ｏでの使用状況履歴ＧはＧ＝（ｔ、（Ｏ，Ｄ）ｔ）　と書くことができる。尚、ｔの時間データおよびＴｔの時制データ列については以下に説明する。

時制データ列は、各時刻での時制データインスタンスの時間軸上で順序付けした集合である。このインスタンスは上の例に示すように具体値に相当するものと実体群（正確にいえば、複数実体のインスタンスの集まり）に相当するものがある。従って、一般に時制データ列Ｔｔは、（ｓ１、ｓ２、・・・、）ｔ　とかくことができる。ここで、（）ｔは（）内の要素が時間データｔで順序付けられていることをしめす。また、ｓ１、ｓ２は各々サロゲートと呼び、データ、実体に共通の概念（オブジェクト）のインスタンスを示す。ｓ１、ｓ２のタイプをＳ１、Ｓ２で示す。このような統一した概念で時制データ列で表現することに関しては３－２で触れる。

次に、時間データについては、種々の性質がある。多くの時制データを調べるとき、時間データの多様性に気がつく。したがって、ここでは格納構造および操作に影響を与えるという面からこの時間データを分類する。まず、第１の視点は、規則性である。時間軸上で時間データが規則性を有して存在するか否か。その規則性も等間隔か否かという点である。例えば、定期点検の結果では等間隔の時間データを持つし、機器の故障履歴では不規則な時間データをもつ。第２の視点は、時間軸上での時間データのデータタイプである。すなわち、時間データの具体値が点であるか、区間であるかである。たとえば、機器の故障履歴では時間データは点であるが、機器の稼働状況の場合には区間と考えられる。しかし、この分類は単に、時間データだけをみること以外に、時制データのレベルで考える必要もある。例えば、温度変化では測定時点は点であるが意味的に時制データは次の測定時点までそのデータ値が継続されているものと考えられる。、従って、実際には、この違いを考慮して点か区間かを選択しなければならない。次に、第３のしてんは、時間データの連続性である。時間データが点の場合は、無条件に離散であるが、区間の場合には連続性の内容を考えなければならない。すなわち、上記温度変化では常に時間データに対応した時制データ列要素が存在するが、機器の故障履歴では、機器の故障期間は全期間の部分集合であり、時間データは区分連続と考えることができる。第４の視点は、時間データの濃度に関するものである。これは、時間データが時間軸上でどの程度互いに近接しているかということである。この視点は、他の時間データとの相対的なものであると同時にデンス／スパースの選択はかなりあいまいではある。例えば、観測条件において、温度と圧力の時間的変化をデータとして保持しようとして、温度については１分おき、圧力は１秒おきに計測されると、圧力データの時間データは温度の時間データよりデンスであるといえる。

これらの４視点は、時間データ及び時制データの意味を掴むためや、格納構造を決定する際に、有用である。尚、時制データの特性のひとつであるが、時制データでは更新・部分削除操作はなくデータ追加がない場合かデータのアペンドが発生する。

以下、３－２では上記の特性を考慮したデータモデリングを述べ、そのデータモデリング結果と上記４視点と関連させた格納構造について３－３で説明する。

３－２．データモデリング

時制データモデルに関する研究は最近非常に活発化している［１９－２８］。一つのアプローチは、関係データモデルの拡張としてタップル間の順序関係［２２］、操作言語へのＷＨＥＲＥ等の時間に関する概念の導入［２３］、属性と時間の関係付け［２５］などが行なわれている。他のアプローチは、ＥＲモデルの拡張であり、ヒストリー概念の導入［２１］がある。第３のアプローチは、時制データを時間、実体（関係）、インスタンス（タップル）の３次元空間で表現するものである［２７］。第４のアプローチはロジックを用いた表現［１９］であり、特別の論理を使用いて時制データの制約、問い合わせを行なうものである。

時制データモデルを利用者への時制情報の伝達（データベース内容の記述）及び、格納構造への決定のためのものという立場にたてば、１－３のアプローチの中から選択しなければならない。ここで、第１のアプローチを考えたとき、関係データモデルの時制データ表現での欠陥が明確となる。例えば、観測と検知器との状況履歴の表現のためには、観測・検知器表（観測Ｎｏ，検知器Ｎｏ，時刻）となるが、これは（観測Ｎｏ，（検知器、時刻））とも（（観測Ｎｏ，検知器Ｎｏ）、時刻）とすべきかが不明確であるし、従来の関係モデルとの違い、効果もあまり明確ではない。一方、ＥＲモデルでは、実体（観測）と実体（検知器）との関係（観測・。検知器）に時刻を属性としてもつ構造となる。このとき、例えば、観測データの構造は、実体内の属性となり上記関係モデルと全く同一の問題となる。また、第３のアプローチでは、実体（観測Ｎｏ，検知器Ｎｏ）Ｘインスタンス Ｘ 時間 の空間を考えるのであるが、実体内に複数の時間依存のデータがある場合の構造の表現には上記空間表現を工夫して定義しなければならない。

時制データが３－１で述べたように時間データと時制データ列で集合であることを考えれば、、第２のアプローチのＥＲモデルにおいて、関係（Relationship）に時制データが関係することがわかる。すなわち、実体は時制データとは関連しない。属性については、３－１の時制データの記述のように時制データ列の要素として属性の具体値があることから、時制データに関連するのであるが、この場合の属性を関係の場合とおなじように扱うことが可能である。つまり、属性を実体と同じレベルで扱える概念を用いる。このために、３－１で用いたオブジェクトが関係する。ＥＲモデルの実体をオブジェクトとするのである。これは、特に、新しい概念の提案ではなく、関数データモデル（Functional Data Model）［３０］の概念と類似したものである。従って、時制データは全て関係（Relationship）に関係し、関係の特殊な場合と考えることができる。これを、時制関係（Temporal relationship）と呼ぶ。この概念が定義された関係については、時制データ特有の操作が可能であり、その特性にあった格納構造が決定でき、また利用者の構造定義、理解が容易となる。

この考え方で定義した構造の例を図１に示す。

ここで述べた時制データの一般表現形式を以下に示す。

・時制データは以下の構造を持つ。

（Ｓ１，・・・Ｓｎ｜ｔ｜Ａ１，・・・Ａｍ）

ここで、Ｓｉは オブジェクトｉのサロゲート（インスタンスを示す識別子）であり、ｔ時刻（時間データ）で、Ａｉはこの関係の属性をしめす。

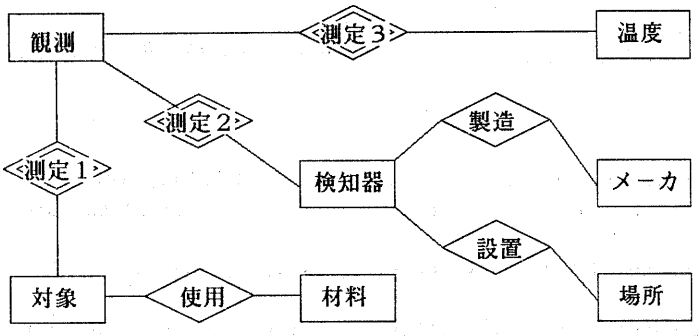
通常は、ｎ＝２、ｍ＝０であるから、（Ｓ１，Ｓ２｜ｔ）と書ける。

時制データ特有の操作としては、Correlation がある。これは、二つの時制データに対して同一時間軸上で時制データ列のマージ、演算等を行なうものであり、以下の式で定義される。

・Correlation 操作

二つの時制データ ＴＳ１＝（ｔ１、Ｔｔ１）とＴＳ２＝（ｔ２、Ｔｔ２）より新しい時制データＴＳ３＝（ｔ３、Ｔｔ３）を生成するCorrelation 操作Ｆは ＴＳ３＝Ｆ（ｔｏｐ（ｔ１，ｔ２）、ＴＯＰ（Ｔｔ１，Ｔｔ２））である。ここで、ｔｏｐは二つの時間データへの新しい時間データ生成の操作である。例えば、ｔ１＝｛１、２、３｝でｔ２＝｛１、４、５、６｝のときｔ３＝｛１、２、３、４、５、６｝とするためのｔ１成分にｔ１にないｔ２の成分を加える操作がｔｏｐに相当する。一方、ＴＯＰは、ｔｏｐで得られた時間データよりＴＳ１、ＴＳ２の結合を行なう操作である。

この操作は科学技術データベースへの操作として有効であると考えられる。たとえば、２つの時系列データの相関や各種演算をこの操作で扱うことができることや応用側での処理からＤＢＭＳ側にすることで処理効率を上げることができるようになる。



測定1～3は、時制関係であり、各々、観測に使用した対象、検知器の利用状況、温度変化の時間データを含む。時制データ特性は 1D(離散，規則，点)，2は(連続，不規則)，3は(連続，規則点)である．尚，ここでは 観測条件のみを示し．観測結果は含まない

図１　時制データの記述例

３－３　格納構造

時制データの形式を既に３－２で説明したが、この一般形式、特にｎ＝２、ｍ＝３の場合の格納構造について以下にしめす。

まず、時制データが（規則的、点、（実体、値）、アペンドなし）の場合を考える。このとき、アクセスとして特定実体に関して時間軸にそった値の列の操作頻度が高いものとする。このとき、格納構造は単純である。保存すべきデータは時間データ、時制データ列全てでなく単に値でよく、この値集合を実体毎にレコードへ格納していく。実体サロゲートと時刻より対応する値をもつ含むレコードのアドレスを計算で求めることができる。このように値のみの保存であるためレコード内の格納率があがり特定実体に関する指定値をとる時間の操作も効率の向上が期待できる。もし、時制データの濃度にばらつきがあり、時間データに対してスパースの場合は適当な圧縮法［１３］によりより少ないレコードで保存可能である。しかし、この場合レコードアドレス算出のための索引レコードを必要とする。

次に、時制データが不規則の場合には不規則な時間データを規則的にする変換情報を加えるだけで上で述べた構造をそのまま使用できる。時制データがアペンドされる場合には、レコードアドレスを論理的なものとして物理ブロックとの対応をとり、レコードフルになったとき新しいブロックを確保する。また、データの増加の仕方がデータによて変わるとき時制データ列のグルーピングが必要である。そして、時制データが区間のときは、時間データの保存が必要であるほか区間［ａ、ｂ］をａ点、ｂ点にわけて保存し、開始点、終了点をしめす必要がある。上の場合に比べてレコードアクセス回数の増加の可能性がある。

時制データ列の要素が二つとも実体の場合には、いずれかの実体からのアクセス頻度が高い場合、他の実体を上の場合の値と同様にして扱えばよい。しかし、アクセス頻度に差がない場合には、実体組ごとに時間軸上での存在情報をレコードへ格納し実体組への索引を考えることにより実現できる。このときは、時制データとして０／１情報のみ保存すればよいため効率的格納が可能である。また、時制データ形式で属性として空間に関係したものが定義されたときは、上で仮定した時間軸にそったアクセス頻度の高さが失われるため、時間データとこの属性を多次元空間として格納する必要がある。このとき、２で述べた多次元データ構造を利用できる。［１８］

ここでは基本形式への格納構造を述べたが、他の形式についてもその拡張あるいは他の格納構造（ハッシングなどの索引技法）との組み合わせで決定できる。このアプローチは、分割格納モデル（Decomposition Storage Model ）［１７］にみられるできるだけ単純な基本形式の組み合わせでデータベースを格納しようとするアプローチに基づくものである。

４。　おわりに

本報告では、ローレンス・バークレイ研究所（ＬＢＬ）でのデータベース研究、特にサイエンティフィックデータベースの状況と筆者がＬＢＬ滞在中におこなった時制データのモデリングについて記述した。時制データについては、１）ＥＲモデルでの関係（Relationship）に全ての時間情報を含めることを提案し、２）時制データの特性と特有操作としてCorrelation を説明し、

3）各特性ごとの格納構造案を提示した。
尚、本報告における時制データに関する部分はＬＢＬのArie Shoshani と共同で得られたものである。

［参考文献］

［１］ John McCarthy, Scientific Information = Data + Meta-data, LBL MEMO,1985
［２］ Frank Olken, Physical Database Support for Scientific and Statistical Database Management , LBL MEMO, 1985
［３］ Arie Shoshani, Frank Olken and H.K.T. Wong, Characteristics of Scientific Data bases, LBL MEMO LBL-17582, 1984
［４］ Arie Shoshani, Data Management Issues of Statistical Databases, LBL Memo, 1985
［５］ Arie Shoshani and H.K.T. Wong, Statistical and Scientific Database Issues, LBL-19841, 1985
［６］ A LBL PERSPECTIVE ON STATISTICAL DATABASE MANAGEMENT, LBL-15393, 1982
［７］ K. Kawagoe, Modified Dynamic Hashing, ACM SIGMOD'85,1985
［８］ Edited by M.L. Brodie et. al., ON CONCEPTUAL MODELLING, Springer-Verlag,1982
［９］ D. Dewitt, M. Stonebraker, F. Olken and L. Shapiro, Implementation Technique for Large Main Memory Database Systems, SIGMOD'84, 1984
［１０］ F. Olken, hopt: A Myopic version of the STOCHOPT automatic File Migration Policy, SIGMETRICS'83, 1983
［１１］ A. Shoshani, Characteristics of Scientific Databases, 10TH VLDB, PP147-160, 1984
［１２］ P. Chan and A. Shoshani, Subject: A Directoruy Driven System for Organizing and Accessing Large Statistical Databases, VLD'80,1980
［１３］ S.J. Eggers, F.Olken and A. Shoshani, A Compression Technique for Large Statistical Databases, VLDB'81, 1981
［１４］ H.K.T. Wong, F. Olken, D. Rotem and A. shoshani, Bit transposed file, VLDB'85, 1985
［１５］ H.K.T.Wong and I.Kuo, GUIDE:Graphical User Interface for DB Expoloration, VLDB'80,1980
［１６］ J. Mylopoulous, P.Bernstein nad H.K.T.Wong, A Language Facility for Designing DB intensional Applications, ACM TODS 5,2,1980
［１７］ G.P. Copeland and S.N. Khoshafian, A Decomposition Storage Model,SIGMOD'85
［１８］ S. Khoshafian et.al. , A Performance and directed Taxnomy for Single/Multi key File Structures, MCC tech. memo,1985
［１９］ J. Clifford and D. Warrenn, Formal Semantics for Time in Databases, ACM TODS, 1983
［２０］ 田中、組系列にもとづく履歴データベースモデルとその完全性制約、情処学会誌、１９８５
［２１］ M.R. Klopprogge,TERM An approach to Include the Time Dimension in the Entity-Relationship Model, Proc. of ER Conf., 1983
［２２］ V. Lum et. al., Designing DBMS Support for the Temporal Dimension, SIGMOD'84, 1984
［２３］ R. Snodrass, The temporal Query Language TQuel, PODS'84,1984
［２４］ R. Snodrass,A Taxonomy of Time in Databases, SIGMOD'85, 1985
［２５］ J. Clifford and A.U. Tansel, On An Algebra for Historical Relational Databases : Two Views, SIGMOD'85, 1985
［２６］ T.L. Anderson,Database Semantic of Time, PhD Th., CSD, U. of Washngton, 1981
［２７］ G. Ariav and H.L. Morgan, MDM:Handling the Time Dimension in Generalized DBMS, Tech. Rep., U. of Pennsylvania, 1981
［２８］ B. Rubenstein, Indexes for Time-Ordered Data, Tech. Memo, 1985
［２９］ A. Shoshani and K. Kawagoe, Temporal data model, Technical Memo, LBL, 1985