

マルチデータベース環境におけるメタ情報収集方式の検討

7K-1

町原宏毅 加納直哉 岡田英二

NTT情報通信研究所

1. はじめに

ネットワーク技術の進展と情報サービスの急速な発展により、各企業では種々のデータベースが構築され、数多く存在するようになってきた。また最近では、このような複数のデータベースに蓄積された情報を経営分析等の目的で有効活用したいという要求も増してきた[1]。

しかしながら、ネットワークに接続した複数のデータベースにアクセスするためには、分散しているメタ情報（データの属性、格納場所、サイズ等のデータに関する情報）を一元的に管理するGlobal Information Resource Dictionary（以降Global-IRD）と呼ばれる統合情報資源辞書が必要となるが、その構築はデータ管理者が行っており、多くの稼働を必要としていた[2]。

そこで、本稿では、ネットワーク上に存在する複数かつ、異種のデータベースが保有するメタ情報を自動的に収集し、Global-IRDに蓄積するメタ情報収集方式を提案する。

2. マルチデータベース環境における情報資源管理

最近のデータベースの動向としてデータベース毎に閉じて使用していた環境から、インターネットで使われている技術を利用した社内向の情報システムの構築（イントラネット）が注目されている[3]。

代表的な形態には、World Wide Web(WWW)を経由してのデータベース利用がある（図1）。

これにより、数多くのデータベースにアクセスできる（マルチデータベース）環境はできたが、このような環境で検索を行おうとすると以下のような管理をすることが必要となる。

- ・データベースへの接続方法の管理。
- ・スキーマ情報の管理。
- ・テーブル間のリンク情報管理。
- ・データベースの追加や、スキーマ変更の管理。

このような管理をGlobal-IRDで行うことにより、ユーザは、接続方法、データベース名、テーブル名、カラム名等を参照できるようになり、マルチデータベース環境での情報検索が可能になる。

3. 現状の課題

Global-IRDを作成しようとした場合、以下の課題がある。

(1) DBMSの異種性の解消

使用するデータベース管理システム（DBMS）によってアクセス手順やメタ情報の形式が異なるた

め、各データベースのメタ情報を収集して一元管理をしようとする、様々な変換が必要となる。

従来の技術では、個別のデータベースのメタ情報をいったんダウンロードし、次にそれをGlobal-IRDに送り、その後、構造を統一化させ、最後にGlobal-IRDに保存する、といった一連の操作を管理者が行う必要があった。管理するデータベースの数が多くなれば、かかる手間は膨大になってしまう。

(2) 個々のデータベースのメタ情報とGlobal-IRDの情報の同期制御

新たにデータベースが追加されたり、個々のデータベースのスキーマ変更が生じた場合に、Global-IRDの情報と食い違ってしまう。これらの間に違いが存在するとGlobal-IRDを見て検索しようとしても、実データベースにアクセスできなくなってしまう。

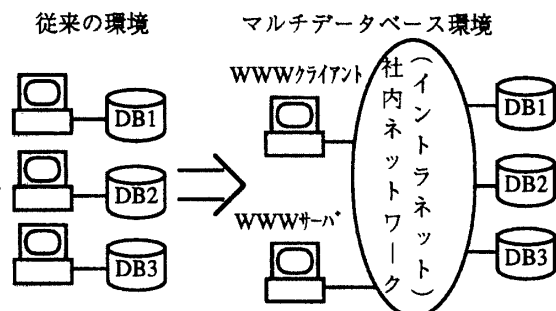


図1：データベース環境の変化

4. メタ情報収集方式

前節に示した課題を解決するためのメタ情報収集方式を提案する。

本メタ情報収集方式は、事前に自動収集するための情報（接続情報）を登録し、各DBMS用のメタ情報収集プログラムを作成しておくことにより、自動化を図ることにした。

4.1. 接続情報

各データベースに接続するための情報として、データベース名、データベースの所在（ホスト名）、DBMS名、ログイン名、パスワード等を管理者が事前にGlobal-IRDに登録する。

接続情報の詳細を表1に示す。

表1：接続情報の例

| DB名   | DB1       | DB2         | DBx |
|-------|-----------|-------------|-----|
| ホスト名  | host1     | host2       | ... |
| DBMS名 | oracle7   | informix5   | ... |
| 接続方式  | oraclenet | informixnet | ... |
| ユーザ名  | DB1       | DB2         | ... |
| パスワード | ...       | ...         | ... |
| 収集契機  | 000**2    | 0011**      | ... |

Hiroki MACHIHARA, Naoya KANOU, Eiji OKADA  
NTT Information and Communication Systems  
Laboratories.  
Meta-Information Collection Method under  
Multi-Database Environments.

- ・DB名：ネットワークにつながったデータベースを一意に識別する名前。
- ・ホスト名：データベースが存在するコンピュータの名前。
- ・DBMS名：データベース管理システムの名前。
- ・接続方式：データベースに接続するための手順に付けられた名前、oracleなら接続識別子の名前、informixやsybaseならデータベースサーバー名を示す。
- ・ユーザ名：データベースの利用者を識別する名前。
- ・パスワード：データベースの利用者を確認するパスワード。
- ・収集契機：メタ情報を収集するためのプログラムを起動するタイミング。

設定フォーマットは、"分(min) 時(hour) 日(day) 月(mon) 曜日(week)"である。

設定する値の内容は、定期的に行われるコマンドの設定を行うUNIXのcrontabファイルと同じにしたもので、週1回、月1回等の設定ができる。

例1：00 0 \* \* 2

毎週火曜日の午前0時に起動する。

例2：00 1 1 \* \*

毎月1日の午前1時に起動する。

参考："\*"設定は、そのフィールドに設定される値の全てを意味する。

曜日の値(1~7、月曜が1)

#### 4. 2. メタ情報収集プログラム

各DBMSで管理しているメタ情報は、DBMS毎にテーブル名、カラム名が固定であり、ネットワーク上のデータベース数が増加してもDBMSの種類はそれほど増えないため、各DBMS毎にメタ情報収集プログラムを用意することとした。

データベース名や接続方式等の情報をパラメータとして受け取るようにし、抽出対象として指定されたデータベースに対してSQL文を発行する。

DBMSによっては、メタ情報の中にカラム数やテーブル行数の無いものがあるため、カウントする処理を行って構造を統一する。

#### 4. 3. 処理内容

接続情報を読み込み、その中の情報から、メタ情報収集プログラムを選択し、接続方法やデータベース名等の情報を渡して検索実行する。

抽出されたメタ情報をGlobal-IRDに登録する。

メタ情報を収集する方式の構成を図2に示す。

##### (1) 制御

メタ情報収集プログラムの選択と起動および、全体の処理を監視制御する。

Global-IRDに登録された接続情報を常に監視し、メタ情報を収集する契機になった時点で、対象となるデータベースのDBMS名からメタ情報収集用プロ

グラムを決定し、DB名、ユーザ名等の情報を渡す。

抽出されたメタ情報(実データベースのメタ情報)とGlobal-IRDのメタ情報を比較し、異なっていた場合、Global-IRDに、その情報を格納させる。収集契機を設定することと、メタ情報の比較を行うことにより、ある程度の自律性を持った同期制御が実現できる。

##### (2) 検索実行/構造統一

抽出対象として指定されたデータベースメタ情報収集プログラムを実行し、メタ情報を抽出しGlobal-IRDに格納できる形式に構造を統一する。

##### (3) 情報格納

抽出してきたメタ情報をGlobal-IRD上で一意に識別できるIDを付与してGlobal-IRDに格納する。

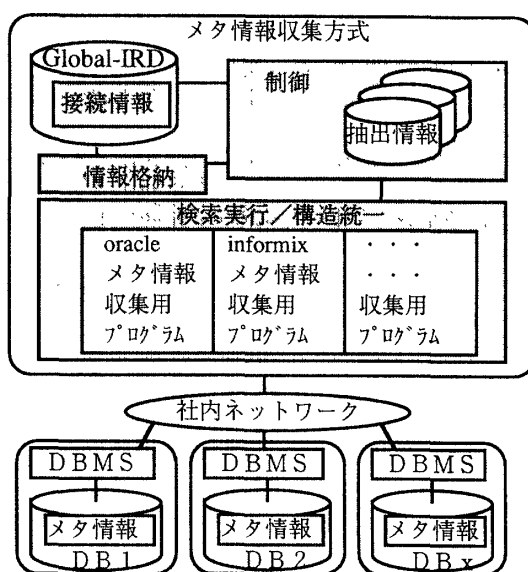


図2：メタ情報収集方式の構成

#### 5. おわりに

事前に管理者が接続情報を登録し、各DBMSのメタ情報収集プログラムを作成しておくことにより、ネットワークにつながったデータベースのメタ情報を自動的に収集する方式を明らかにした。これにより、個々のデータベースに何ら手を加えることなく、自動的にGlobal-IRDを構築できるようになる。

今後は、各DBMSが持っているメタ情報を参照して検索するだけでなく、目的とする情報をキーワード等で効率的に獲得するための仕組みを検討し、この中からGlobal-IRDで管理する情報を明らかにし、その収集方法を検討していきたい。

#### 参考文献

- [1]W.H.Inmon, 初めてのデータウェアハウス構築, トムソンパブリッシング, 1996
- [2]鈴木源吾, 他, イントラネットにおける情報資源管理技術の検討, 1996電子情報通信学会 情報・システムソサイエティ大会講演論文集, D-45, 1996
- [3]Sun World 5月号, 特集イントラネット, IDG communications, 1996