

DB利用性の向上技法

情報システム化技術演習－3
(データベースシステム設計)

第4回

URL <http://homepage3.nifty.com/suetsuguf/>

アドレス: fwhy6454@mb.infoweb.ne.jp

作成者 末次文雄 ©

復習：データ正確性の保証

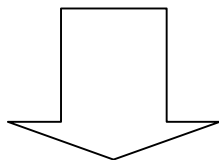
- 入力データの検証では、
 - 全テーブルの入力順序を決めておく
 - ER図の基数を参考にする
- タイムスタンプでは、
 - 登録日は、そのテーブル行の作成日付である
 - 更新日は、重要なデータ項目に対応する
- データの有効日では、
 - そのテーブル行の1レコードの有効期間を示す
 - データ項目ごとに持つのは間違い

補足(有効日設定済みのテーブルの主キー)

方法1: (元のキー+有効日開始)の連結。

欠点として、有効日の変更があった場合、一意で無くなる場合が発生する可能性。

商品コード	価格	有効日(開始)	有効日(終了)
AAAAA	100円	20020101	20020501
AAAAA	110円	20020501	20020515
AAAAA	120円	20020515	20020601
AAAAA	130円	20020601	



110円を「実施せず」に変更。

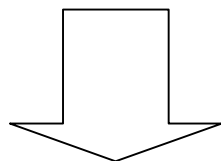
商品コード	価格	有効日(開始)	有効日(終了)
AAAAA	100円	20020101	20020501
AAAAA	110円	20020501	20020501
AAAAA	120円	20020501	20020601
AAAAA	130円	20020601	

補足(有効日設定済みのテーブルの主キー)

方法2: 有効日設定通知番号を設けて、

(元のキー+有効日開始通知番号)の連結。

商品コード	価格	有効日(開始)	有効日(終了)	通知番号(開始)	同(廃止)
AAAAA	100円	20020101	20020501	X01	X02
AAAAA	110円	20020501	20020515	X02	X03
AAAAA	120円	20020515	20020601	X03	X04
AAAAA	130円	20020601		X04	



110円を通知番号X03で「実施せず」に変更。

商品コード	価格	有効日(開始)	有効日(終了)	通知番号(開始)	同(廃止)
AAAAA	100円	20020101	20020501	X01	X02
AAAAA	110円	20020501	20020501	X02	X03
AAAAA	120円	20020501	20020601	X03	X04
AAAAA	130円	20020601		X04	

目次

1. データ利用の推進
 ＜演習問題4. 1＞
2. Usage情報のデータベース化
 ＜演習問題4. 2＞
3. 情報系のデータベース化
4. 企業内全文書のデータベース化
 ＜演習問題4. 3＞
5. 情報検索の技術的側面
6. DB運用管理上の工夫
7. DB設計の留意点
 ＜演習問題4. 4＞

1. データ利用の推進

①EUDの推進（ユーザー自身による開発）

- ・外部スキーマ(VIEW)を準備する---演算結果の仮想表
- ・ヒナ型SELECT文の準備
 - ー単なる検索
 - ー小計、合計、件数カウントetc.
- ・サブルーチンの提供
 - ー時点セレクト機能(有効日、指定日)
 - ー複雑なDB検索(部品構成展開etc.)
- ・INDEXの充実
- ・コード表のDB化(コード、用語・・・)

- ・検索用DBのコピー準備
 - ー基幹業務用とは分ける
- ・データのダウンロード(ユーザーはEXCEL等で加工する。)

補足: SQL(外部スキーマの定義)

- ・複数テーブルから作成することが出来る(ただし、参照のみが許される)
- ・定義時には定義するだけで、**実行時**に仮想のテーブルが作成される。

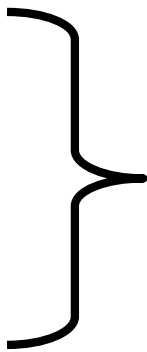
```
CREATE VIEW ビュー名  
    (データ項目,データ項目,...)  
AS SELECT   計算式  
      FROM   表名  
      WHERE  条件式
```

問合せ文(検索文)と同形式

ビューの効用:

- ・**ユーザー利用**
- ・開発効率面
- ・セキュリティ面

② DBの充実

- Usage情報の付加
 - 情報系データベースの作成
 - 通常文書のDB化
 - 関連システムの情報統合（社内システムのDB）
 - 外部DBの利用契約（新聞記事、R&D・・・）
 - DBの利用方法公開（マニュアル、url一覧・・・）
- 
- （後述）

<演習問題 4. 1>

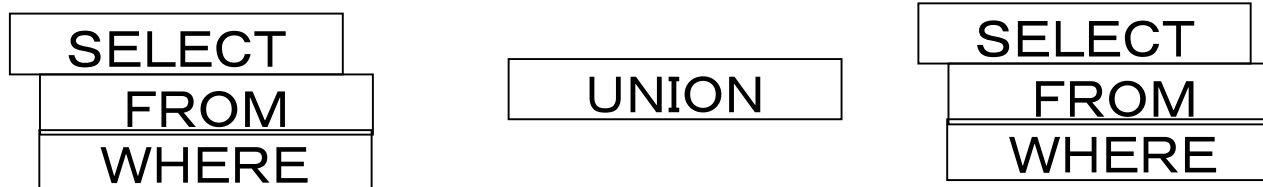
以下の課題について外部スキーマの定義をなさい。
(つまりVIEW作成のSQL文の作成)

課題

- ①課題2、社員の保有スキル(カレント+履歴)に社員氏名が含まれるVIEW。
- ②上記VIEWを使用した、検索用のSQL文を作成せよ。
(あるスキルを保有する社員の取り出し。)

補足：課題4. 1のヒント

- 社員氏名を含むviewが必要
 - 派遣候補者テーブルと保有スキルテーブルの結合 (join)
 - 履歴派遣候補者テーブルと履歴保有スキルテーブルも同様
- Viewはシンプルにし、使いやすくする
 - 上記二つのテーブルを一つにまとめる
 - そのためには、和集合を使う (UNION)



2. Usage情報のデータベース化

2. 1 Usage情報とは？

- ・製品、部品、ソフト、データなどの特徴や用途を明示するデータである(Usage Condition 情報)

2. 2 対象

- ・重要なもの、重要なデータに関して付与する

2. 3 Usage情報が必要な理由

- ・情報の流用、用途を拡大することによる
開発のダブリを防止し、資源の有効活用を図る。
- ・いわば、ノウハウの公開にあたる。
- ・設計者、開発者、作成者は当然熟知しているが、
利用者には分らないものである。
- ・通常、設計者、開発者は当然過ぎて、記録しない。

2. 4 Usage情報の事例

- 自動車の構成部品

トランスミッションA＝

車種Y＋北米向け＋2000ccエンジン＋オートマチック＋酷寒向け

- 日用品

ズボンB＝

若者向け＋腰周り78センチ＋長さ80センチ＋2タック＋
4ポケット＋ベルト通し有り＋撥水＋形状記憶＋茶色

- ソフト部品

和暦から変換＝

西暦4ケタ＋年＋月＋日＋過去1000年間＋将来30年間

- 資料類

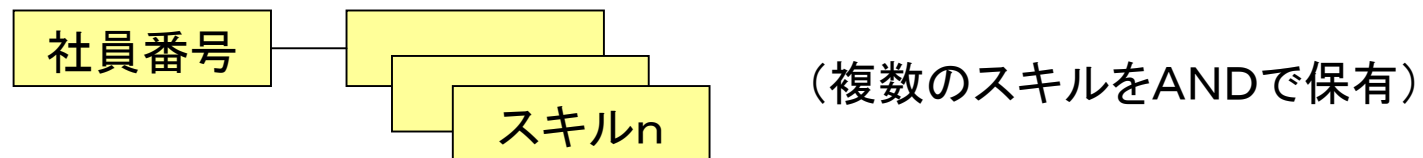
経済分析書C＝

日本＋1990年以降＋10年間＋消費需要予測＋一般向け

2. 5 Usage情報のDB構造

①保有スキルテーブルの場合

- ・このテーブルそのものがUsageDBである。



- ・簡単な、「複数スキルを保有する社員」の取り出し方法

(例)レベル4以上のスキルA+B を保有する社員の取り出し

- ・スキルのインデックス作成(スキル分野コード+スキル項目)
- ・スキルA、レベル4以上の社員番号を取り出し一時保管
- ・その社員番号で、スキルB、レベル4以上の社員を取り出す
(二つのAND条件であれば「副問合せ」句でもOKであるが、
それ以上であれば複雑になる。)

2. 5 (続き) Usage情報のDB構造

②一般的な場合

- ・対象(アイテム)に、項目(位置を固定)ごとに、使用条件を持たせる



(ヨコ・列方向はAND条件)、(タテ・行方向はOR条件)

(符号には、「NOT」がある)

- ・「複数の条件を指定して、アイテム番号を取り出す方法

(例) 項目1・条件A + 項目3・条件C を有するアイテム番号の取り出し

- ・項目1は必ず指定する。項目1にはインデックス作成
- ・項目1・条件Aおよびに項目3・条件Cに合致するアイテム番号を取り出す

<演習問題 4. 2>

以下の課題についてUsage情報のデータベース構造図、検索方法を考えてください。

課題

- ①自分が関心があるものは何か。
- ②それについて、データベース構造図の作成
- ③データ検索の方法

3. 情報系のデータベース化

① サマリーDBの作成

- ・基幹系DBからサマリーする
- ・過去の実績データを集める
- ・時系列のデータに変換する

例示: (売上の日別、週別、月別合計・・・)
(売上の地域別、店舗別・・・)

② データウェアハウスに蓄積

- ・基幹系の明細データを収集
- ・多次元解析(期間別、顧客別、製品別、地域別・・・)
- ・統計処理(数値解析・分析)
- ・データマイニング(相関、規則性の発見)

OLAP

補足：サマリーDBの作成例

店舗別売上明細から月別売上表を作る。

店舗別売上明細テーブル

店舗コード	年月日	顧客コード	売上商品コード	売上個数	登録日
-------	-----	-------	---------	------	-----

商品テーブル

商品コード	商品名	商品価格	有効日(開始)	有効日(廃止)	更新日
-------	-----	------	---------	---------	-----

毎月末に、上記のテーブルを
使って売上集計をして
下記のテーブルを生成する

店舗別・月別売上テーブル(サマリーDB)

店舗コード	年月	売上金額	登録日
-------	----	------	-----

これがサマリーDB構造

(補足) 基幹系DBと情報系DB

	基幹系DB	情報系DB
利用目的	主活動の遂行 (開発、製造、販売、 購買、経理・・・)	戦略立案 企画 分析
情報の特徴	詳細データ	集約データ 時系列データ
データ量	小、中	膨大
ユーザー数	多い	少ない

<演習問題 4. 3>

以下の課題についてサマリー情報のデータベース構造図を設計してください。

課題

- ①課題1、部署コード別の時系列給与総額データ
- ②課題2、スキル別の月別派遣社員数の時系列データ

4. 企業内全文書のデータベース化

①企業内文書の分類・体系化

(まず部門内で実行→次いで全社に水平展開)

②文書は細大もらさず対象にする

(いわゆるペーパーレス化の実施、パソコン1人1台)

(議事録の類も対象。ホット情報であり即日公開原則)

③基幹系システムのアウツプットも対象

(アウツプットのEXCEL化を促進)

④機密管理

(情報の機密等級を決めて、アクセス権限付与)

⑤企業内情報のポータルサイト化、体系化

(全ての情報アクセスの入口)

補足：企業内データの機密管理

- ① セキュリティ・ポリシーを規定（セキュリティを優先、個人情報保護の方針）
- ② 入室制限（コンピュータ室、事務室）
- ③ 重要データは、暗号化する。バックアップを採り保管場所を変える。
- ④ 端末起動時の制限
パスワード、パスワード付きのスクリーンセーバ、指紋検証
- ⑤ データのアクセス権限を制限
職務に応じた権限付与、外注先の人も同様
 - ・重要データは、都度の申請・許可方式
（常時アクセス権を付与しない）
 - ・重要データ（個人情報、財務情報、機密）
 - ・移動時、退職時は削除
- ⑥ その他
 - 情報漏えい防止ソフトウェア、監視カメラ設置（心理的効果）
 - アクセスログ（記録）は長期保管
 - 外部とのアクセス管理（ファイヤウォール）
 - 企業用の電子認証システムの導入
 - ネットワークに繋がらない（最高機密用コンピュータ）
 - 社員教育
 - 監査を受ける（内部監査、外部監査）、公的なセキュリティ認定を受ける

補足：職務に応じたアクセス権限

- 機密等級を決める(5ランク程度)
 - ・個人情報、財務情報、製品開発情報はレベルを上げる
- 情報を分類する(開発、製造、販売、経理・・・)
- 社員ごとにアクセス権を設定
- すべてのファイルの機密等級、分類をする
- 上記をデータベース化する

ファイル名	分類コード	機密等級	有効期間
-------	-------	------	------

社員番号	分類コード	機密等級	有効期間
------	-------	------	------

社員番号	パスワード
------	-------

補足：企業内ポータルサイト

- インターネット技術を使用（イントラネット）
- 通常のWebサイトと同様の設計
- 機密管理のしくみを持つ
 - 社員番号、パスワードを入力
- 企業内の全文書を対象とする
- サーチ手段
 - ハイパーリンク
 - インデックス
- 検索場所
 - 企業内
 - 社外から
- 実際には、各部門のサイトを結合する形式

5. 情報検索の技術的側面

①分類

コード検索

- ・データ検索ともいう
- ・コード、数値、書誌
- ・事前にコード体系を規定
- ・索引には、このコードを使用
- ・検索結果のズレが少ない

テキスト検索

- ・概念検索、テーマ検索ともいう
- ・文献、ドキュメント
- ・検索用語の体系(シソーラス)の事前の検討が必須
- ・テキスト中の単語を索引とする(手動、自動あり)
- ・検索結果のズレが大きい
- ・他の絞込み方式を併用(アクセス頻度など)

マルチメディア検索

- ・画像、音声、音楽
- ・キーは図形、画素、輪郭、音波、メロディなど
- ・未だ実用化できず(出だしのメロディ程度)

5. 情報検索の技術的側面(続き)

②手法

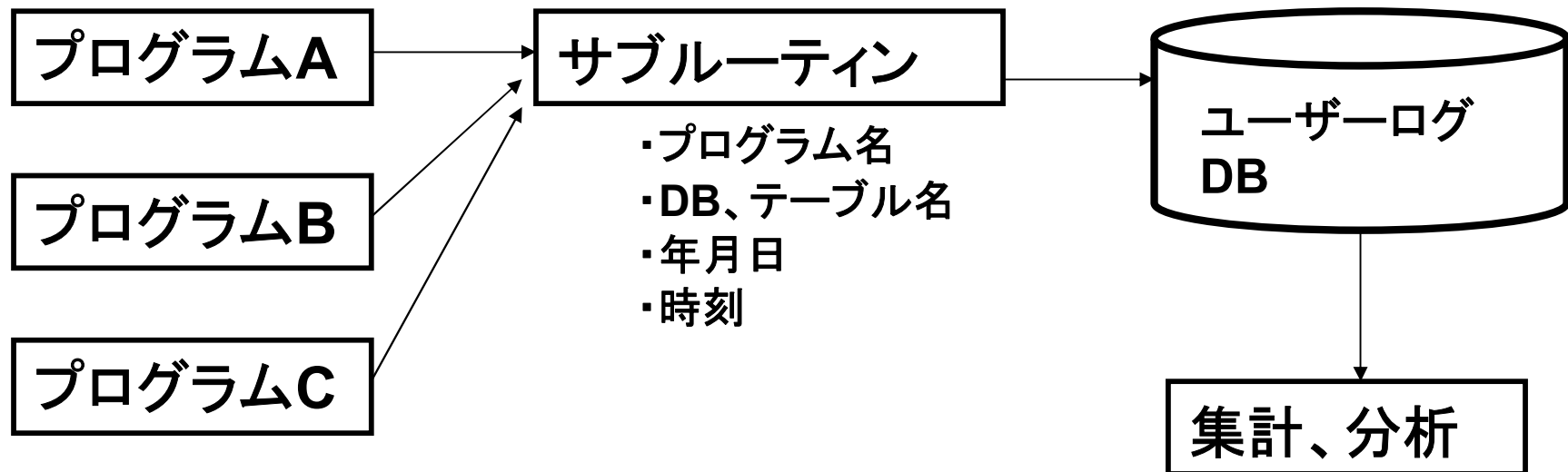
- 索引方式
 - ・索引用語の表現力が高いこと(適確)
識別力が高い(類似の区別)
- トランケーション
 - ・幅を持たせて索引を拾い上げる手法
 - ・省略指定(* □□ * △△ *)
- 論理演算子
 - ・検索条件式(AND、OR、NOT)
- ハイパーテキスト
 - ・リンクを貼る方式
- プルダウンメニュー
 - ・検索キーを提示する
- 文字列検索
 - ・検索時間が遅い(商用では実用例少)
 - ・通常は、単語を索引にする

6. DB運用管理上の工夫

- 6. 1 DB利用状況の把握
- 6. 2 アクセス・スピードの計測
- 6. 3 DBスペース状況の把握

6. 1 DB利用状況の把握

- どのDB、テーブルを誰が、何のために利用したのかを把握する。
- DBMSのログファイルは、煩雑な形式であり、アプリ責任でログを採取する方が現実的。
- 但し、ログへの書込みオーバーヘッド発生。



6. 2 アクセス・スピードの計測

- ① 実地に計測
 - ストップウォッチでTATを計測
- ② システムを利用
 - システムの各部分に一時的にトラップを入れて、個々の経過時間を計測
- ③ ユーザーログDBを使用
 - ユーザーログDBに、アクセス開始、終了のタイムスタンプを書き出す

6. 3 DBスペース状況の把握

- ①定期的に、DASDスペースの使用状況を調査
 - ープライマリー領域
 - ーあふれ領域(オーバーフロー領域)
- ②残スペースが少なくなれば、拡張する
- ③DB、テーブルの再編成
 - ー併せて、不要データの削除
 - ーカレント、ヒストリカルデータに分割
 - ー実施サイクルはデータ量増分に応じる。

7. DB設計の留意点

①DB構造変更に柔軟に対応できるシステム

- ・アプリプログラムは必要なデータ項目のみ取出すこと
- ・外部スキーマ(View)の利用

②パフォーマンス設計

- ・データ量分析、データ量増分予測を励行
- ・データ利用の増大を予測し、対処する

③データの精度保証

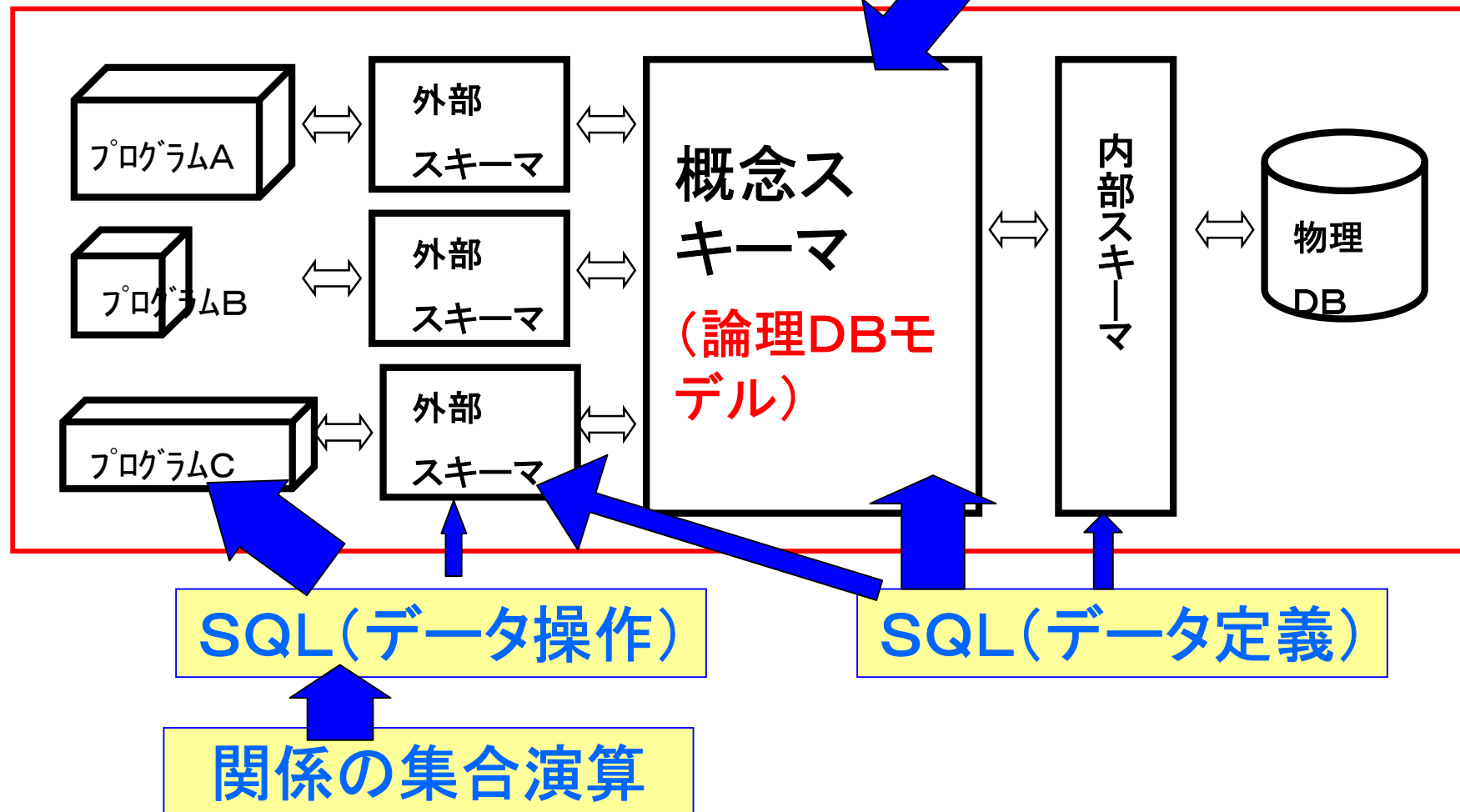
- ・継続的なデータチェック機能の強化、アプリ保守

④トレード・オフの勘案

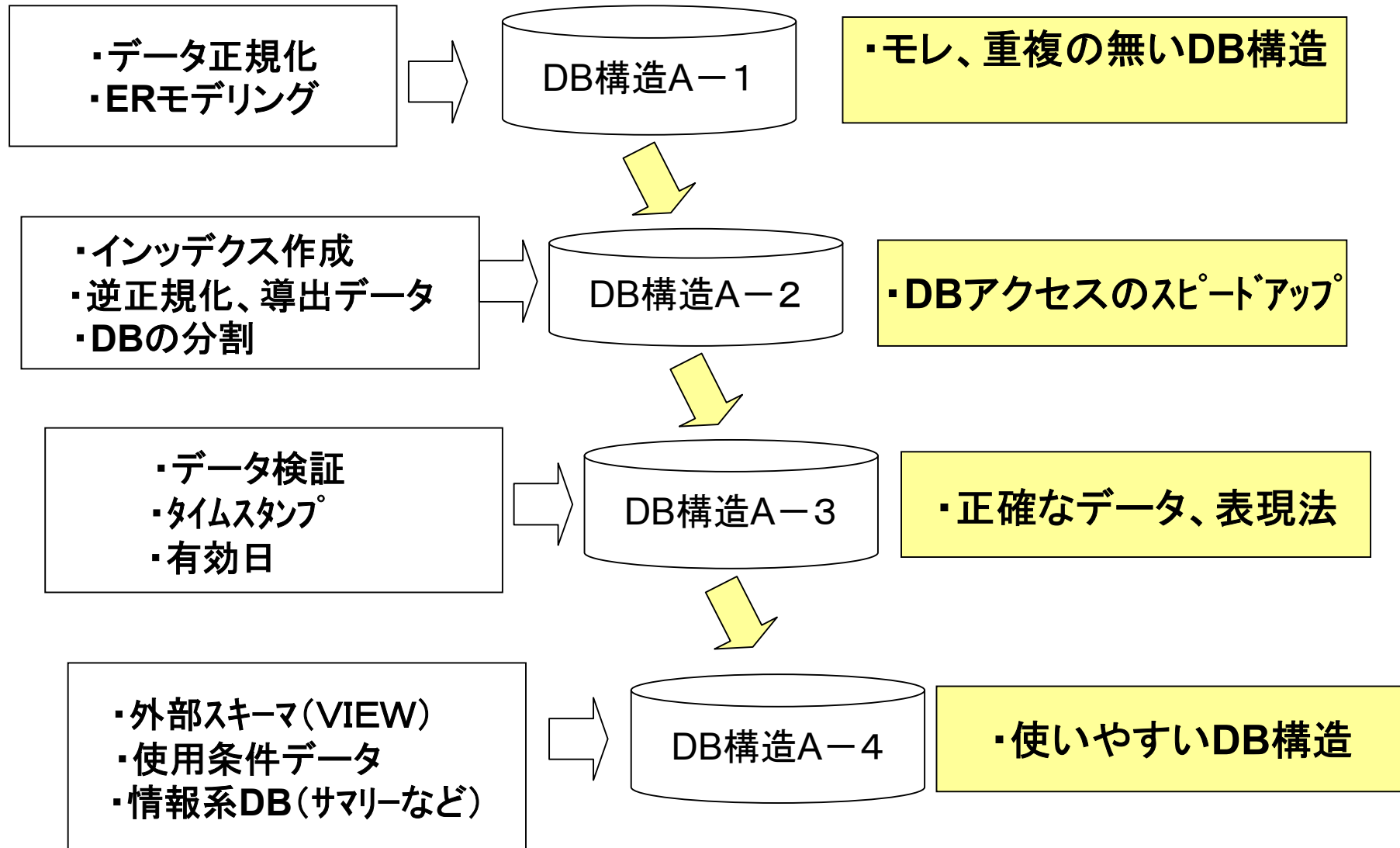
- ・正規化のレベル、インデックスと更新時間
- ・リカバリー時間と2重化

(補足)スキーマと技法・SQLの対応

関係の集合演算 → データ正規化 → ERモデル技法



(演習のまとめ)



<演習問題4. 4>

当演習の感想文（内容レベル、役に立ったかどうか、などを含めて、何でも結構です。）

- ・メール(fwhy6454@mb.infoweb.ne.jp)
- ・期限： 1週間以内
- ・解答例：ホームページ上に公開（ただし4. 4除く）

参考

- 仕事の進め方
 - －何事も計画を立ててから進める
 - －書き出してみる(リスト、関連図)
 - －自分の考えを話し、人からアイデアをもらう
 - －5W3H(What, Why, Who、When、Where, How to, How much, How many)
- SEの心構え
 - －好奇心→もっと知る→対象業務を好きになる
 - －忍耐力、持久力、スタートしなければゴールに着かない
 - －プラス思考
- トレンド
 - －プロジェクト・リーダー経験者
 - －データベースのスキルを持つエンジニア
 - －業務系システムの構築・運用経験者など

“力のある”エンジニアへの需要の伸びは顕著