



IBM SWG

『データマイニングによる診療情報分析のご提案』



2008年7月16-18日

日本アイ・ビー・エム株式会社
SW事業 IM事業部

1. データマイニングとは
2. データマイニングを実現する技術
3. 産業界でのデータマイニング事例
4. 診療情報、治験情報などへの利用提案
5. データマイニングを実現するIBMソリューション

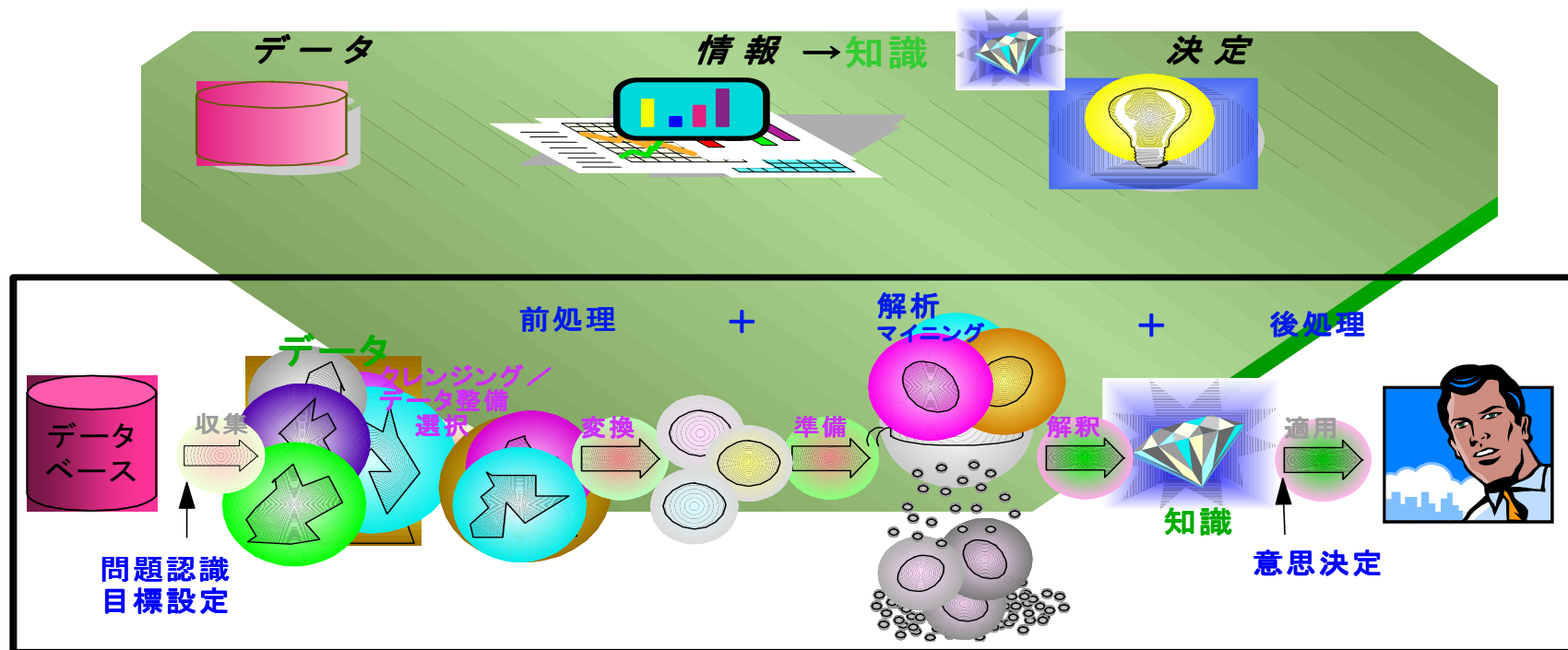


1. データマイニングとは

データマイニングとは、

「以前には知られていなかった、有効で、活用できる**知識**を大規模データベースから抽出し、その**知識**を用いて重要なビジネスの決定を行うプロセス」である。

(P. キャベナ他著、河村佳洋他監訳、データマイニング活用ガイド、トッパン、1999)



「 風が吹けば桶屋が儲かる！？」

この関連をデータマイニングで解けるか？



1. 風で土ぼこりが立つ
2. 家の中で宴会をする人が増える
3. 三味線の需要が増える
4. 猫皮が必要になり、ネコが殺される(三味線は猫皮が使用されていたらしい)
5. ネコが減ればネズミが増える
6. ネズミは箱をかじる
7. 箱の需要が増え箱屋が儲かる



⇒ 風と土ぼこり、猫の数とねずみの相関、ねずみの箱かじり力等

説明変数と目的変数のデータ蓄積があれば予測は可能かもしれない。。。。

この説明変数と目的変数の関連を浮かび上がらせる技術がデータマイニング。。。。

2. データマイニングを実現する技術

基礎統計： 要約統計、回帰分析

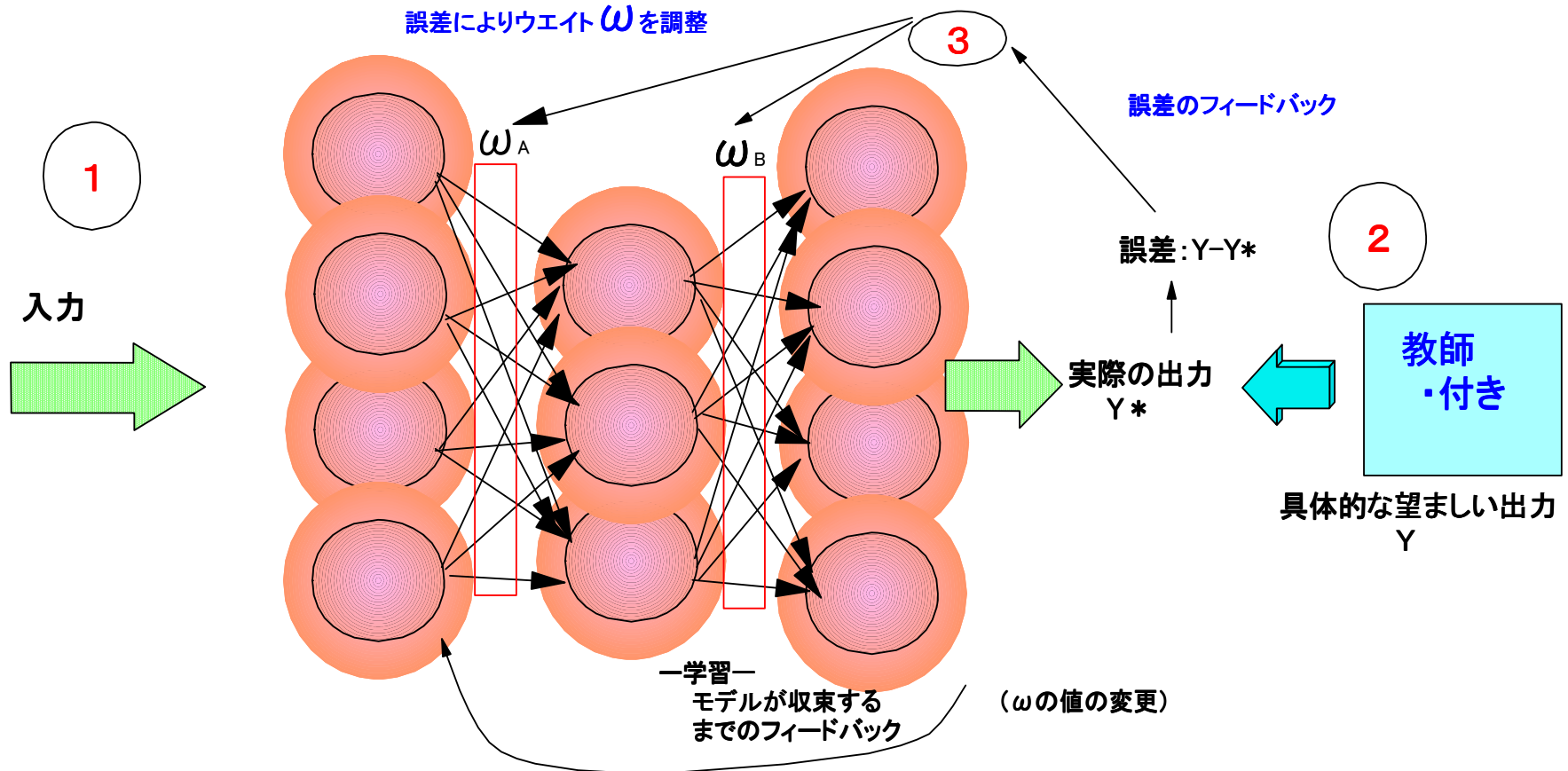
ニューラルネットワーク：

- 教師つき学習（バックプロパゲーション）
- 教師無し学習（コホーネンマップ）

データ並列処理

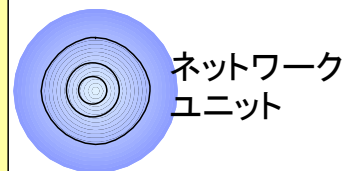
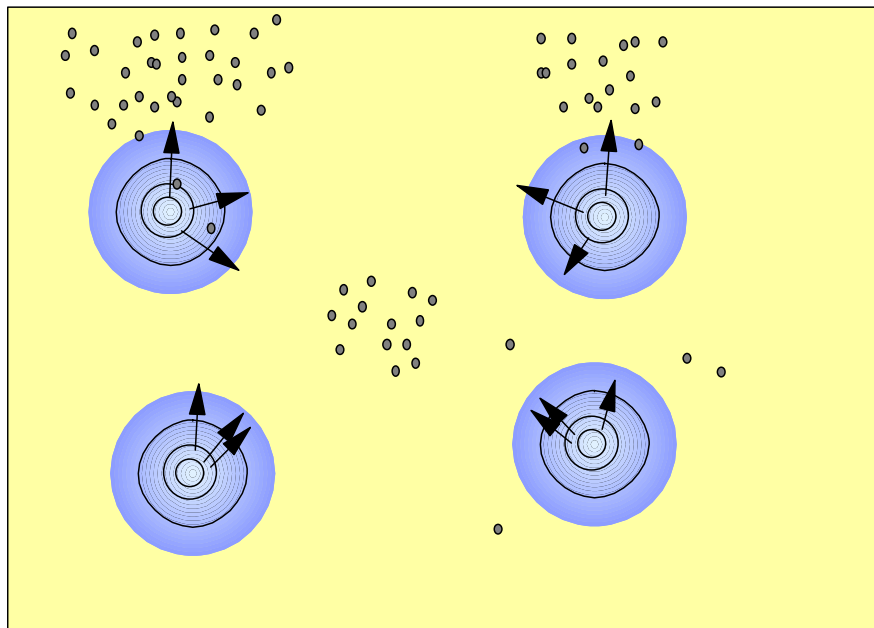


ニューラルネットワーク：教師付学習（プロパーゲーション）

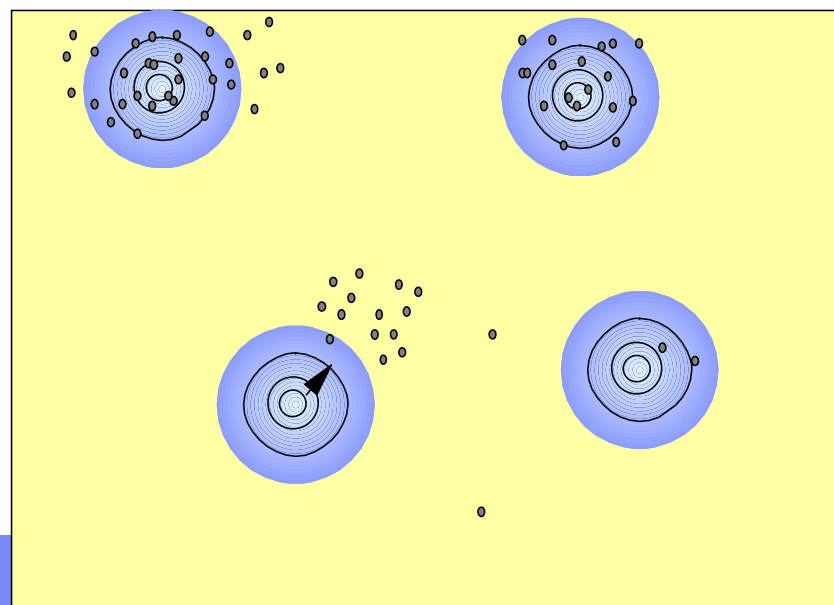


ニューラルネットワーク：教師なし学習/自己組織化（コホーネンマップ）

初期段階



- ● データ
 - ● レコード
- 後期段階

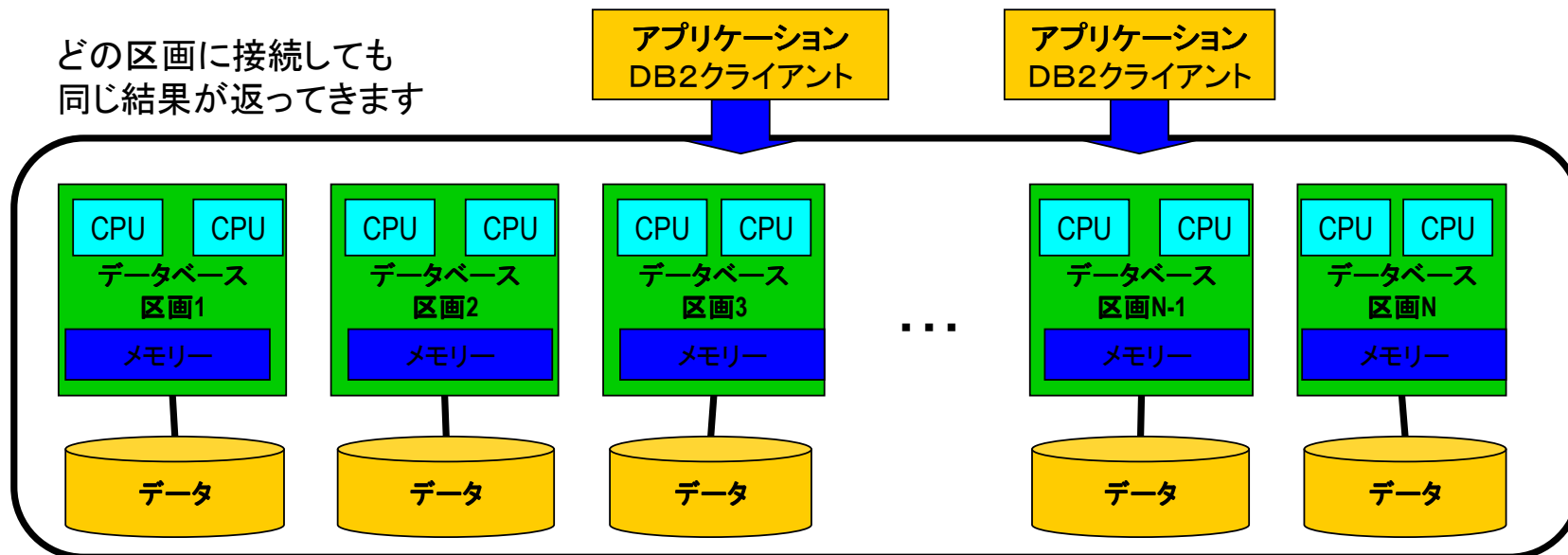


並列処理技術

Database Partitioning Feature (DPF)

- データベースの区分を持つサーバー群
 - ▶ 全体で1つのデータベース
 - ▶ パラレル・オプティマイザーと高速通信経路で接続
- ハッシングによる均等なデータ配置

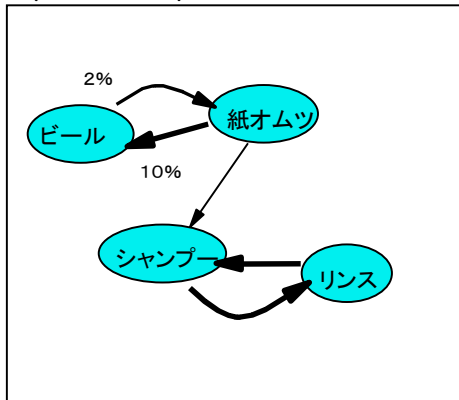
1000区分までのスケーラビリティ
透過的データアクセス



単一データベース

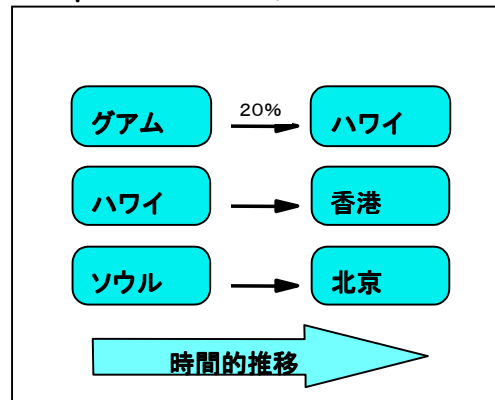
IBM データマイニングの主な機能

① 相関関係 (Association)



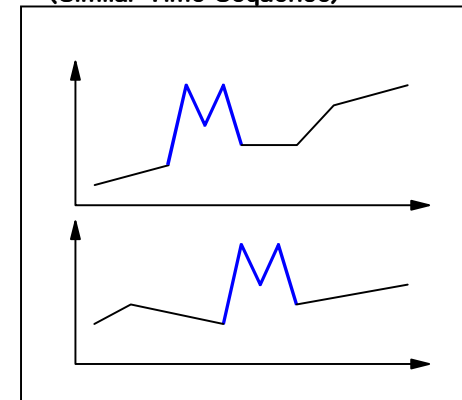
紙オムツを買う顧客の10%は同時にビールも購入する

② 時系列順パターン (Sequential Pattern)



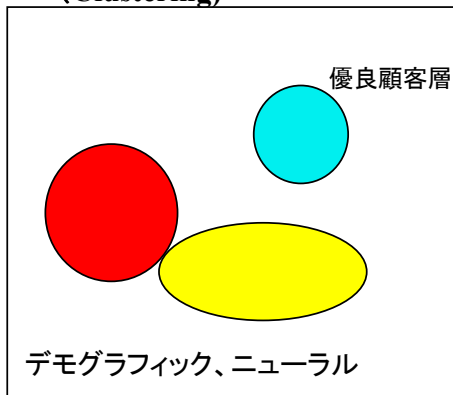
ガムに旅行した若い女性は、その次は20%の確率でハワイに行く

③ 類似時系列 (Similar Time Sequence)



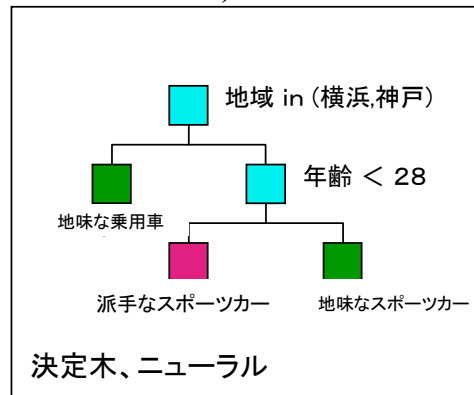
過去の商品別売上時系列データから新商品の売上パターンを予測する

④ クラスタリング (Clustering)



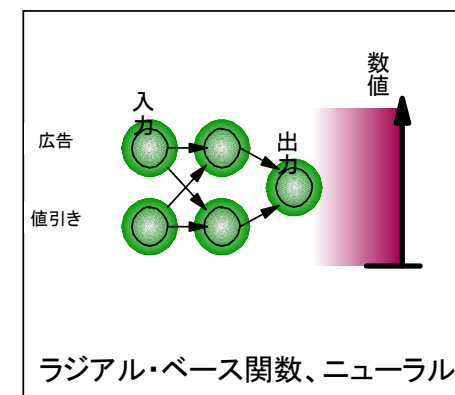
デモグラフィック、ニューラル
100万人の顧客を50のグループにセグメンテーション

⑤ クラス判別 (Classification)



決定木、ニューラル
派手なスポーツカーを買うのは郊外に住む若い人たちである

⑥ 予測 (Value Prediction)



ラジアル・ベース関数、ニューラル

広告、値引き、陳列品目をパラメータとしてインスタントコーヒーの販売量を予測する



association

1



「相関関係分析」

sequential pattern

2



「時系列パターン分析」

similar time sequence

3



「類似時系列分析」

clustering

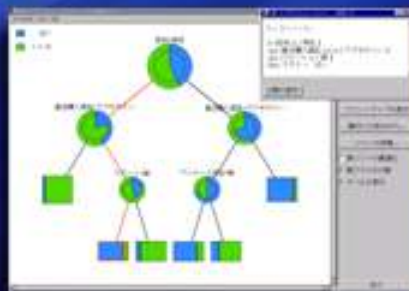
4



「クラスター分割」

classification

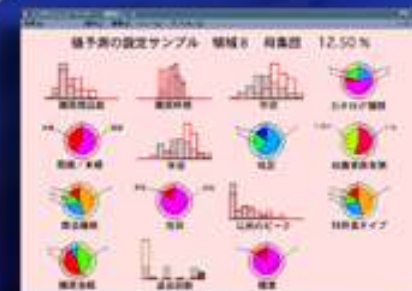
5



「クラス判別」

prediction

6



「予測」

3. 産業界でのデータマイニング事例

■データマイニングはビジネスのさまざまな分野で活用されています。

●金融業界

- ・顧客のチャンネルセグメンテーション
顧客と金融機関との取引パターン(取引種類、デリバリチャンネル...)をベースに、チャンネルに着目した顧客の新しいセグメントを見出す。
- ・倒産確率予測
過去3~5年のBS/PLから1~3年後の倒産確率を予測し、融資の評価尺度の一つとして利用。
- ・与信スコアリング
消費者金融を利用しているお客様の属性(年齢、性別、職業...)と過去の取引データから、新規のお客様に対する融資限度額を算出するモデルを構築。

●流通/小売り業界

- ・新商品の売上げ予測
過去に販売している商品の属性(形、重さ、色、素材...)と売上げデータから、新規投入商品の数週間後の売上げを算出する予測モデルを構築。原材料手配に利用。
- ・出店計画
既存店の立地条件(駐車場スペース、商圈人口、競合店状況...)と売上げデータから、新規店舗の売上げ算出モデルを構築。出店計画に反映。
- ・カタログ送付先の選定
通信販売を利用している会員のデータベースの中から「どの人にカタログをおくれば、より効果的か」というモデルを構築。

●通信業界

- ・顧客離反分析
「どんな客層がどんなタイミング(年齢、性別、利用期間、利用回数、利用金額...)で離反確率が高いか」を分析し、新サービス紹介等のつなぎ止め作戦に反映する。

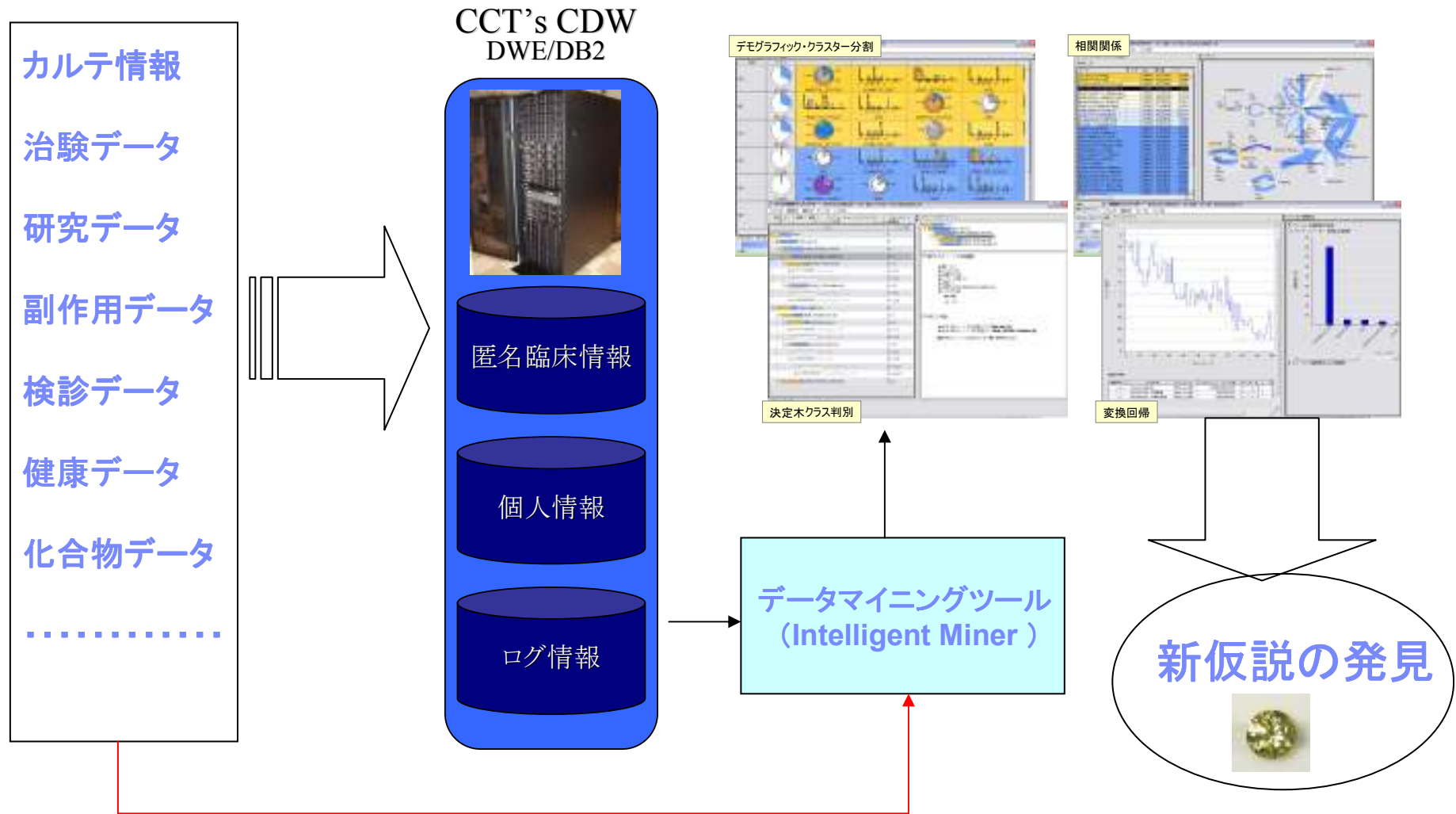
●製造業界

- ・品質管理
製造工程の各過程で計測された品質管理データと最終検査における可否との関係を分析し、プロセスの改善に役立てる。

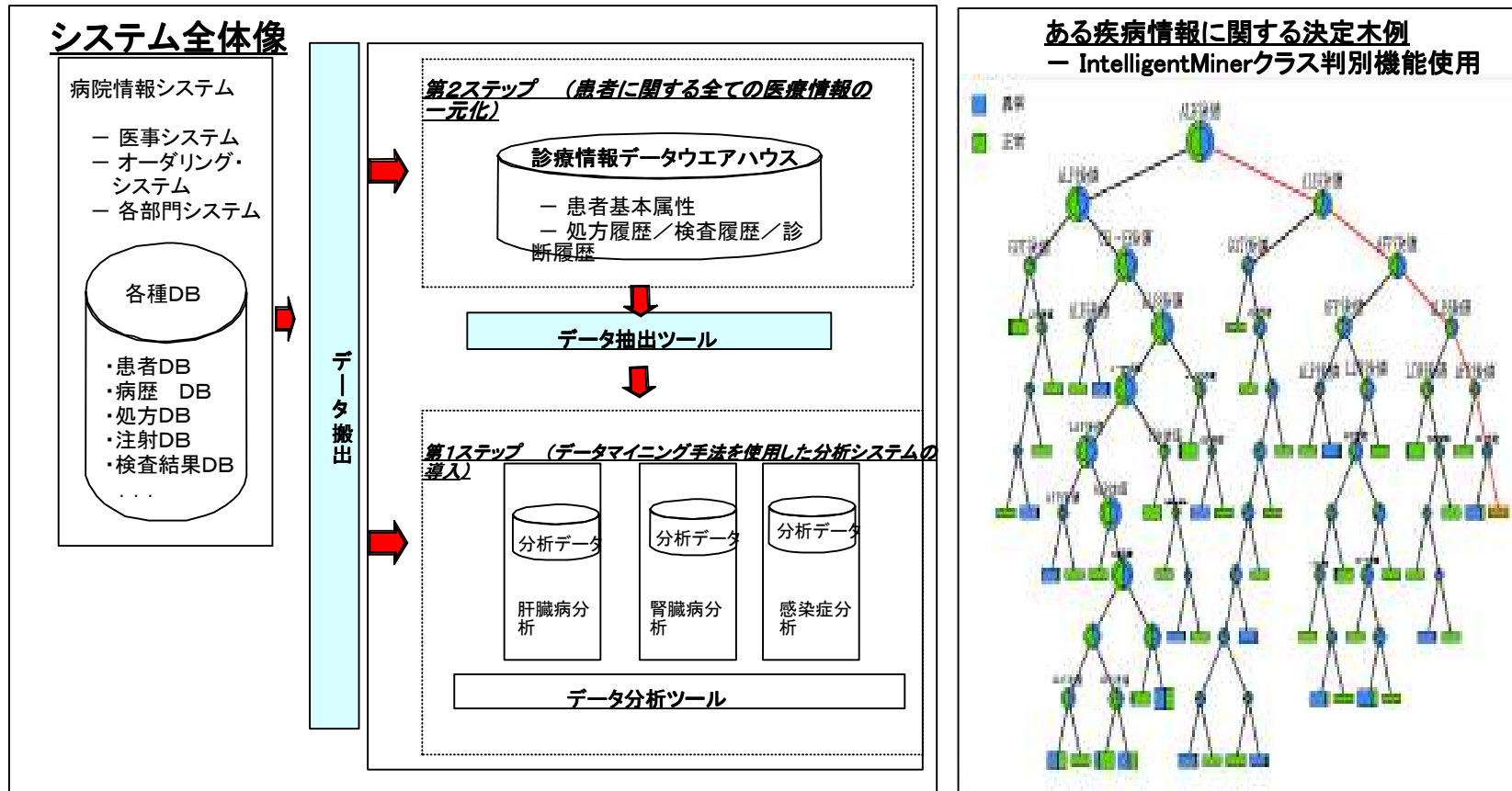
●その他業界

- ・競馬(勝ち馬)予測
JRA発表の公式データを用い、距離別に芝、ダート、障害の7モデルを開発。走破タイムを予測し、2002年1月一般の競馬ファンに情報提供開始。

4. 診療情報、治験情報などへの利用提案



病院の臨床研究支援ーデータマイニングによる疾患判別モデルの作成

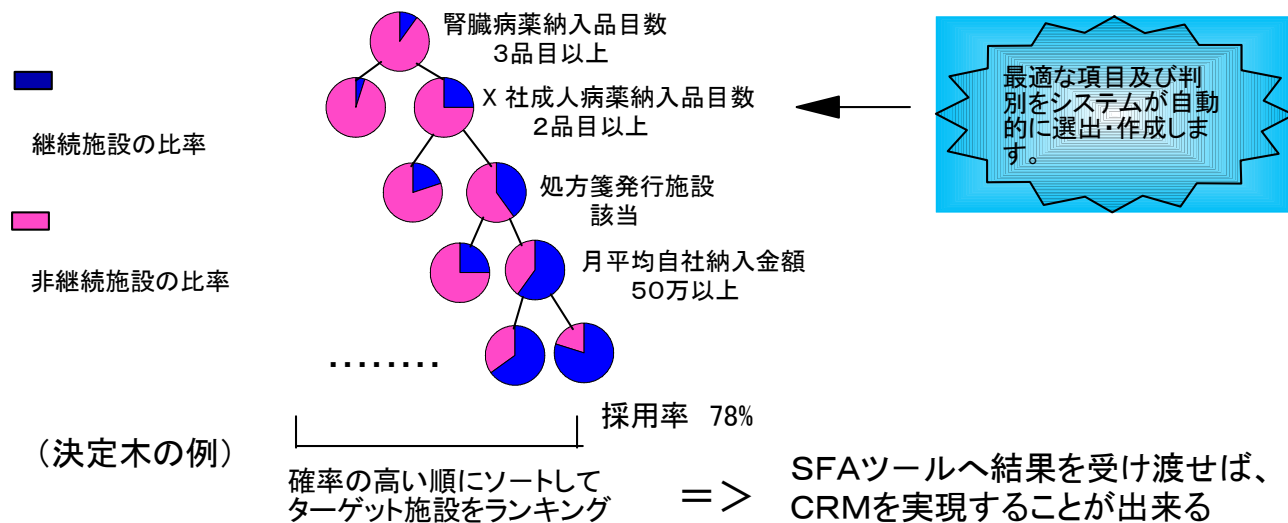


どのような患者のどの治療が効果的か過去の事例から、事実を浮かび上がらせます。

製薬業の営業支援例

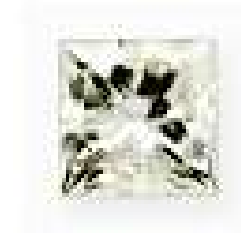
目的： 新薬発売後、どういう施設が他剤から新薬への乗換えを行ったかの分析を行う

- 手法：
- ・新薬発売前と発売後の一定期間の各施設の薬剤納入額を算出
 - ・新薬への乗換え、継続(非乗換え)に関する施設要因(施設属性、医師属性等)を洗い出す



その他診療情報を使った仮説発見の可能性。。。。。

- XX患者の平均余命予測モデル
- XX薬の服薬期間と効果との相関分析
- 体質別投薬効果分析
- 新薬と従来薬との効果差別分析
- 遺伝子情報のクラスタリングによる治療品質の向上



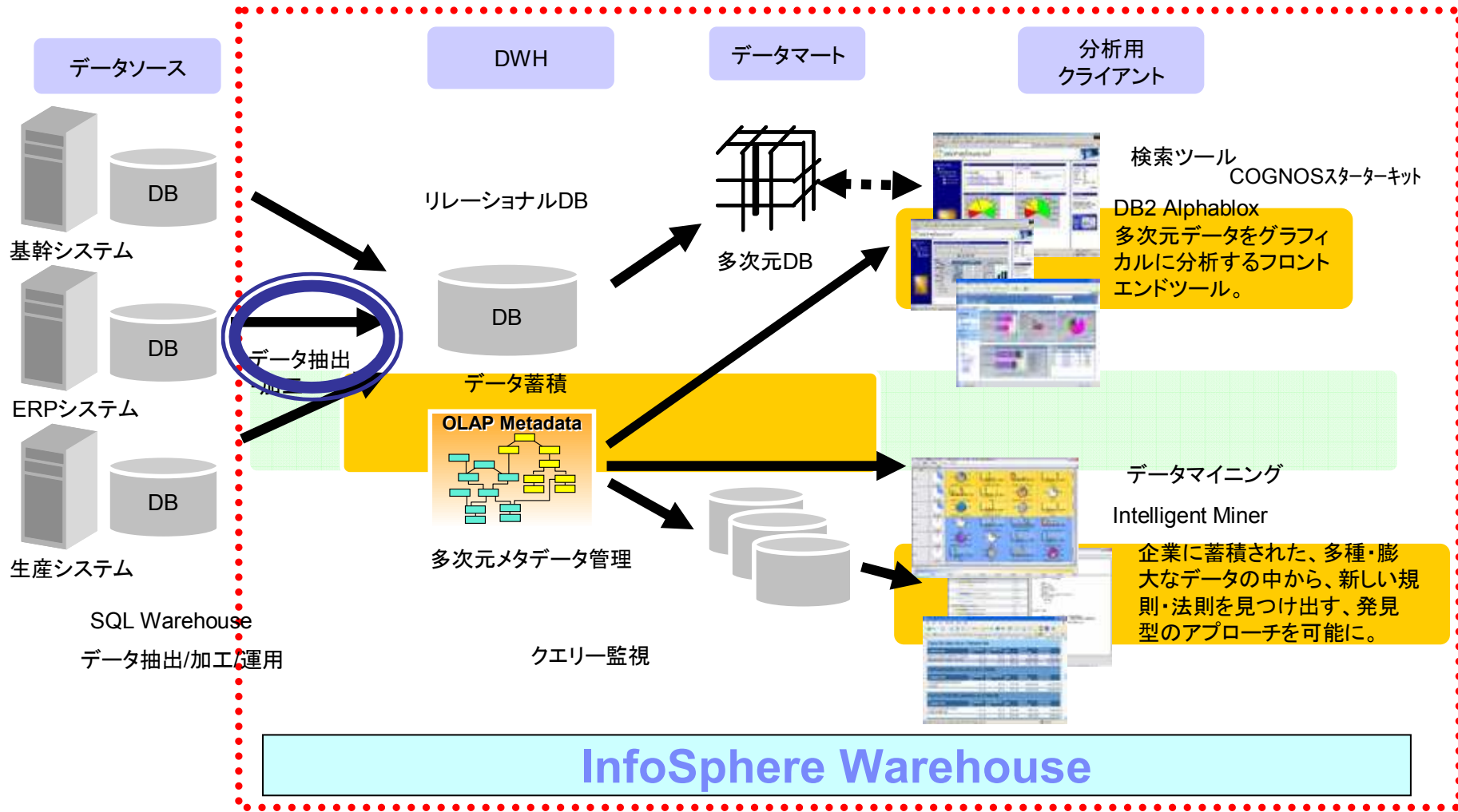
..... 等



従来の統計で実現できなかった仮説発見の可能性がります

5. データマイニングを実現するIBMソリューション

InfoSphere Warehouse (BIとDWHの機能をオールインワンで提供)





- ニュース
- プレスリリース
- イベント / セミナー
- キャンペーン

プレスリリース

渋谷にBIソリューションセンターを開設

2008年7月7日

渋谷にBIソリューションセンターを開設

—意思決定に役立つ情報活用をデモから技術支援までトータルにサポート—

日本IBM(社長兼会長:大歳卓麻、NYSE:IBM)は、ビジネス・インテリジェンス(BIソリューション)の導入を検討されるお客様、およびIBM製品を活用してBIソリューションを販売・導入されるパートナーを対象に、デモから技術支援まで一貫したサービスを提供する施設「IBM BIソリューションセンター」を、本日、日本IBM新渋谷事業所内(東京都渋谷区)に開設しました。

BIとは、企業内に散在するデータを蓄積、加工して意思決定に活用するための手法です。本センターは、BI導入にあたってビジネス効果やシステム環境を相談したいというお客様やパートナーの要望にお応えし、IBMのBIソリューションに関する総合的な支援を提供します。2008年1月に統合されたCognos製品をはじめ、IBM® DB2®関連製品、企業内に蓄積されたデータを抽出・加工・変換しデータウェアハウスに格納するInformation Server関連製品、データマイニング関連製品を支援対象とし、今後対象製品を順次拡大する予定です。BIに精通した専任のソフトウェア、ハードウェアの技術者と営業担当者が連携して、お客様やパートナーを支援していきます。

本センターでは、BIソリューション環境を構築するための技術支援として、最適なソフトウェア構成の提案や推奨ハードウェア構成の規模測定を行います。問い合わせを受けた際に、一ヶ月程度かかる場合があったシステム構成の最適化提案も、独自に用意した分析シートに沿った質問項目に答えていただくことで、3営業日での回答が可能です。

また、多様な業務や業界におけるデモを通じて、BIソリューションの効果を理解いただくことができます。例えば、「営業パフォーマンス・オペレーション改善支援」では、支店および営業担当者毎の売上、シェア、案件情報を可視化したり、業績が優れた社員の行動特性を共有することで、企業全体の営業力向上に役立てることができます。また業界に特化したデモも幅広く展開する予定で、製造業界向け「ワランティコスト削減ソリューション」では、製品保証にかかわるクレームの発生状況を分析し、将来予測に基づいたコストの引当と各ディーラーからの請求額の妥当性を検証、クレーム処理にかかるコストの削減に貢献します。

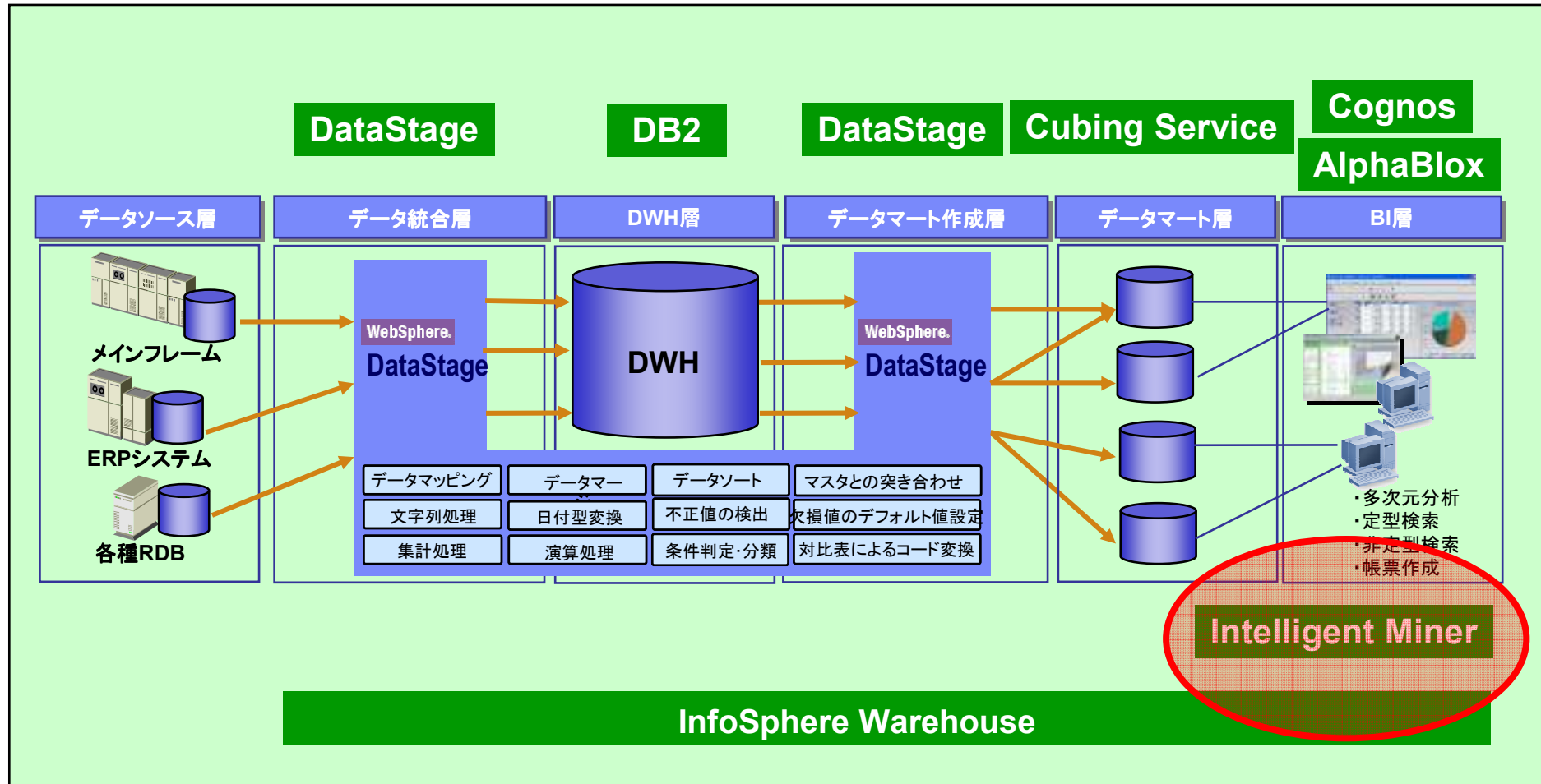
本センターの概要は、以下のとおりです。

ご支援内容

ページ・オプション

ページを印刷する

IBM BI ソリューションセンター支援範囲 (本年7月より実働開始)



本日はお忙しい中、ご静聴ありがとうございました。



IBMブース内でデモを実施していますので御時間ありましたら、お立ち寄りください。



IBM データマイニングソリューションに興味をもっただければ幸いです。ありがとうございました。

