





SYBASE®

数理システムユーザコンファレンス2006講演資料



製品品質向上と在庫削減の ためのデータ及びツールの 活用法 - Sybase IQのご紹介 -

2006年11月22日
サイベース株式会社

サイベース社のご紹介



サイベース社はデータセンターから現場のアクションポイントまでの広範囲にわたる情報管理ソフトウェアを提供するソフトベンダーとしては世界最大規模です。

- 経営理念: 高い誠実性と堅実性
 - 2005年収入8.2億ドル
 - 2005年ライセンス売り上げ2.9億ドル
 - 2005年キャッシュバランス8.7億ドル
 - 2005年業務キャッシュフロー1.7億ドル
- グローバル企業
 - 本社米国カリフォルニア州ダブリン市
 - 営業拠点: 60カ国
 - 開発拠点: 米国、カナダ、中国、インドなど
 - 従業員数 3,800人
 - 企業ユーザー 12,000社以上
- ソフトウェア業界のリーダー
 - 20年の経験に基づく革新的なテクノロジーの提供
 - オープンかつ安全なソフトウェアソリューションの提供

データ管理における要件と課題 (1)

大容量履歴データベース、統合データベースを構築するにあたって

要件:

- 多量な属性を持つ、大量データの管理
- 上記データからの必要な情報の有効な応答時間での取り出し
- 上記が現実的なコストで実現可能

課題:

- 大量データの取り込み、蓄積及び統合(装置, 製品, 進捗, 品質)
- 傾向分析のため長期間のデータ保管
 - ローディングのスピード
 - ディスク容量
 - コスト
- 大量データからの検索
 - 現実の使用に耐えうるパフォーマンス

データ管理における要件と課題 (2)

大容量履歴データベース、統合データベースを構築するにあたって

要件:

- 多量な属性を持つ大量のデータが格納可能であること
- 上記データ検索が有効時間内に終了すること
- 非定形な検索の性能が予測可能かつ実用可能であること
- 標準インターフェースが利用可能であること

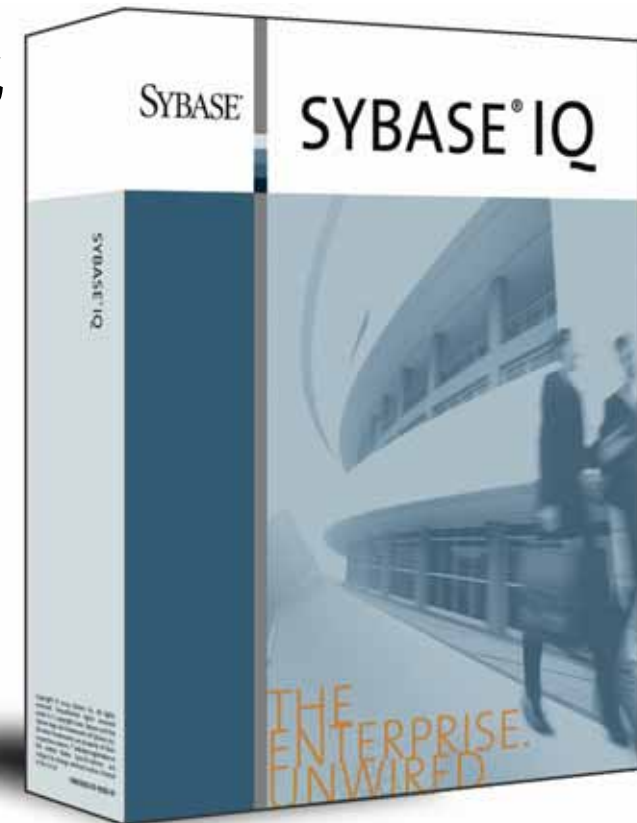
課題:

- 汎用RDBMS仕様からくる制約
 - 表当たりには持てる属性(カラム)数の制約
 - 非定形検索のチューニングが難しい
 - 正規化の問題
 - 結果的にアプリケーション開発、システム運用への負荷

Sybase IQとは？

Sybase IQ

- 多種類, 大量データの登録, 検索に最適化された超高速クエリーエンジン
 - 慣れ親しんだSQLインターフェース
 - 独自のデータ格納形式 & アクセス手法
 - 従来型RDBMSの10 ~ 100倍のスピード
 - データ圧縮
 - 従来型RDBMSの5倍の格納効率
 - 均一なパフォーマンス, データ構造非依存
 - 自由度の高いテーブル設計



Sybase IQの適用分野例

POSデータ分析

会計帳票等, 各種レポーティング処理

アクセスログ, セキュリティログの検索・分析

企業統治の為の取引履歴管理

製造工場における製品・品質履歴情報管理

SYBASE®

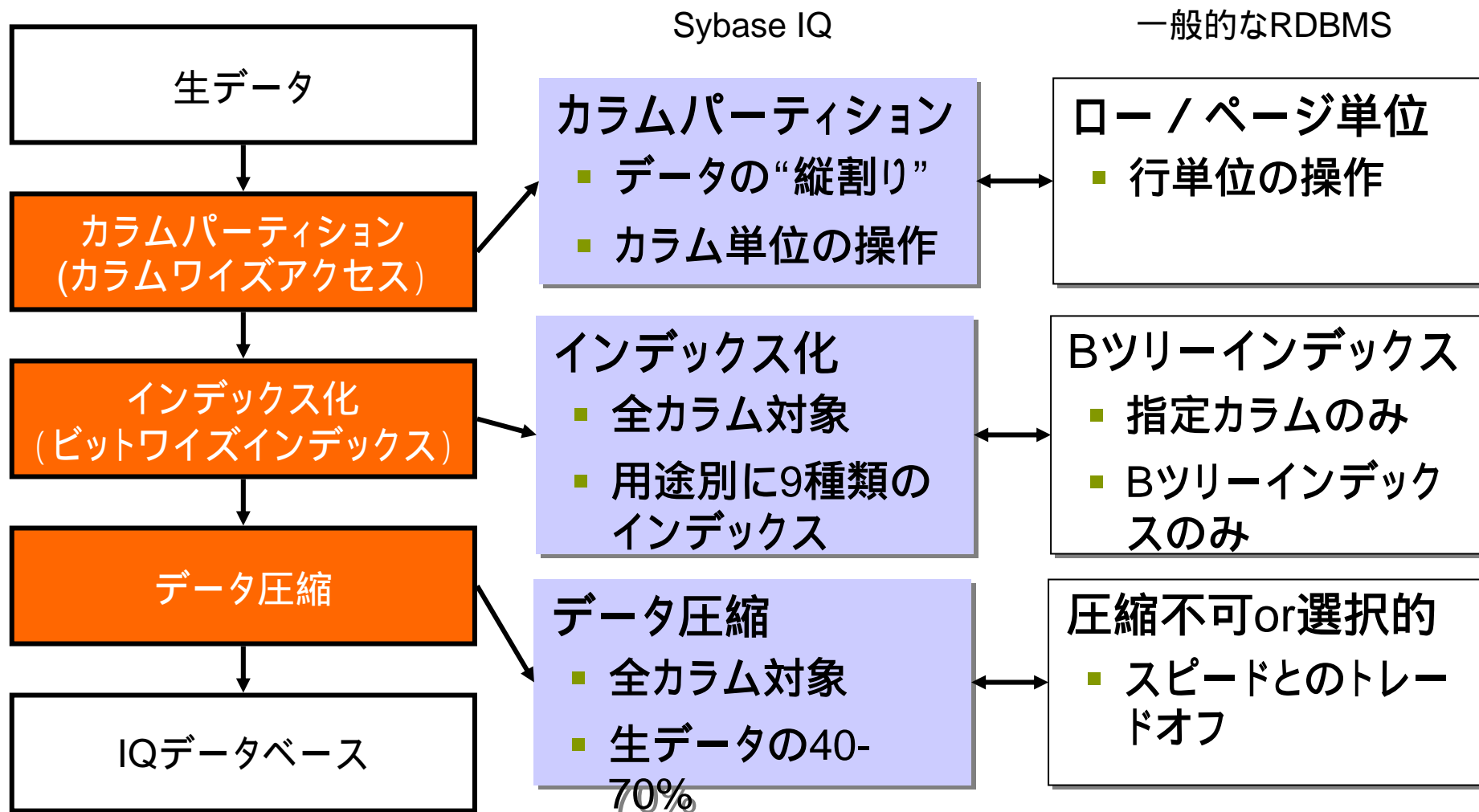


Sybase IQの特徴

SYBASE®



IQ固有のデータ構造



一般的RDBMS

データ格納、アクセス方式

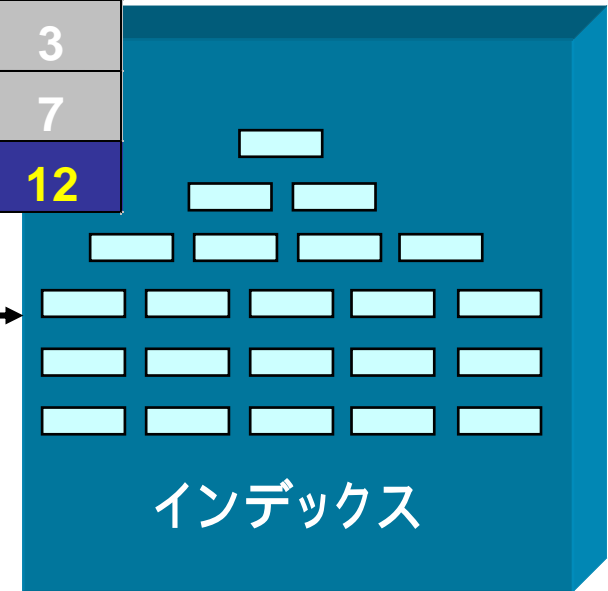


リレーショナルテーブル

日付	店舗	州	クラス	数量
3/1/96	32	NY	A	6
3/1/96	36	MA	A	9
3/1/96	38	NY	B	5
3/1/96	41	CT	A	11
3/1/96	43	NY	A	9
3/1/96	46	RI	B	3
3/1/96	47	CT	B	7
3/1/96	49	NY	A	12

- インデックスとデータは別に作成
- B-ツリー型のインデックスが一般的

指定カラムのみ
インデックス作成



SybaseIQ カラムパーティション

データ格納、アクセス方式

各カラムは独立して
保存・管理される

ローIDによりカラム間
の関係は管理される

日付	店	州	クラス	数量	カテゴリ
1998/3/1	32	NY	A	6	Gen
1998/3/1	35	CT	A	9	Spec
1998/3/1	36	CT	C	11	Gen
1998/3/1	39	SD	D	8	Gen
1998/3/1	42	KY	A	5	Spec
1998/3/1	43	VT	C	14	Spec
1998/3/1	47	GA	A	31	Gen
1998/3/1	51	MD	A	4	Sub
1998/3/1	55	DC	D	16	Gen

入出力の単位

メリット:

- 必要なカラムだけのアクセスで、ディスクI/Oを平均90%低減
- 同じデータ型のデータだけで構成される為、圧縮の効果が高い
- クエリの処理スピードが10-100倍高速

カラムパーティション 効果

Q: カリフォルニアで保険に加入していない, 男性の人数は?

一般的なRDBMS

1000 万件

性別	州	加入
M	NY	Y
M	CA	Y
F	CT	N
M	MA	Y
M	CA	N
-	-	-

800 バイト/行

$$\frac{800 \text{ バイト} \times 1000 \text{ 万件}}{16 \text{ KB Page}} = \mathbf{500,000 \text{ I/Os}}$$

- 大量の不要データ処理が発生
- テーブルスキャンが発生するケースが多い

Sybase IQ

1000 万件

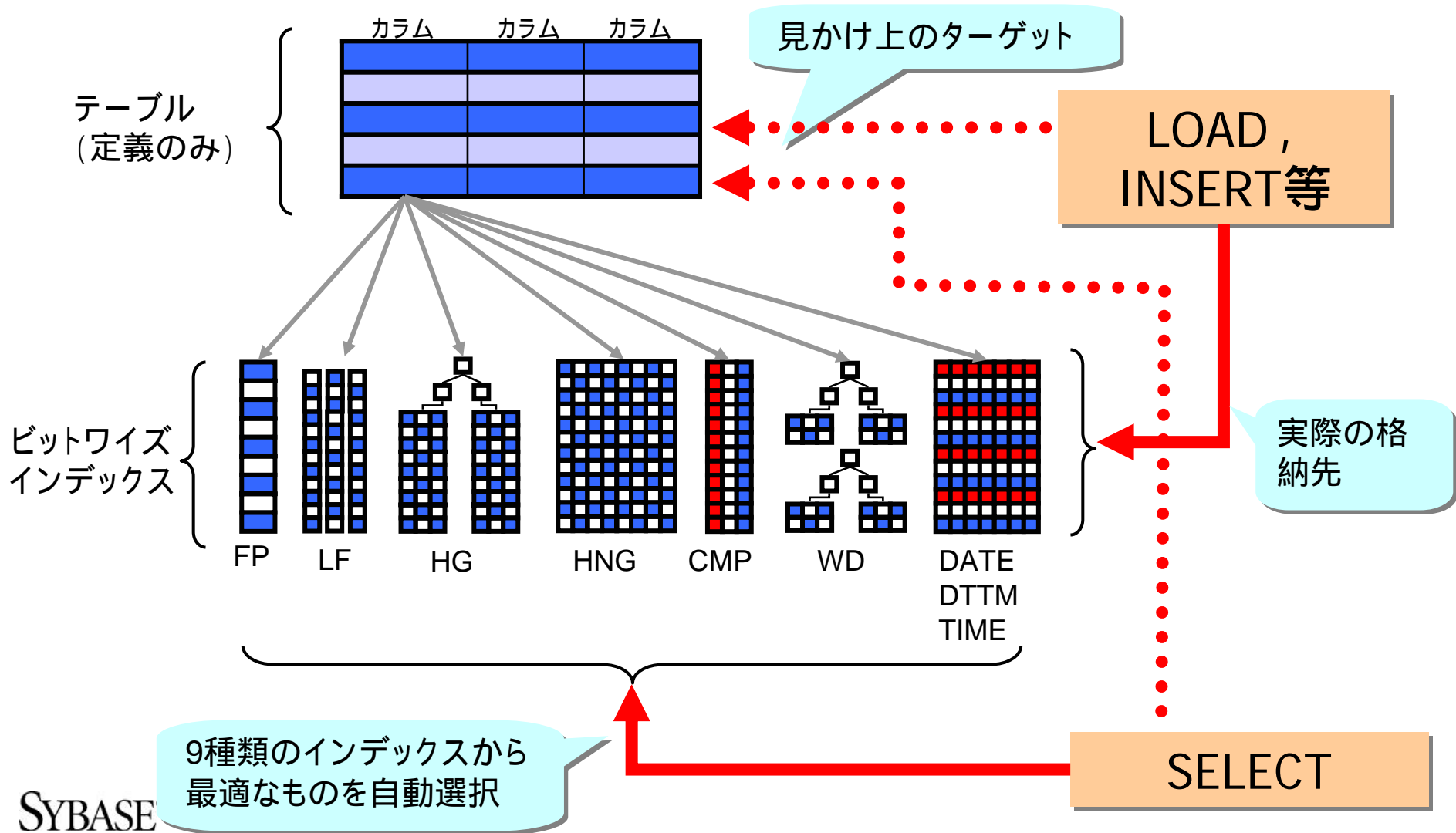
性別	加入	州
M	Y	CA
M	N	CA
F	Y	NY
M	N	CA

1バイト 1バイト 2バイト

$$\frac{(1+1+2) \text{ バイト} \times 1000 \text{ 万件}}{16 \text{ KB Page}} = \mathbf{2,500 \text{ I/Os}}$$

- 圧縮により, 実際のI/O容量は更に減少
- 構造上, テーブルスキャンが有り得ない

ビットワイズインデックス



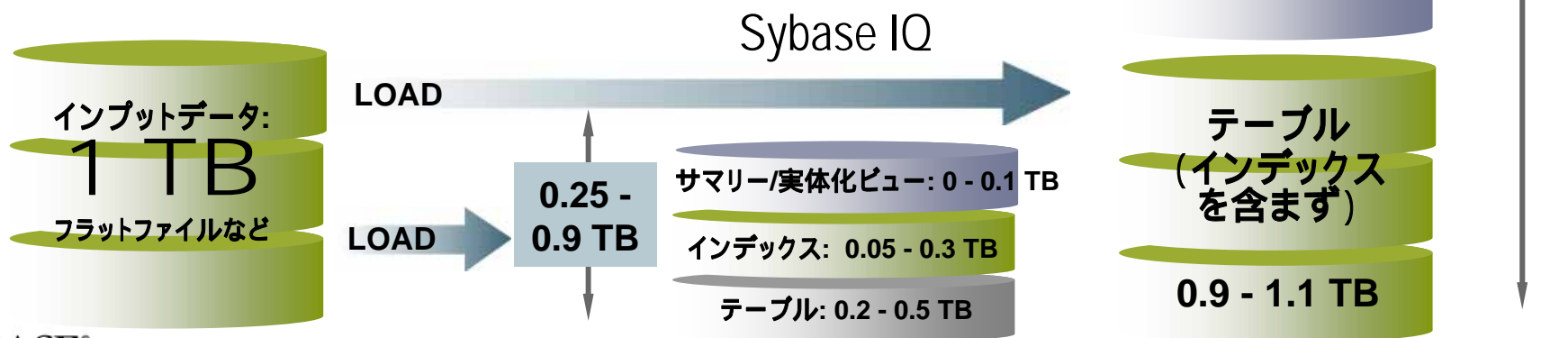
Sybase IQ インデックスの種類

インデックス名称	略記	用途
Fast Projection	FP	射影, 集合関数を伴う射影, 文字列検索等。 デフォルトで生成
Low Fast	LF	カーディナリティ1,500以下。 結合, group by, count(), min(), max(), <>, IN リスト, 範囲検索など
High Group	HG	カーディナリティ1,500超。LFに同じ
High Nongroup	HNG	範囲検索, 集合関数を伴う条件句
Compare	CMP	2つのカラム間の大小比較
Word	WD	文字列検索(like)
Date	DATE	date型データの範囲検索
Datetime	DTTM	datetime型データの範囲検索
Time	TIME	time型データの範囲検索

Sybase IQ データ圧縮

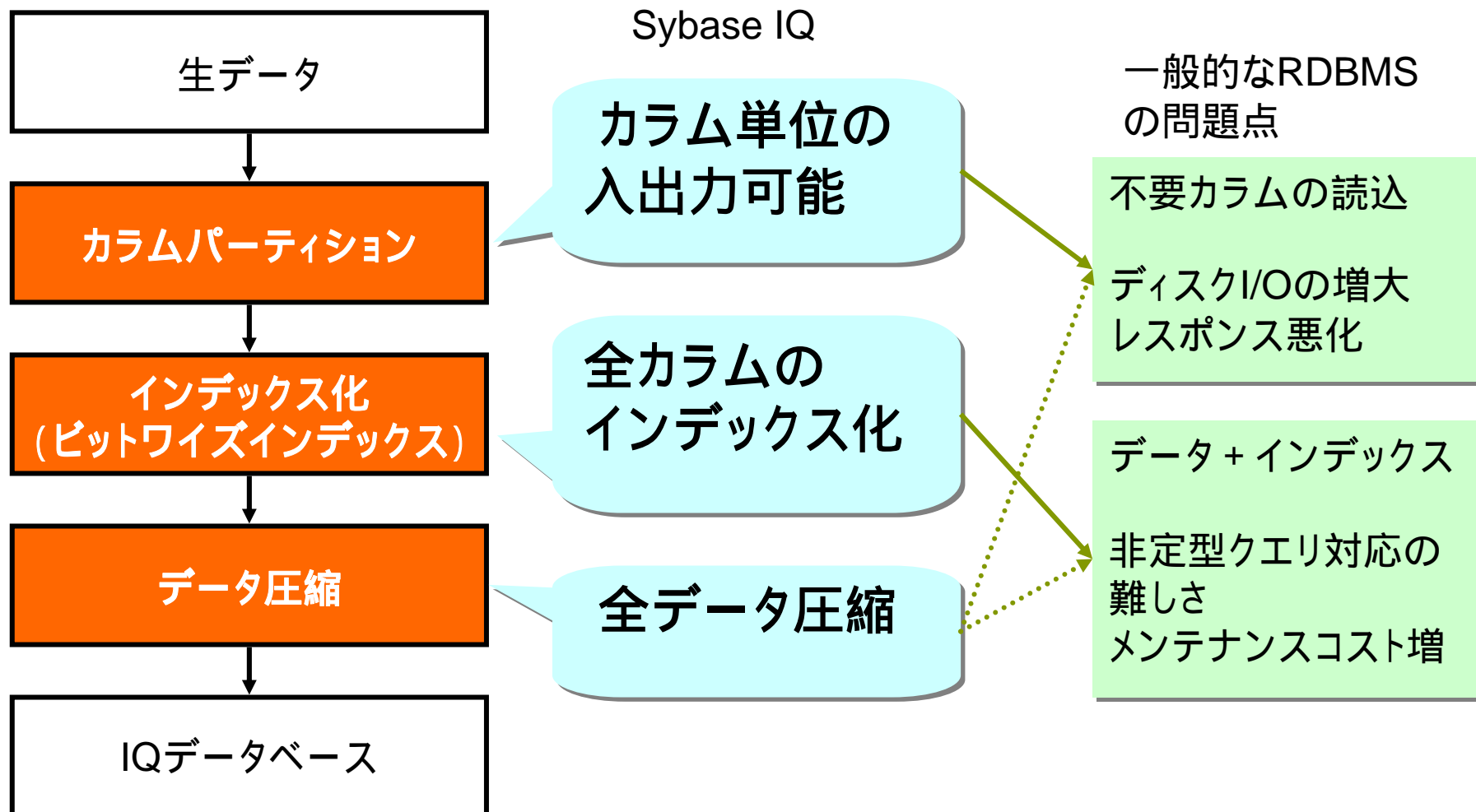
データ圧縮

- 全ビットワイズ化データを圧縮し格納
- ストレージ消費量
 - 生データ比 約50%
 - 従来のRDBMS比 約20%
 - 時系列データの場合, 従来型RDBMSの5倍の期間分の保存が可能
 - ロード, 検索時のディスクアクセス量が 1/100~1/1000

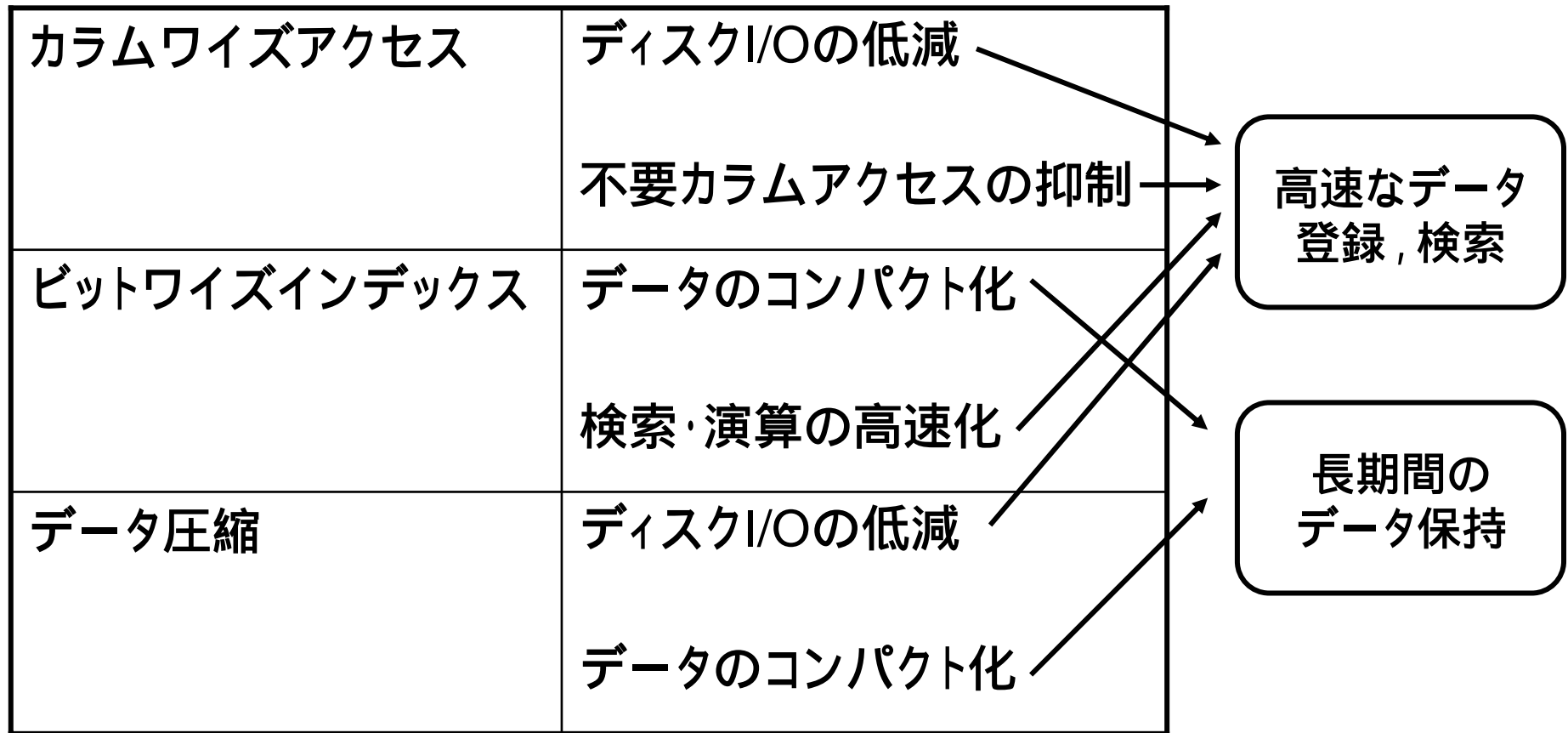


Sybase IQ

多数カラム, 大量データの管理に適した理由



まとめ



Sybase IQは、多様で大量なデータから、必要な情報を取得するための最適な基盤を提供します