

テラバイト級の大容量データでも 高速検索できるXMLデータベース **TX1**[®]



2006年10月4日

東芝ソリューション株式会社
TOSHIBA SOLUTIONS CORPORATION
プラットフォームソリューション事業部

Agenda



1. **TX1**の特長
2. **TX1**の開発環境
3. まとめ

1. **TX1**の特長
2. **TX1**の開発環境
3. まとめ

XMLデータベース **TX1** とは

東芝ソリューションが開発した「テラバイト級の
大容量データでも高速検索できるXMLデータベース」



非定型データのデータベース化



大量に存在する多種多様の非定型データを的確に管理・活用するデータベースに求められることは？

柔軟なデータ構造が採れる

XMLの柔軟性・拡張性をそのまま活かすことができる

XMLデータベース **TX1** は、
非定型データの管理に適したデータベース

大容量データを管理できる

エンタープライズレベルの
大容量でも高速に検索できる

信頼性を確保できる

企業システムとしての
信頼性を確保できる

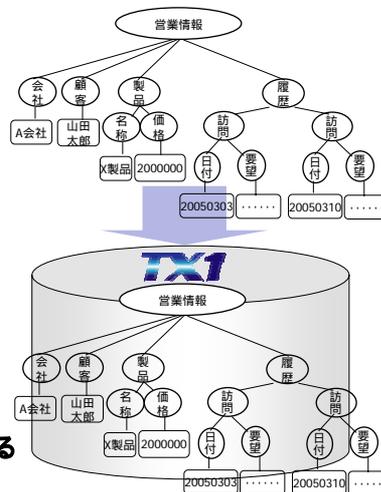
柔軟なデータ構造 **TX1** のデータ管理



- XMLの階層構造をそのまま格納でき、スキーマの定義も不要
- データ構造に変更があっても、更新データをそのまま格納するだけで、DB構造の変更は不要
- 複雑な構造のデータもそのまま格納可能



データ構造の変更があっても柔軟に対応できる
XMLの柔軟性・拡張性をそのまま活かせる



階層構造をそのまま格納可能

大容量データでも高速に検索 TX1の高速検索技術



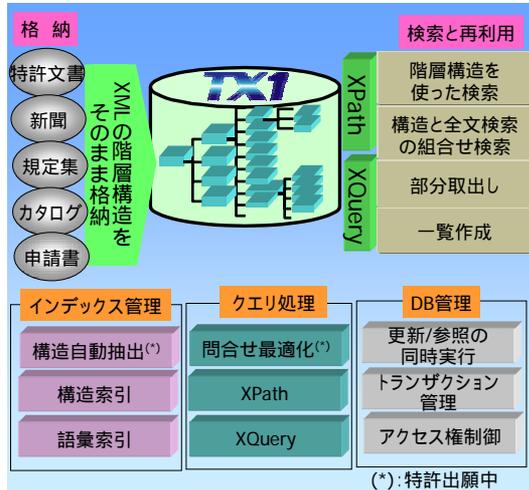
独自に開発した技術で、高速検索を実現

■ 大容量データでも高速検索

- ・独自に開発した**構造自動抽出技術(*)**と**問合せ最適化技術(*)**で、大容量データでも高速に検索
- [構造自動抽出技術]
階層構造の照合を高速化するための索引化技術
- [問合せ最適化技術]
複雑な問合せでも、最適な問合せプランを生成する技術

■ 高いスケーラビリティ

- ・マルチスレッド対応により、多人数で同時に検索しても、検索性能が低下しない高いスケーラビリティ



大容量データでも高速に検索 TX1の高速検索技術



XMLデータを検索する際には、階層構造をたどったり、様々な条件を満足する部分階層を抽出したりするために、“階層パターン照合”が必要

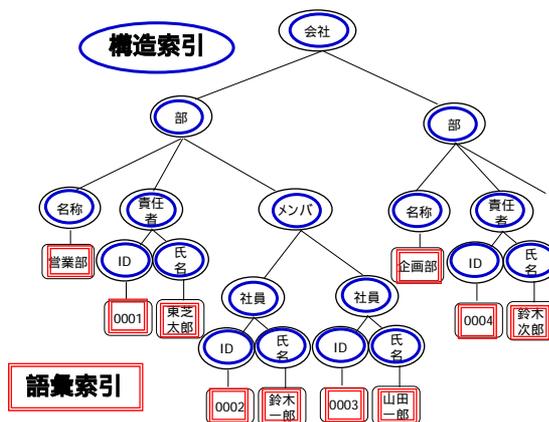
1. 部の責任者の氏名が「東芝太郎」の部の名称は？
営業部 東芝太郎
 2. 氏名に「鈴木」が含まれる人が属している部の名称は？
営業部 鈴木一郎
企画部 鈴木次郎
- < 高速化するために >

■ 構造自動抽出と語彙索引

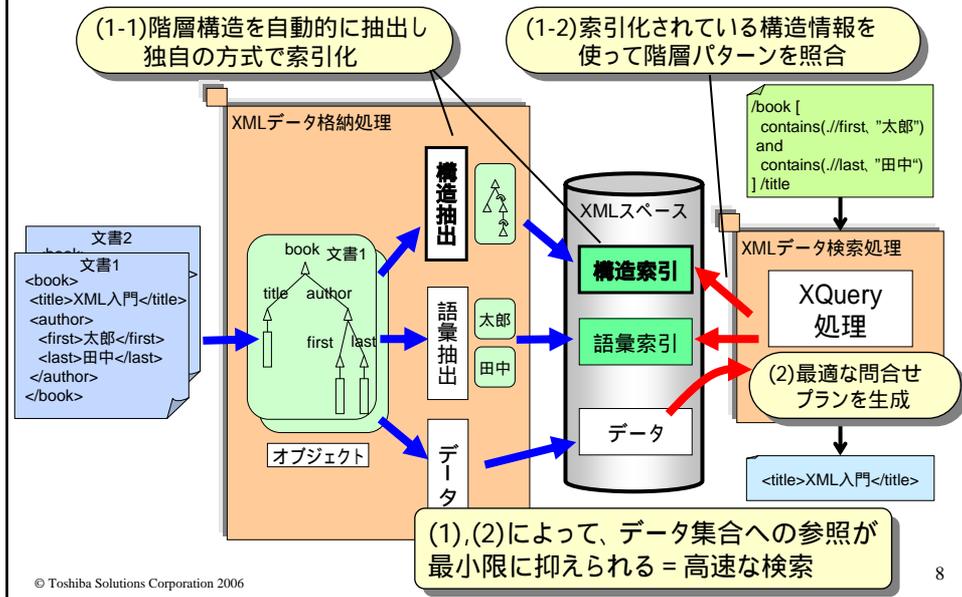
- ・大容量データでも、“階層パターン照合”を高速に行うため、独自の索引方式で索引化(構造索引)
- ・要素内容の語彙を全文検索用の索引として索引化(語彙索引)

■ 問合せ最適化

- ・抽出した構造情報と語彙情報を統計的に分析して、最適な問合せプランを生成



大容量データでも高速に検索 構造索引の生成とクエリによる利用



8

大容量データでも高速に検索 TX1の検索性能



検索性能

- 特許公報8年分300万件(100GB)から、1万件の大量ヒットでも、検索時間は約1秒

測定環境

- PCサーバ: Xeon 3GHz × 2、メモリ 2GB
- ディスクアレイ装置: ArrayFort
- OS: Windows Server 2000

特長

- 特許公報のような複雑な構造のXMLデータもそのまま登録して、高速に検索可能
- テキストデータの部分一致検索でも高速に検索
- 大量ヒットでも件数の取得が高速
- 構造とキーワードを組み合わせた検索でも高速に検索



データベースに求められる信頼性を確保するためのデータベース管理機能をサポート

管理機能

更新/参照の同時実行	トランザクション管理	リカバリ機能	アクセス権制御
------------	------------	--------	---------

■ MVCC方式 (Multi-Version Concurrency Control: 多版型同時実行制御方式) による同時実行制御

各トランザクションがデータベースに対して問合せるときに、別のトランザクションが更新中の最新データでなく、更新前のデータを参照して処理する方式

■ アクセス権制御

XML文書を最小単位としてXPath形式で指定するパスごとに権限の設定が可能

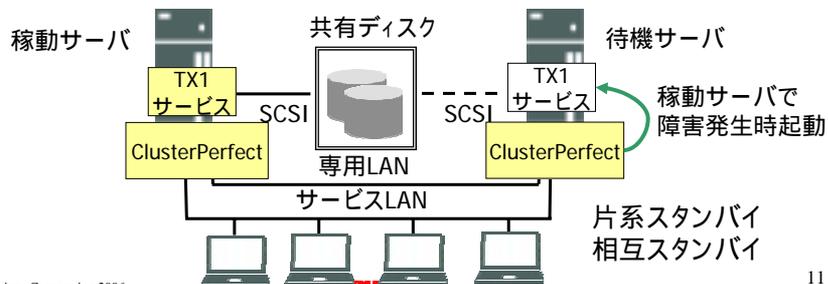


運用機能

オンラインバックアップ	データベースのインポート/エクスポート
-------------	---------------------

可用性システムの構築

当社ClusterPerfectなどのクラスタソフトウェアとの連携



TX1 の特長 (まとめ)



TX1 は、高速性と信頼性の両立したXMLデータベース

柔軟なデータ構造が採れる

XMLの柔軟性・拡張性を
そのまま活かすことができる

大容量データを管理できる

エンタープライズレベルの
大容量でも高速に検索できる

信頼性を確保できる

企業システムとしての
信頼性を確保できる

1. **TX1** の特長

2. **TX1** の開発環境

3. まとめ



標準問合せ言語XQueryで、 効率的なアプリケーション開発が可能

- 標準問合せ言語XQueryに対応
XMLデータの取出しに適した標準
問合せ言語で、効率的なアプリ
ケーション開発を行うことが可能
- JDBC、ODBCをベースにしたAPI
をサポート

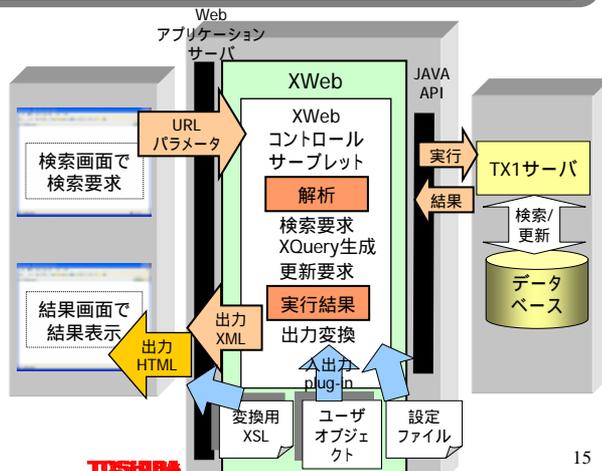
```
FOR $sale IN document("販売.xml")//販売情報
LET $a:= $sale/明細/商品
WHERE contains($sale/日付,"200503")
RETURN
<sale> {
  $sale/顧客/氏名,
  <商品数>{count($a)}</商品数>
}</sale>
```

XQueryの例:
「2005年03月」に販売した顧客名と
商品数の一覧を表示

Webアプリケーション開発支援機能 XWeb

- URLパラメータを用いてTX1にアクセスできるサブレットを提供
- XWebを利用することで、HTMLページやスタイルシートの作成と簡単な設定
だけで、TX1にアクセスするWebアプリケーションを開発可能

- 基本アクセス機能
- クエリ作成支援機能
- 出力結果のHTML変換
支援機能
- ユーザオブジェクトの
plug-in機能



TX1の動作環境



■ 稼働環境

プラットフォーム	OS	CPU	メモリ容量
Windows	Windows 2000 Server(SP4) Windows Server 2003	Pentium 800MHz(最小) Xeon 3.0GHz, 2CPU(推奨)	256MB(最小) 2GB(推奨) (データ量、同時実行数に依存)
Solaris	Solaris9 12/02以降(64bit) Solaris10(64bit)	UltraSparc 600MHz(最小) UltraSparc 1.3GHz, 2CPU(推奨)	512MB(最小) 4GB(推奨) (データ量、同時実行数に依存)

■ 開発環境

開発言語	OS	SDK	XMLパーサ
JAVA	Windows 2000 Server (SP4) Windows Server 2003 Windows 2000 Professional (SP4) Windows XP Professional (SP1)	JDK/JRE 1.4.2 (Sun Microsystems)	JAXP 1.1 (Sun Microsystems:JDK/JREに 含まれる)
	Solaris9 12/02以降(64bit) Solaris10(64bit)	JDK/JRE 1.4.2 JDK/JRE 5.0 (Sun Microsystems)	JAXP 1.1 JAXP 1.3 (Sun Microsystems:JDK/JREに 含まれる)
C++	Windows 2000 Server (SP4) Windows Server 2003 Windows 2000 Professional (SP4) Windows XP Professional (SP1)	Microsoft Visual C++ 6.0 (SP6) Microsoft Visual C++ .NET 2003	MSXML 3.0 MSXML 4.0
Visual Basic	Windows 2000 Server (SP4) Windows Server 2003 Windows 2000 Professional (SP4) Windows XP Professional (SP1)	Microsoft Visual Basic 6.0 (SP6) Microsoft Visual Basic .NET 2003	MSXML 3.0 MSXML 4.0

1. TX1の特長
2. TX1の開発環境
3. まとめ



まとめ



1. TX1は、非定型データの管理を目指して、設計・開発されたデータベースです。
2. 非定型データを管理するデータベースに求められることは、
 - (1)柔軟なデータ構造が採れること
 - (2)大容量のデータを管理できること
 - (3)信頼性を確保できること
3. TX1は、上記を満たす「非定型データの管理に適したデータベース」です。



- 拡張性の高いシステムを構築するには、XML技術を利用したシステムを検討することが有効です。
- 当社は、XMLデータベース TX1を核として、拡張性や利便性の高いソリューションをご提供致します。

ご清聴ありがとうございました

お問い合わせは：
東芝ソリューション株式会社
プラットフォームソリューション事業部
e-mail: xml@toshiba-sol.co.jp

商品情報ホームページ
<http://xml.toshiba-sol.co.jp/>



TOSHIBA



東芝ソリューション株式会社
TOSHIBA SOLUTIONS CORPORATION

プラットフォームソリューション事業部

Intel, Pentium, Xeonは、アメリカ合衆国およびその他の国におけるインテルコーポレーションまたはその子会社の商標または登録商標です。
Sun, Sun Microsystems, Java, Solaris は、米国 Sun Microsystems, Inc. の米国およびその他の国における商標または登録商標です。
Microsoft, Windows, Windows Serverは、米国Microsoft Corporationの米国およびその他の国における登録商標です。
本資料に掲載の商品の名称は、それぞれ各社が商標として使用している場合があります。
本内容は予告なく変更する場合があります。詳細は、仕様書あるいは説明書をご覧ください。

© 2006 Toshiba Solutions Corporation