ORACLE®

ORACLE®

ORACLE THIRTY YEARS 30

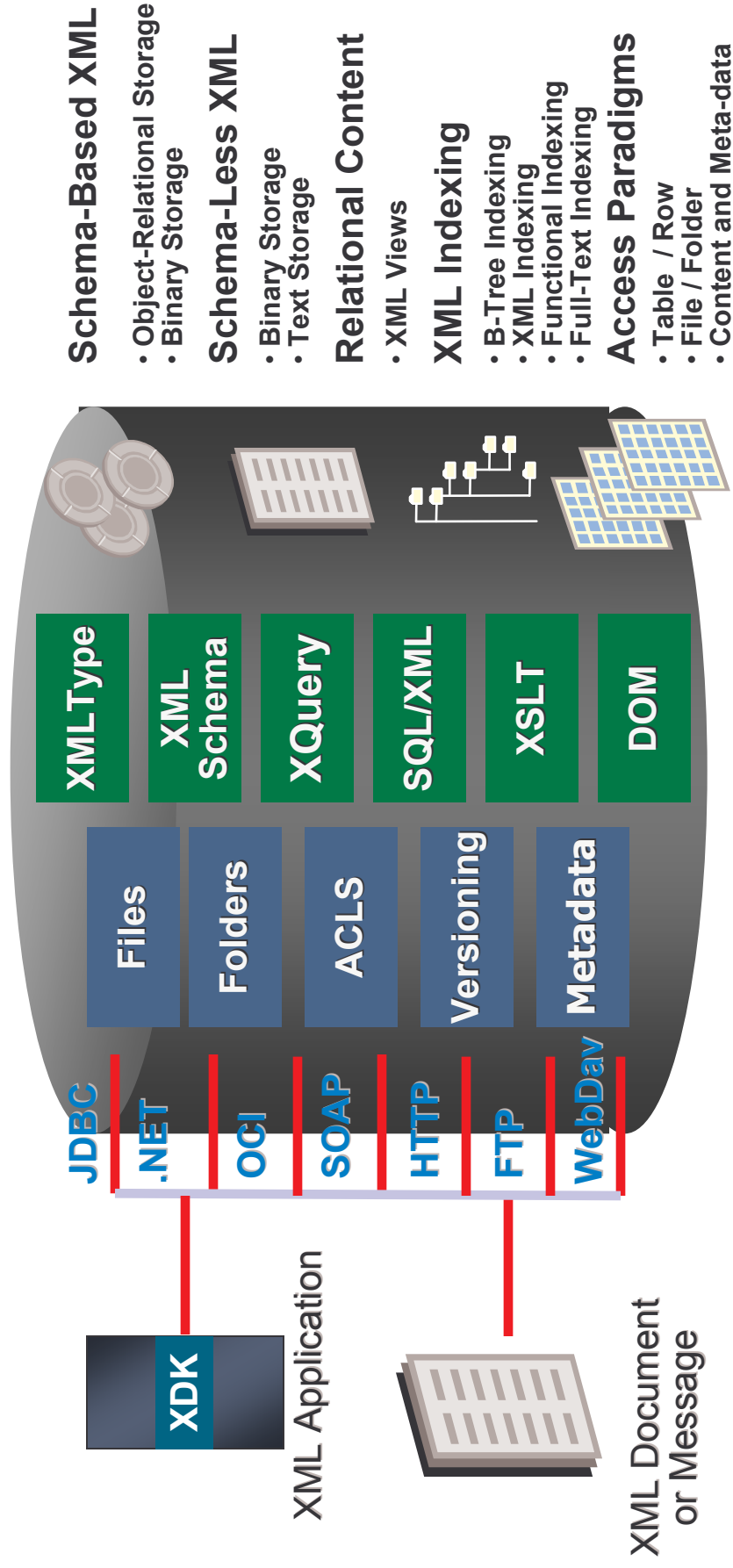# Compound Document Management with Oracle XML DB

Geoff Lee, Principal Product Manager
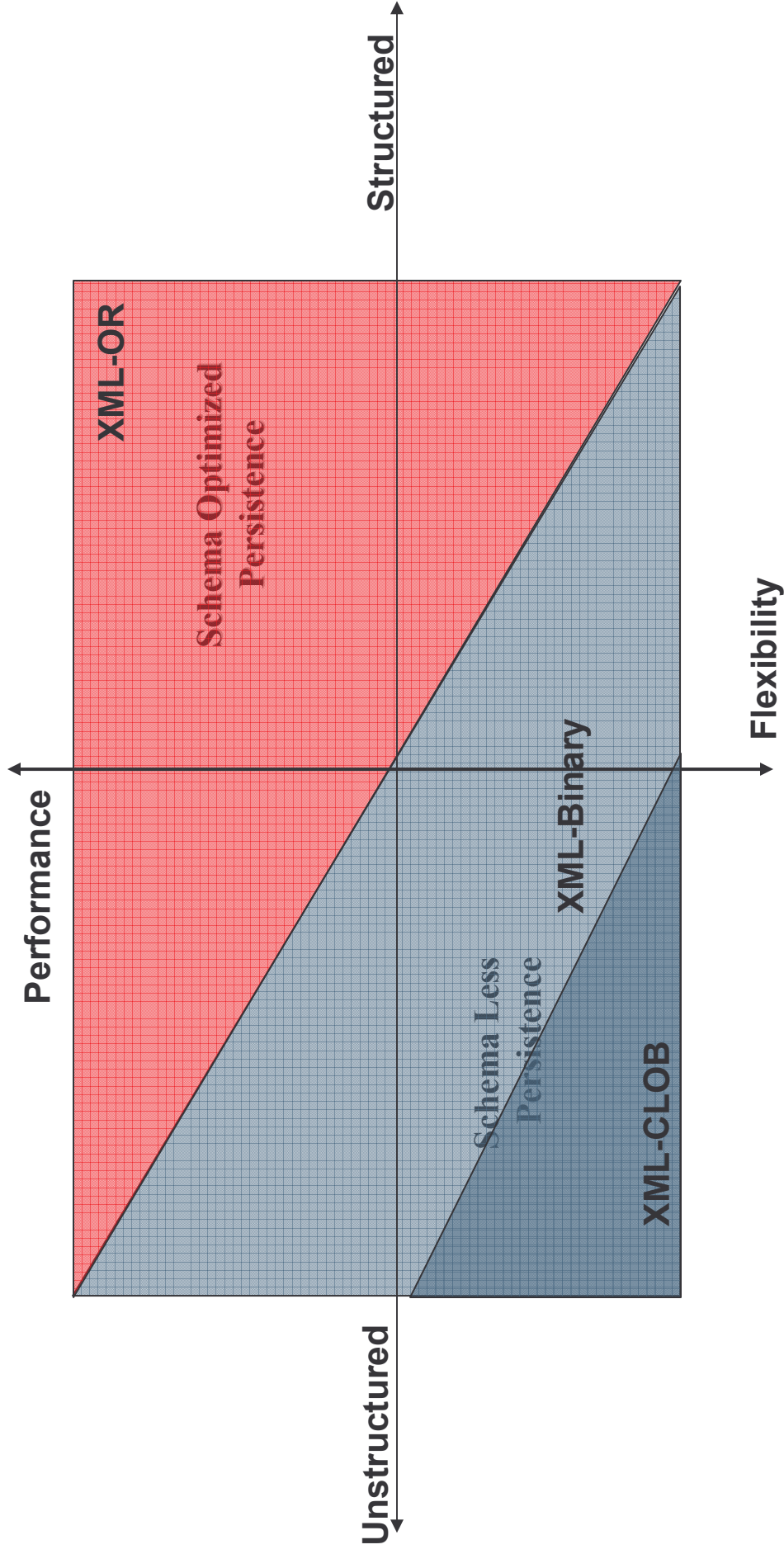Oracle Corporation

Introduction

- Growing volumes of Content
- Unstructured – Free form documents
  - Images
  - Web pages
  - Text, PDF files
- Semi-structured – XML Documents with extensible structure
  - Web 2.0 content (RSS feeds, SOA, etc)
  - Microsoft Office and Open Office use XML for content
  - Resumes
  - Biometrics information
  - User documentation
  - Vertical industry standard messaging formats

ORACLE

Oracle XML DB Overview

Schema-Based XML
- Object-Relational Storage
- Binary Storage

Schema-Less XML
- Binary Storage
- Text Storage

Relational Content
- XML Views

XML Indexing
- B-Tree Indexing
- XML Indexing
- Functional Indexing
- Full-Text Indexing

Access Paradigms
- Table / Row
- File / Folder
- Content and Meta-data

XMLType
XML Schema
XQuery
SQL/XML
XSLT
DOM

Files
Folders
ACLS
Versioning
Metadata

JDBC
.NET
OCI
SOAP
HTTP
FTP
WebDav

XDK

XML Application

XML Document or Message

ORACLE

XML Content Classification

ORACLE

Content Management Support in Oracle XML DB

- Access: Hierarchical repository/Protocols/PLSQL
- Management: Compound Documents/Events/User Defined Metadata
- Security: ACLs
- Version-control: XML DB Versioning
- Query-based Search: Web Services/XQuery
- Semi-structured content: Binary XML, XMLIndex
- Unstructured content: Secure Files

ORACLE

Compound Documents – W3C XLink and XInclude Standards

- XLinks –
  - Models arbitrary relationships between documents
  - Support for Simple links
  - Resource configuration controls its behaviour
- XInclude –
  - Includes multiple XML or text documents in a single infoset
  - Document can be decomposed on the way in and rebuilt on the way out
  - Support for versioning, locking, and access control

Compound Documents – XLink and XInclude

Example:

XInclude - A book is made of several chapters
XLink - Author's information is better stored separately

```
<Book
 xmlns:xi="http://www.w3.org/2001/XInclude"
 xmlns:xlink="http://www.w3.org/1999/xlink" >
<Author xlink:type="simple"
 xlink:href="/public/author/dir/paul.xml">
        A. Paul
</Author>
<xi:include href="toc.xml"/>
<xi:include href="chapter1.xml"/>
<xi:include href="chapter2.xml"/>
<xi:include href="index.xml"/>

</Book>
```


XML DB Repository

```
<Book>
<xlink.../>
<xinclude>
</Book>
```

```
<author>
.....
</author>
```

```
<chap10>
.....
</chap10>
```

```
<chap2>
.....
</chap2>
```
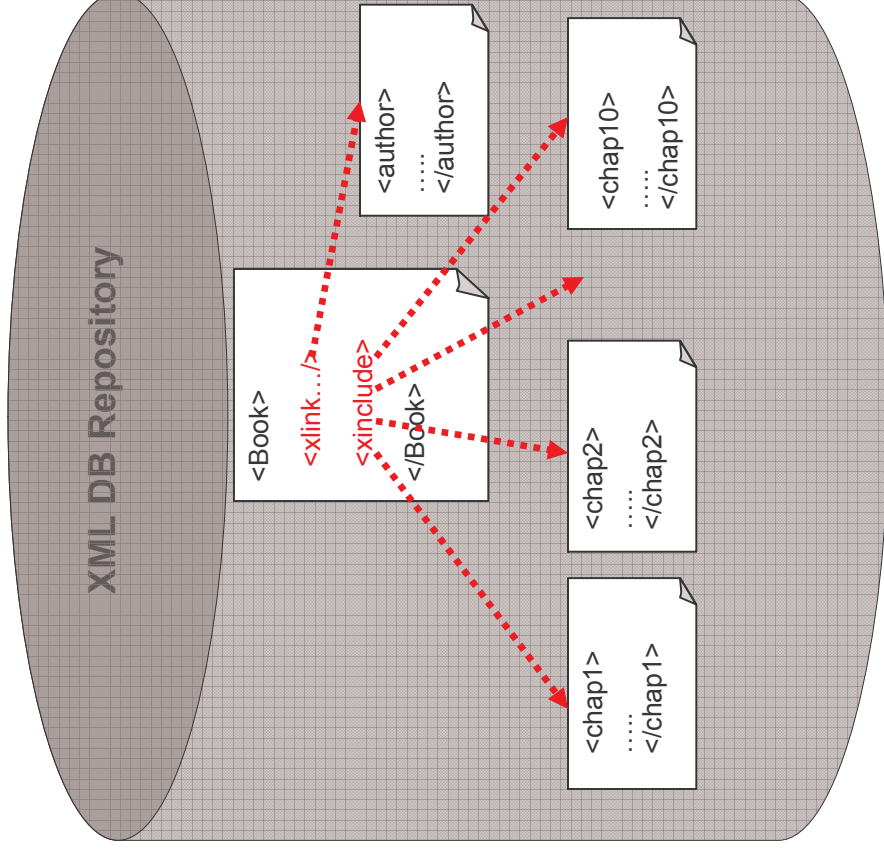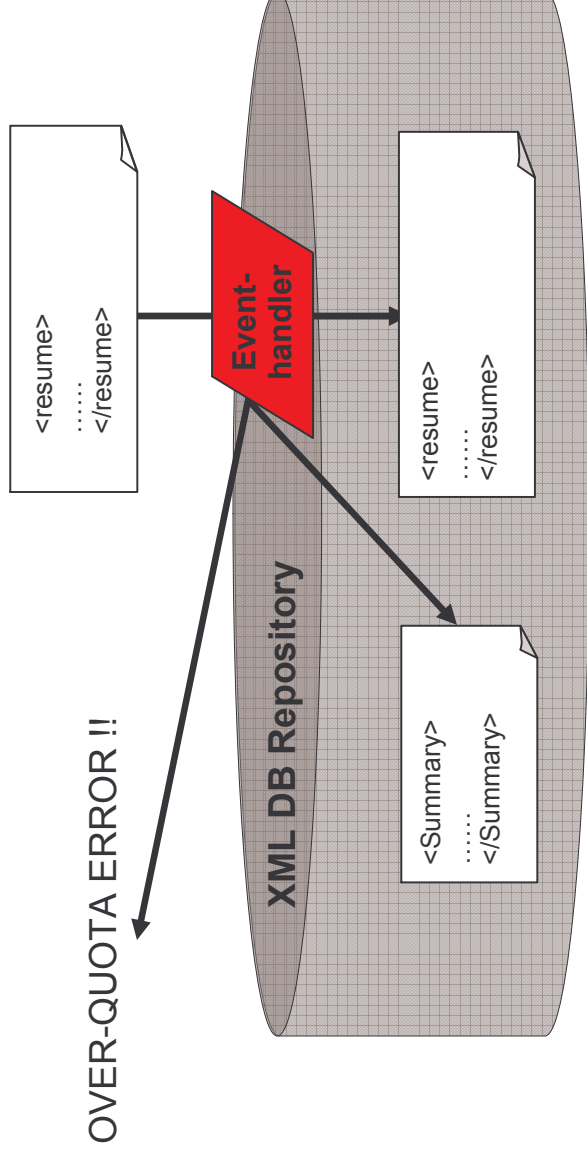
```
<chap1>
.....
</chap1>
```

Repository Events

- Repository event handlers are triggers for repository operations
- Attached to either individual resources or the entire repository
- The handlers can be implemented in Java or PL/SQL
- Event preconditions can be specified

```
<existsNode>
    <XPath>/Resource/[ContentType="text/xml"]</XPath>
    <namespace>xmlns="http://xmlns.oracle.com/xdb/XDBResource.xsd"</n
    amespace>
</existsNode>
```

- Resource configuration can specify default resource properties (ACLs, etc.)

Repository Events

<resume>
......
</resume>

Event-handler

OVER-QUOTA ERROR !!

XML DB Repository

<resume>
......
</resume>

<Summary>
......
</Summary>

Typical applications

- Extra security checks
- Quota
- Recycle bin
- Data transformation
- Logging

ORACLE

Access Control

- Access Control Lists (ACLs) protect repository resources
- Based on the DAV ACL standard
- ACLs define the privileges a user has on a resource
- Integrated with database security
- Extensible framework to suit various application needs
  - User-defined aggregate privileges
  - Security-class
  - Inheritance and constraints

Content storage

- PL/SQL, SQL, and network protocols can be used to store data
  - FTP, HTTP(S), NFS and WevDAV protocols are supported
- Binary XML provides efficient storage
- Content versioning
  - API to create a version controlled resource (VCR)
  - APIs to checkin-checkout versions
  - Supported for schemaless or schema-based resources if the schema tables have no associated triggers or constraints for access control

Content Retrieval and Search

- Retrieval – navigational or path-based access
  - Resource_view and Path_view for SQL access to the repository
  - PL/SQL and Protocol access
  - Automatic ACL checks
- Search across documents
  - XQuery supports repository access (fn:doc(), fn:collection(), etc.)
  - Web services support
  - XMLIndex provides query performance
  - Search resource properties by using resource_view and path_view
  - User defined metadata

ORACLE

Semi-structured content

- Binary XML
  - Streaming evaluation
  - Works across multiple tiers, supported in Java and C
  - Handles semi-structured content efficiently
- Secure Files
  - Support piecewise updates
  - Faster LOB read and write
  - Data compression
- XMLIndex
  - Efficient querying of content
  - Efficient querying of repository metadata
  - Efficient DML with Path Subsetting

Summary

- Oracle XML DB meets all your Content Management needs
  - High performance content storage and retrieval
  - Content management
  - Content search
  - Content security

The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

ORACLE®

ⓘ | ORACLE IS THE INFORMATION COMPANY