Information Retention and Oracle Database

Kevin Jernigan
Senior Director
Oracle Database Performance Product Management
The following 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.
Program Agenda

- What is Oracle Database?
- Oracle Database features related to retention
- Information Retention and Oracle Database
What is Information Retention?

- Keep information available and find-able
  - Structured: databases
  - Semi- and un-structured: documents

- What about:
  - Bit rot
  - Format changes
  - Software and hardware changes
What is Oracle Database?

- A Relational Database Management System – RDBMS
  - With 30+ years of innovations
  - With dozens of world record audited performance benchmarks

- Store and manage more than “just” structured data
  - XML
  - DICOM
  - RDF / OWL / SPARQL
  - Spatial / Graph
  - etc
Oracle Database Features Related to Retention

- Partitioning & Compression: storage tiering & ILM
- SecureFiles & DBFS: semi- and un-structured data
- Flashback Data Archive & CQN: events / streams & history
Oracle Database Partitioning & Compression

- Partitioning
  - Transparently split up data in tables
  - Improves query performance AND admin flexibility

- Compression
  - Advanced Compression
  - Advanced LOB Compression & Deduplication
Advanced Compression Option
Minimizes database storage, improves performance

- Compress data partitions/tables
  - OLTP and DW
- Compress unstructured data
  - Compression and deduplication
- Compress backups
  - Faster RMAN compression
  - Data Pump compression
- Compress Data Guard Redo log transport
  - Network compression
- Securely store historic data
  - Flashback Data Archive now part of Advanced Compression (in 11g)
Oracle Database Partitioning & Compression

- **Partitioning**
  - Transparently split up data in tables
  - Improves query performance AND admin flexibility

- **Compression**
  - Advanced Compression
  - Advanced LOB Compression & Deduplication
  - Hybrid Columnar Compression – Exadata, ZFSSA, Pillar Axiom
About Hybrid Columnar Compression

• Hybrid Columnar Compressed Tables
  – Compressed tables can be modified using conventional DML operations
  – Useful for data that is bulk loaded and queried

• How it Works
  – Tables are organized into Compression Units (CUs)
    • CUs are multiple database blocks
  – Within Compression Unit, data is organized by column instead of by row
    • Column organization brings similar values close together, enhancing compression
Oracle Database Partitioning & Compression

- **Partitioning**
  - Transparently split up data in tables
  - Improves query performance AND admin flexibility

- **Compression**
  - Advanced Compression
  - Advanced LOB Compression & Deduplication
  - Hybrid Columnar Compression – Exadata, ZFSSA, Pillar Axiom

- **Partitioning + Compression = Information Lifecycle Management (ILM)**
  - Keep more data in database for longer, with better performance
Oracle Database ILM

As data ages:
- Activity declines
- Volume grows
- Older data primarily for reporting

This Quarter
- Row Store for fast OLTP

This Year
- Compressed Column Store for fast analytics
- 10x compressed

Prior Years
- Archive Compressed Column Store for max compression
- 15x compressed

As data cools down, convert data to columnar compressed
Oracle Database SecureFiles & DBFS

- **SecureFiles**
  - High performance implementation of ANSI-standard LOB interfaces
  - Performance equal to or better than standard networked file systems
    - Throughput, response time, scalability

- **DBFS – Database File System**
  - Make Oracle Database behave like a file server
  - Access to files inside the database is now same as file system access
  - Performance and scalability: 1 billion file test
Database-Enabling File-Based Tools

- DBFS allows access to files in the db using file system interfaces
  - File operations translated into SQL operations
  - Directories and path names are derived from key columns in tables
  - Enables access by existing file-based tools

**DBFS Client**
- Acrobat Reader
- Network

**DBFS Server in DB**
- Path Names, Directories
  - /Customers
  - /Lucas
  - /Contract
  - /Photo

**SQL Access**

<table>
<thead>
<tr>
<th>ID#</th>
<th>Name</th>
<th>Address</th>
<th>Contract</th>
<th>Photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>Lucas</td>
<td>......</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Select Contract from Customers where Name = ‘Lucas’
Oracle DBFS HSM

- DBFS Hierarchical Storage Management
  - Transparently migrate a SecureFiles LOB to tape or cloud
  - Replace the LOB in the database with a DBFS Link
  - DBFS Link looks like the LOB to applications and users
    - Slower “time to first byte”
DBFS HSM Store

- A DBFS HSM store allows archiving files to tape
- Application migrates older files to HSM store (e.g. old invoices)
  - A DBFS Link replaces the LOB, LOB reads on links are transparent
  - A LOB can be easily migrated back to the table for updates
- HSM store: disk staging area for storing recently accessed files
  - Seldom accessed files are migrated to tape, brought back on reference

<table>
<thead>
<tr>
<th>Order#</th>
<th>Customer</th>
<th>Year</th>
<th>Invoice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>Lucas</td>
<td>2003</td>
<td></td>
</tr>
</tbody>
</table>

Sales Table

```
"/HSM/Old_Invoices/Invoice_1234"
```

DBFS Link

```
/HSM
```

```
/Old_Invoices
```

```
/Invoice_1234
```

LRU

Tape
Oracle DBFS HSM

- DBFS Hierarchical Storage Management
  - Transparently migrate a SecureFiles LOB to tape or cloud
  - Replace the LOB in the database with a DBFS Link
  - DBFS Link looks like the LOB to applications and users
    - Slower “time to first byte”

- SecureFiles + DBFS + DBFS HSM
  - Keep metadata in Oracle Database
  - Store most files offline, but still part of the database
Oracle Database Flashback Data Archive & CQN

- Flashback Data Archive – FDA
  - Transparently store history of changes
  - Automatically delete data that is older than specified retention period
  - Works with structured AND unstructured – SecureFiles – data
  - Beginnings of integrated transparent provenance

- CQN – Continuous Query Notification
  - Register SQL – actually CQL – queries in database
  - Each registration includes notification method / code
  - Fully contextual event processing
  - Combine with FDA and SecureFiles
Oracle Database and Information Retention

- Keep metadata and files together in Oracle Database
  - Unified backups
  - Unified security model
  - ILM and HSM
  - Leverage other database features

- Application developer agility
  - Much less work to implement content management applications
  - Just like SQL and RDBMS’s, reduces the amount of “wheel reinvention”