



IBM Software Group

DB2 Data Archive Expert

Catherine Chochoy

2 Février 2005

DB2. Data Management Software



© 2004 IBM Corporation

IBM Software Group | DB2 Data Management Software



DB2 Data Archive Expert Agenda

- Overview
- Terminology
- DB2 Grouper - a key component
- Product Architecture Overview
- Summary



© 2004 IBM Corporation

The Need for Archiving

As the size of a database grows, the percentage of inactive data often grows as well. Some of the major problems that occur with maintaining large volumes of inactive data include:

- Performance
- Manageability
- Extensive hardware and storage costs

DB2 Data Archive Expert - Purpose

- Archive related sets of information across multiple tables
- Selectively identify data to archive and retrieve
- Exploit less expensive media for your archived data
- Access archived data with SQL
- Access archived data with minimal or no application changes
- Defer the delete portion of the archive process
- Compress archived data with hardware data compression
- Define non DB2-enforced referential constraints
- Discover relationships between tables from definitional sources as well as access sources



IBM Software Group

DB2 Data Archive Expert - Terminology

DB2. Data Management Software



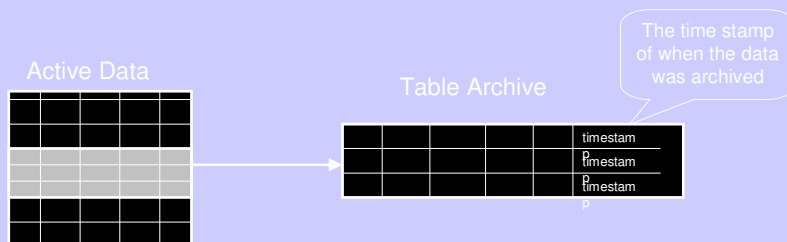
© 2004 IBM Corporation

IBM Software Group | DB2 Data Management Software



Definition: Archiving

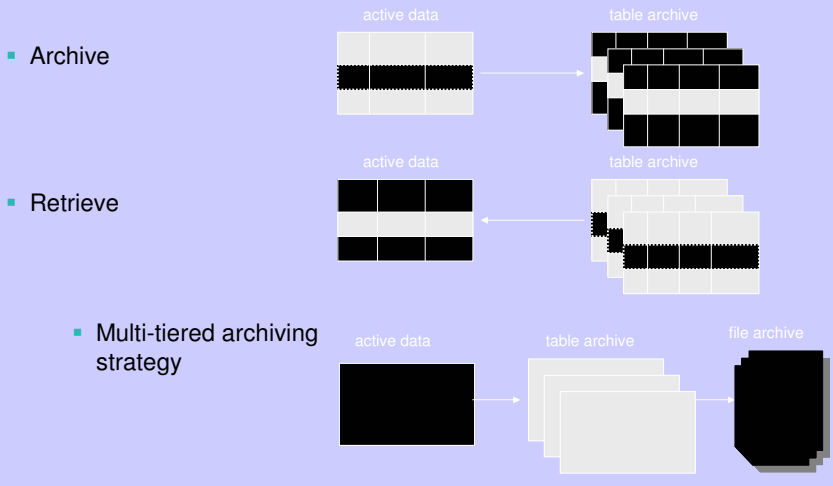
Archiving is the process of moving inactive data to another storage location that can be located and accessed when necessary.



© 2004 IBM Corporation

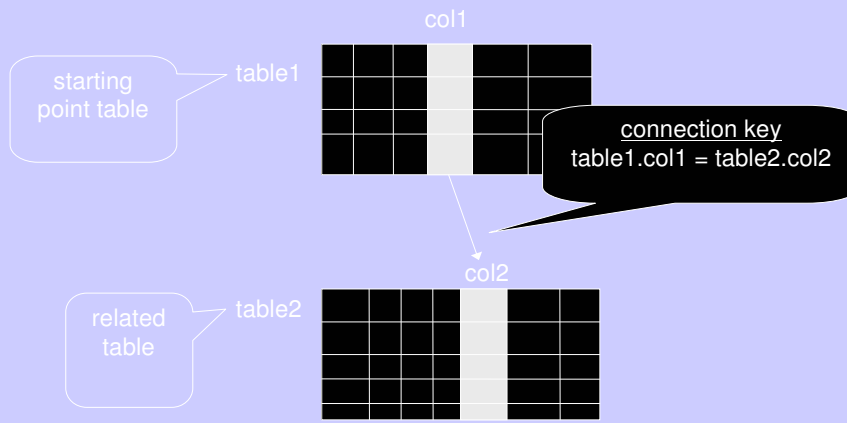


Archiving Terminology



Archiving Terminology...

- Archive Unit





IBM Software Group

Overview of Archiving with DB2 Data Archive Expert

DB2. Data Management Software



© 2004 IBM Corporation

IBM Software Group | DB2 Data Management Software



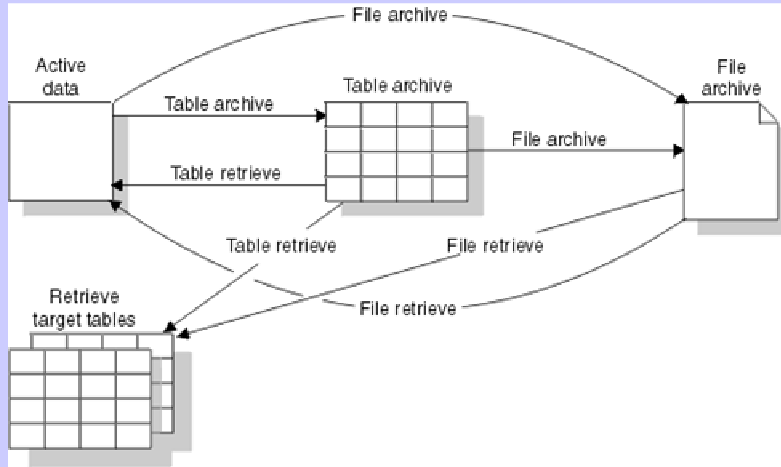
Benefits of Using DB2 Data Archive Expert

- Archive data across a set of related tables.
- Choice of archiving strategies.
- Choose when to remove the archived data from the source.
- Captures information about the archive.
- Allows you to schedule archiving activity.
- Robust meta-data management.
- Choice of relational or sequential archive target
- Deferred deletion with changed data detection to ensure maximum level of data integrity
- Automatic schema validation and change detection



© 2004 IBM Corporation

DB2 Data Archive Expert Model



DB2 Data Archive Expert Components

- An ISPF interface
- Specifications



- DB2 Data Archive Expert metadata
- A callable Application Programming Interface (API)
- DB2 Grouper



IBM Software Group

DB2 Grouper - a component of DB2 Data Archive Expert

DB2 Data Management Software



© 2004 IBM Corporation

IBM Software Group | DB2 Data Management Software



The Solution - Grouper

- Grouper is a component for discovering, recording, and managing groups of related objects (tables) that comprise a business application.
- Grouper has the flexibility to augment and edit the information stored within it to support business needs.
- Grouper is autonomic at a predictive level; when users input a table to Grouper, it returns all of the tables that are related to it. Users can then choose either all or a subset of the tables returned for the operation being performed.
- Grouper is reusable:
 - Groups of related tables can be defined for multiple uses
 - The Grouper component can be integrated with various DB2 tools

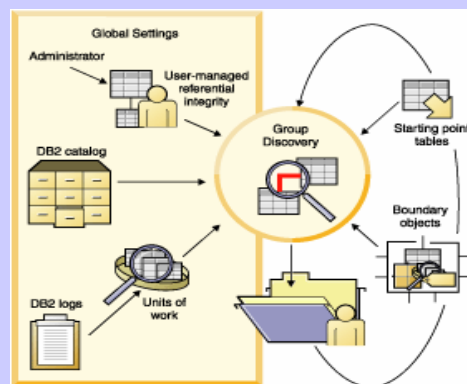
© 2004 IBM Corporation

Types of Table Relationships

- Referential Constraints (RI)
 - DB2 System-enforced
 - Non-enforced
 - Can consider RI constraints only for grouping purposes (such as archiving)
- Trigger
- Packages (tables in an application that contains static SQL)
- LOBs (z/OS)
- Table Hierarchy and Reference Columns (Linux/UNIX/Windows)
- Tables within the same commit scope (dynamic SQL)
- Additional specified relationships based on knowledge about the system

Group Administration (GUI)

- Maintain sets of groupings
- Specify global relationships (non-enforced RI)
- Kick off dynamic (UOW) relationship discoveries
- Specify input values to the group discovery process
 - Group discovery scope (starting points)
 - Boundary objects
 - Additional relationships
- Launch the group discovery process
- Browse through the groupings
- Edit the groups





IBM Software Group

Archiving with DB2 Data Archive Expert

DB2. Data Management Software



© 2004 IBM Corporation

IBM Software Group | DB2 Data Management Software 

Archive Unit

PROJECT

| PROJNO | PROJNAME | DEPTNO | PENDDATE |
|--------|-------------------|--------|----------|
| F200 | USER EDUCATION | C01 | 2/1/98 |
| M010 | WL PROGRAMMING | D11 | 2/1/98 |
| M011 | WL PROGRAM DESIGN | D11 | 12/1/92 |
| M012 | WL PROJECT DESIGN | D11 | 12/1/92 |
| M014 | WL PROJECT PRELIM | D11 | 12/1/92 |
| OP00 | OPERATION | E11 | 2/1/98 |

starting point table: PROJECT
 connection keys:
 DEPARTMENT.DEPTNO = PROJECT.DEPTNO
 DEPARTMENT.DEPTNO = EMPLOYEE.WORKDEPT
 related tables: DEPARTMENT & EMPLOYEE

DEPARTMENT

| DEPTNO | DEPTNAME |
|--------|------------------------|
| C01 | INFORMATION CENTER |
| D11 | MANUFACTURING SYSTEMS |
| D21 | ADMINISTRATION SYSTEMS |
| E11 | OPERATIONS |

EMPLOYEE

| EMPNO | LASTNAME | WORKDEPT |
|-------|-----------|----------|
| 60 | STERN | D11 |
| 130 | QUINTANA | C01 |
| 150 | ADAMSON | D11 |
| 180 | PIANKA | D11 |
| 140 | NICHOLS | C01 |
| 170 | YOSHIMURA | D11 |
| 310 | SETRIGHT | E11 |

© 2004 IBM Corporation



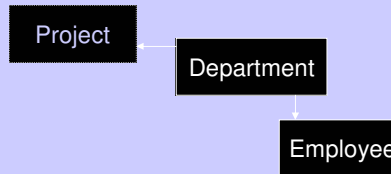
Steps in Archiving

Create an archive specification. The archive specification contains all of the information that is required to copy data from the archive source to the archive target.

1. **Select a starting point table.**
A starting point table designates the basis for the archive unit upon which a row filter is specified to determine the rows to be archived.
2. **Find related tables.**
Archive data is copied as an archive unit. If the find related tables function is used, then the archive unit is presented, but can be modified.

PROJECT

| PROJNO | PROJNAME | DEPTNO | PREKDATE |
|--------|-------------------|--------|----------|
| P000 | USER EDUCATION | D01 | 21/1/90 |
| MR010 | WL PROGRAMING | D11 | 21/1/90 |
| MR011 | WL PROGRAM DESIGN | D11 | 12/1/92 |
| MR012 | WL PROJECT DESIGN | D11 | 12/1/92 |
| MR013 | WL PROJECT FINISH | D11 | 12/1/92 |
| OP000 | OPERATION | E11 | 21/1/90 |



Steps in Archiving...

3. **Identify connection keys.**
Connection keys are columns in the tables of the archive unit that provide relationships between the tables. If the find related tables function is used, these columns are presented, but can be modified.
4. **Specify junction tables.**
A junction table is a table in the archive unit that contains connection keys to other tables in the unit, but is a table for which you do not want its data to be archived.

PROJECT (starting point table)

| PROJNO | PROJNAME | DEPTNO | PREKDATE |
|--------|-------------------|--------|----------|
| P000 | USER EDUCATION | D01 | 21/1/90 |
| MR010 | WL PROGRAMING | D11 | 21/1/90 |
| MR011 | WL PROGRAM DESIGN | D11 | 12/1/92 |
| MR012 | WL PROJECT DESIGN | D11 | 12/1/92 |
| MR013 | WL PROJECT FINISH | D11 | 12/1/92 |
| OP000 | OPERATION | E11 | 21/1/90 |

DEPARTMENT
(junction table)

| DEPTNO | DEPTNAME |
|--------|------------------------|
| D01 | INFORMATION CENTER |
| D11 | MANUFACTURING SYSTEMS |
| D21 | ADMINISTRATION SYSTEMS |
| E11 | OPERATIONS |

EMPLOYEE

| EMPNO | LASTNAME | WORKDEPT |
|-------|-----------|----------|
| 60 | STERN | D11 |
| 130 | QUINTANA | D01 |
| 150 | ADAMSON | D11 |
| 160 | PIANKA | D11 |
| 140 | NICHOLS | D01 |
| 170 | YOSHIMURA | D11 |
| 310 | SETRIGHT | E11 |

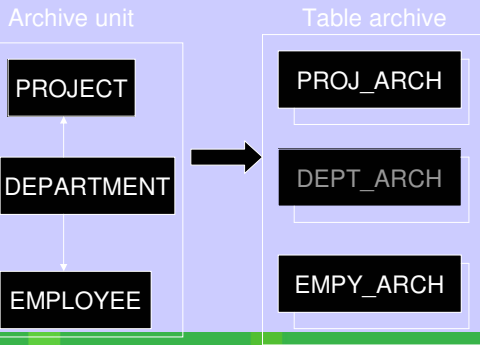
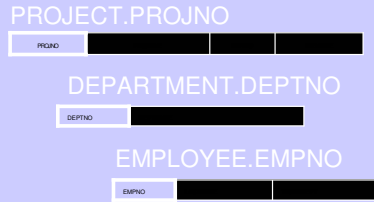
↑ department.deptno = project.deptno

↓ department.deptno = employee.workdept



Steps in Archiving...

- Specify row uniqueness.**
 In most cases DB2 Data Archive Expert requires row uniqueness. DB2 Data Archive Expert can locate any primary keys or unique columns that exists. If one does not exist, the user must then specify unique columns.
- Specify table archive targets.**
 Specify one or more table spaces in which to place archive tables, or you may choose to use the DB2 Data Archive Expert default archive tables.



Steps in Archiving...

- Specify a row filter.**
 A row filter is the criteria applied to the starting point table that selects the rows to be archived.

WHERE PROJECT.PRENDATE = '12/1/1982'

PROJECT

| PROJNO | PROJNAME | DEPTNO | PRENDATE |
|--------|-------------------------|--------|-----------|
| | | | |
| | | | |
| MA211 | WL PROGRAM DESIGN | D11 | 12/1/1982 |
| MA212 | WL PROJECT DESIGN | D11 | 12/1/1982 |
| MA213 | WL PROJECT CONF PROCESS | D11 | 12/1/1982 |
| | | | |

DEPARTMENT

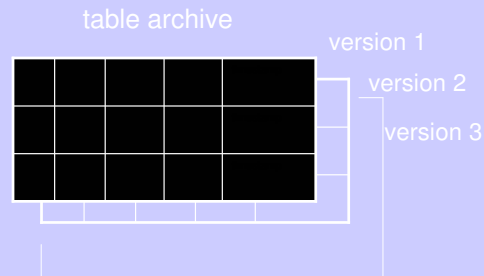
| DEPTNO | DEPTNAME |
|--------|------------------------|
| 001 | INFORMATION CENTER |
| 011 | MANUFACTURING SYSTEMS |
| 021 | ADMINISTRATION SYSTEMS |
| 011 | OPERATIONS |

EMPLOYEE

| EMPNO | LASTNAME | WORKDEPT |
|-------|-----------|----------|
| 60 | STERN | D11 |
| 130 | GUENTHER | D01 |
| 150 | ADAMSON | D11 |
| 160 | PANKA | D11 |
| 140 | HORNOLLS | D01 |
| 170 | YOSHIMURA | D11 |
| 310 | SE TRIGHT | D11 |

Steps in Archiving...

8. Running an archive specification.
 Each time that an archive specification is run, a new set of archive data is created. This new set is called an archive data version. Many versions of archived data can exist.
- Different row filters can be used for each run of the archive specification.



Summary : how does DB2 Data Archive Expert for z/OS help customers?

- By improving
 - Performance
 - Manageability
 - Hardware and storage costs
 and thus reducing IT costs
- By freeing up developers from writing customized archiving software
- By providing an SQL callable interface for ease in scheduling
- By discovering related tables using the DB2 Grouper component
- By allowing the data to be removed/deleted from the source independently from the copy to the archive
- By working with data hardware compression
- By capturing all pertinent information about the archive
- By providing a choice of archiving strategies
 - To table
 - To file
 - To both (multi-tier)

Unique product capability

IBM Software

IBM

SG24-7080-00

Draft Document for Review January 5, 2004 5:20 pm

IBM DB2 Data Archive Expert for z/OS: Put Your Data in Its Place

- Reduce disk occupancy by removing unused data
- Streamline operations and improve performance
- Filter and associate data with DB2 Grouper



Paolo Bruni
Walter Huth
Ernie Mancill
Iain Warnock

SG24-7080
Authors: Paolo Bruni, Walter Huth, Ernie Mancill, Iain Warnock



© 2004 IBM Corporation

ibm.com/redbooks


Redbooks

IBM Software Group | DB2 Data Management Software

IBM

References

- Other publications on DAE performance & implementation
 - DB2 for z/OS and OS/390: Ready for JAVA* - SG24-6435-00
 - Cross Platform DB2 Stored Procedures: Building and Debugging* - SG24-5485-01
- Redbooks
 - DB2 Data Archive Expert for z/OS: Put Your Data in Its Place* - SG24-7070
 - Squeezing the Most out of Dynamic SQL with DB2 for z/OS and OS/390* - SG24-6418
 - DB2 Java Stored Procedures - Learning by Example, SG24-5945* for WLM definition and JAVA stored procedure performance
- Using UNION in VIEW
 - Zero Latency Enterprise Data Store Techniques for DB2 for OS/390 and z/OS*
 - <http://www.idug.org/idug/member/journal/Sept01/articl06.cfm>
 - SQL Techniques to Solve Design Challenges*
 - <http://www.idug.org/idug/member/journal/aug02/sqltech.cfm>



© 2004 IBM Corporation