DB2 Information Management Software

IBM Software Group

Protect, Retain and Comply with your DB2 Data

Cécile Benhamou
Catherine Chochoy

DB2 Technical Sales

© 2006 IBM Corporation

Life is not easy

Regulation	Date	Purpose	Implications
Basel II	2007	Improve measurement of total risk and strengthen ability to determine capital needed	Devise and implement new methodologies and systems to measure and monitor credit and operational risk (only a select number of top institutions required to comply, but others choosing to do so voluntarily)
Sarbanes-Oxley	Enacted	Strengthen financial reporting, internal controls by fixing responsibility within companies' management	CEO and CFO certifications of annual and quarterly SEC reports; evaluation of the effectiveness of internal controls; rapid disclosure of material changes in financial conditions or operations. Set up automatic controls repository to identify deficiencies
HIPAA	Enacted	secure medical records (lifetime), prove how they have been used & who has used them	Prohibit use or disclosure of protected information. Track authorizations by individual or explicitly authorized by the Regulations
Patriot Act	Enacted	To prevent usage of the financial system to support illegal activities, particularly terrorism	Customer Identification programs and anti-money laundering programs
Various anti-money laundering (AML)	Enacted	Prevent the laundering of money derived from illegal activities	Build or acquire automated risk management systems, CRM systems capable of identifying account holders' activities
Gramm-Leach-Bliley	Enacted	Protection of personally identifiable financial information	Ensure the security and confidentiality of customer records; protect against anticipated threats and unauthorized access

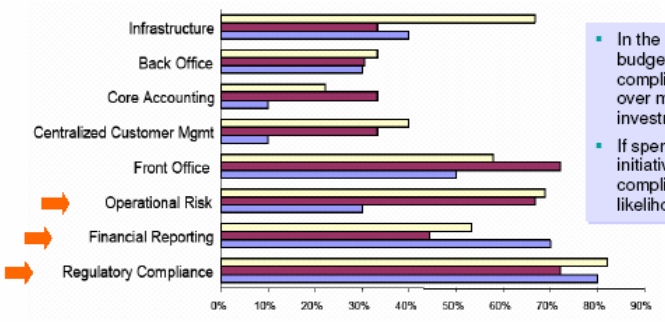
© 2006 IBM Corporation

Nor getting easier…

- Japanese Protecting Personal Freedom Act
- Department of Defense - 5015.2
  - requires certified application or technology to manage records (retention)
- SEC Rule 17a-4
  - requires brokers to preserve communications with clients (6 years)
- Corporate Information Security Accountability Act of 2003
  - requires audit of IT security and reporting
  - security infrastructures meet minimum standards
- California Bill 1386
  - a bill that protects data concerning California Residents in all computers across the United States
- European Union
  - various countries are working on proposed bills to protect data concerning EU Residents
- ... and more to come

© 2006 IBM Corporation

Top Business Priorities



- In the event of unforeseen budget constraints, compliance will take priority over many other strategic IT investments.
- If spending for other IT initiatives can be tied to compliance, may increase likelihood of approval.

Source: Financial Insights, "2004 IT Spending Survey Results: Diversity Among Banks?" May 2004.

© 2006 IBM Corporation

The Bottom Line – Improving Internal Control

Regulators have multiple goals. .

- Security of the national and international services infrastructure
- Improved risk management across the enterprise
- Integrity of financial reporting processes and related business practices
- Customer information security

. . . which drive investment in several areas

- People: Professionals with regulatory experience will be hired to enable firms to meet and anticipate new regulatory requirements
- Process: More robust processes and procedures will enable top management to monitor and enhance regulatory compliance
- **Technology: Significant investment will be made to do the following:**
  - Encrypt sensitive data
  - Protect sensitive production data
  - Save data for future audits and to comply with retention rules
  - Auditability - discover who did what, where and when
    - Real time
    - Historically
  - Engage in real-time monitoring of operations

© 2006 IBM Corporation

Encrypting your Data

© 2006 IBM Corporation

Encryption: fonctions construites

- Encryption des Données et des Index
- Encryption au niveau de la colonne
  - Colonne définie en VARCHAR et FOR BIT DATA
    - CREATE TABLE EMP (EMPNO VARCHAR(32) FOR BIT DATA);
- Mot de passe pour l'encryption
  - SET ENCRYPTION PASSWORD = :hv_pass;
- Utilisation du mot clé ENCRYPT pour insérer des données encryptées dans la table EMP:
  - INSERT INTO EMP (EMPNO) VALUES(ENCRYPT('47138'));
- Afficher une colonne avec les données décryptées:
  - SELECT DECRYPT_CHAR(EMPNO,:hv_pass) FROM EMP;
    - Si mot de passe correct, DB2 retourne le numéro d'employé décrypté
- Remarque: On peut utiliser le registre spécial SET ENCRYPTION PASSWORD
- Utilisation de GETHINT: fournit un indice permettant de se rappeler le mot de passe

© 2006 IBM Corporation

Protecting Sensitive Production Data

IBM DB2 Test Database Generator

© 2006 IBM Corporation

Test Data Base Generator

- A powerful tool that provides several methods of **generating test data for DB2 easily**
  - From scratch or from existing data sources
- Maintains **referential integrity** while extracting data sets from source databases
- Create test data in **new** or **existing** tables
- Copy a **slice** of data instead of all of the data
- Create a restructured database for testing
- **Useful for regulatory compliance and data protection**

© 2006 IBM Corporation

DB2 Grouper - a common component

- There are **many relationships** between the various DB2 objects, such as tables, in a business application. **Some** of these relationships can be **discovered easily**, while **others are hard to find and manage**.
- These relationships can be “dynamic” or “transient”, they can be hidden in low use (monthly or year-end applications) making them difficult to see, complicating managing the business applications efficiently.
- As the complexity of data relationships increases, it becomes increasingly difficult to discover and keep track of this information.
- **Grouper is a component for discovering, recording, refining, and managing groups of related objects (tables) that comprise a business application.**
  - Grouper utilizes this relationship (grouping) information as the basis for consistent administrative management activities.

© 2006 IBM Corporation

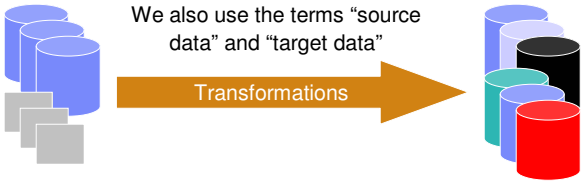
DB2 Grouper :Types of Table Relationships

- Referential Constraints (RI)
  - DB2 System-enforced
  - Non-enforced
- Triggers
- Packages (tables in an application that contains static SQL)
- Table Hierarchy and Reference Columns (LUW)
- LOBs (whether in the same or different table space)
- *Tables within the same commit scope (dynamic SQL)*

© 2006 IBM Corporation

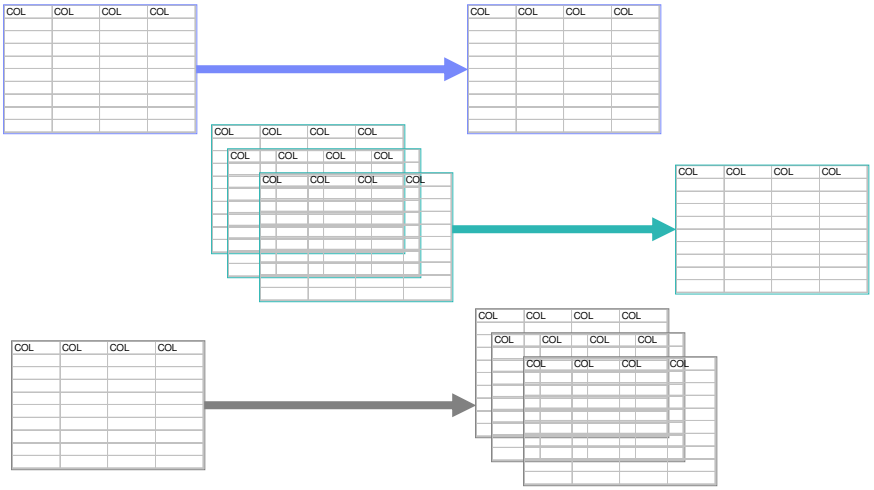
Introducing the TDBG computing model

1. **Start with data that exists somewhere in your enterprise**
2. Leverage knowledge of data relationships
3. Apply transformation rules
4. Create test data



© 2006 IBM Corporation

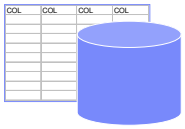
Start with data that exists means more than just make a duplicate copy…



© 2006 IBM Corporation

Data sources

- Data sources can be database tables or views or files

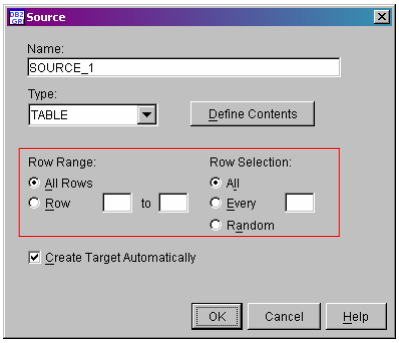


- File types supported include
  - CSV (comma separated value) file
  - Text file with any delimiter
  - Text file of fixed width
- NOTA: Starting with existing data is an optional step
  - You can choose to direct TDBG to create test data “from scratch”

© 2006 IBM Corporation

Identifying the data you want to start with

- Multiple levels of filtering
- WHERE clause to limit result set
  - When source is a table
- Range
  - All
  - From/to
- Row selection within the range
  - Every row sequentially
  - Every nth row
  - Every row randomly



© 2006 IBM Corporation

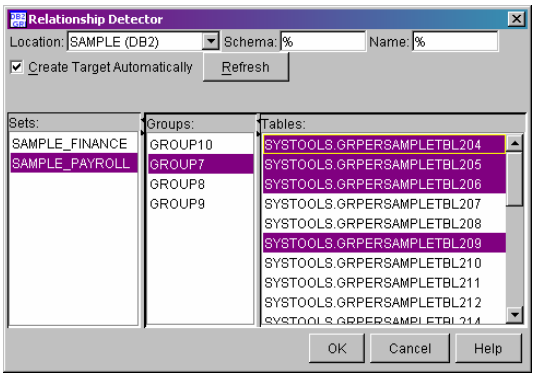
Introducing the TDBG computing model

1. Start with data that exists somewhere in your enterprise
2. **Leverage knowledge of data relationships**
3. Apply transformation rules
4. Create test data

© 2006 IBM Corporation

Would you like to include related tables?

- Related tables are detected
  - RI defined
- You can choose to include any number of related sources



© 2006 IBM Corporation

Copy related data

- Copy all rows in a related group
  - Copy all rows that are related across an entire related set of tables
- Copy a slice of data across a related set of tables
  - Start with a specific customer and copy that row and all rows related to that customer across all related tables

© 2006 IBM Corporation

Introducing the TDBG computing model

1. Start with data that exists somewhere in your enterprise
2. Leverage knowledge of data relationships
3. **Apply transformation rules**
4. Create test data

© 2006 IBM Corporation

Transformation rules

- Transformation rules define the target test data
  - How to generate test data from source data
  - How to generate test data from scratch
- *Examples*

Create a target column PHONE which is the combination of a country code (derived from COUNTRY file), an area code from TABLE1, and a phone number from TABLE2

Create a target column ACCT_BALANCE which is a random number that falls within a specified range

Create a target column that is exactly the PIN column with the 3rd and 5th positions replaced (masked) with the letter X

© 2006 IBM Corporation

Scopes and sets

- You define your test data one target column at a time
- The scope of a transformation rule set is target column
- Multiple rules can be specified for each target column
- Transformation rules are applied in order
- Each rule can modify, replace, append, or preface the previous value to allow for incremental building of a target column

© 2006 IBM Corporation

Transformation rules

1. Static Values
2. Source Column Values
3. Data Lookup
4. Data Masking
5. Expressions
6. Random Values
7. Pattern Generation

© 2006 IBM Corporation

Introducing the TDBG computing model

1. Start with data that exists somewhere in your enterprise
2. Leverage knowledge of data relationships
3. Apply transformation rules
4. **Create test data**

© 2006 IBM Corporation

What can TDBG create?

- TDBG can generate test data as ….
  - Comma separated value (.csv) file
  - Text file with fixed width columns
  - Delimited text file
  - File of SQL INSERT statements
  - DB2 for z/OS internal load format file
  - XML file
- TDBG can also …
  - Generate test data and directly insert it into DB2

© 2006 IBM Corporation

You can also choose…

- How many rows of test data to generate
- The exact structure of the target table(s)
  - Inherit from source (auto-target feature)
  - Specify column definitions
- To automatically generate DDL needed to create the target table(s)
  - Includes primary and foreign keys
  - Can be modified

© 2006 IBM Corporation

Summary : DB2 Test Database Generator

- **For when you need to…**
  - **Create a restructured database for testing**
  - **Create test data in a variety of output formats**
  - **Create test data in new or existing tables**
  - **Copy a slice of data instead of all of the data**
  - **Mask and censor sensitive data for testing**
- **Provides several methods of generating test data for DB2**
  - **Can create scripts that can be run now, later, and repeatedly**
- **Supports multiple interfaces**
  - **GUI(Java), Command line, ISPF(Z/OS only)**
- **Supports DB2 across multiple platforms** - assuming suitably licensed
  - **Unix (AIX & Sun), Windows, Linux, z/OS**

© 2006 IBM Corporation

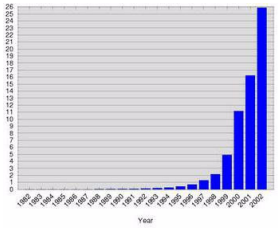
Saving Data For Future Audits And To Comply With Retention Rules

IBM DB2 Data Archive Expert

© 2006 IBM Corporation

Inactive (dormant*) Data

- **Inactive** data exists everywhere
  - Transactional histories
  - DW (Data Warehouses)
  - e-mail, ...
- **Inactive** data is not unusable;
  --it just has a lower probability of access
  - Data that is unusable should be eliminated
- Why keep inactive data?
  - Legislated by government
  - Business need
    - Trend analysis,
    - Previous customer history
- As databases grow in size,
  the percentage of inactive data grows →



10GB	10% inactive
100GB	50% inactive
2TB	80% inactive
50TB	95% inactive

*Reference: Inmon

© 2006 IBM Corporation

Inactive Data - COSTS YOU

- IT executives are concerned how to manage complex enterprise databases that keep growing larger
- Expanding capacity is not always the most effective way to deal with growth
- Allowing databases to grow without bounds can affect bottom line by increasing costs in the following ways:
  - Performance
  - Manageability
  - Hardware and storage costs

© 2006 IBM Corporation

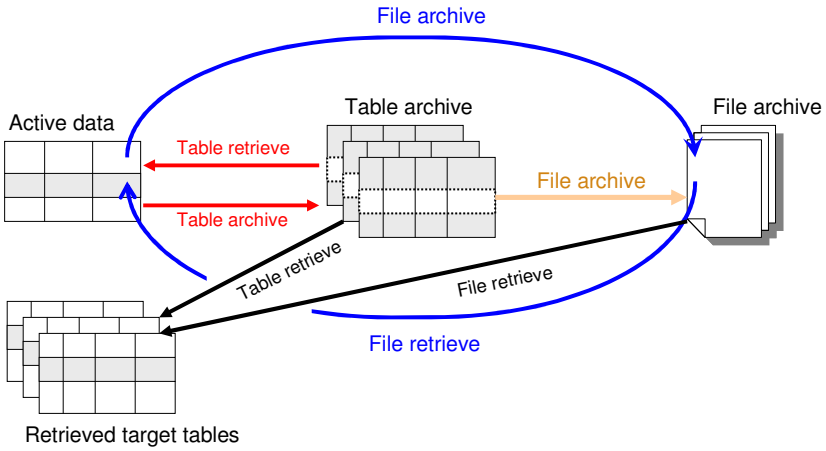
Introduction

- What is archiving?
  - **The process of removing data from active data stores while providing the capability of accessing the removed data at a later time**
  - Archive/Retrieve (as opposed to Backup/Recover)
  - Why isn't an archive the same as a backup?
    - Archives contain sets of selective data
    - Archives are application oriented
    - Backups tend to be data store oriented

However, an archive could be an important element in a backup/recovery scheme

© 2006 IBM Corporation

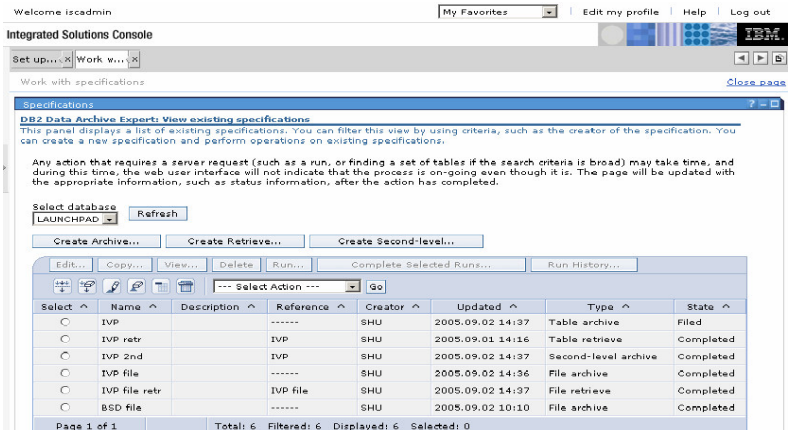
Multi-tiered Archiving Strategy



© 2006 IBM Corporation

IBM DB2 Data Archive Expert for Multiplatform

- Provides a Web interface that helps you to configure and use DB2 Data Archive Expert.



© 2006 IBM Corporation

Archive Unit

starting point table: PROJECT

connection keys:
DEPARTMENT.DEPTNO = PROJECT.DEPTNO
DEPARTMENT.DEPTNO = EMPLOYEE.WORKDEPT

related tables:
DEPARTMENT & EMPLOYEE

PROJECT

PROJNO	PROJNAME	DEPTNO	PRENDATE
IF2000	USER EDUCATION	C01	2/1/1983
MA2110	W L PROGRAMMING	D11	2/1/1983
MA2111	W L PROGRAM DESIGN	D11	12/1/1982
MA2112	W L ROBOT DESIGN	D11	12/1/1982
MA2113	W L PROD CONT PROGS	D11	12/1/1982
OP1010	OPERATION	E11	2/1/1983

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
160	PIANKA	D11
140	NICHOLLS	C01
170	YOSHIMURA	D11
310	SETRIGHT	E11

© 2006 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**

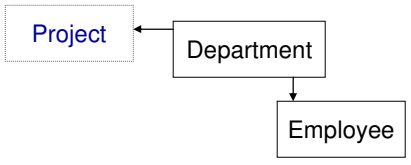
   A **starting point table** designates the basis for the **archive unit** upon which a **row filter** is specified to determine the rows to be archived.

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.

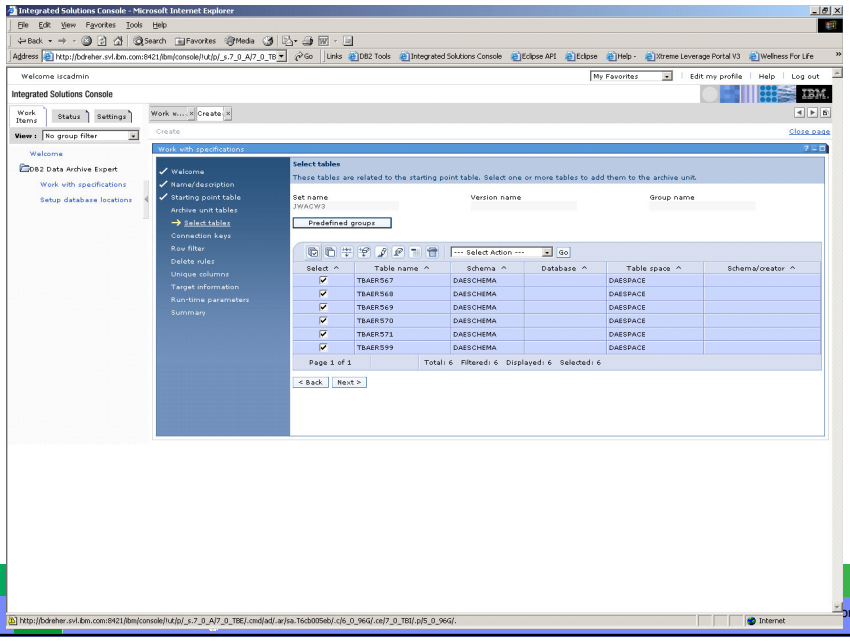
PROJECT

PROJNO	PROJNAME	DEPTNO	PRENDATE
IF2000	USER EDUCATION	C01	2/1/1983
MA2110	W L PROGRAMMING	D11	2/1/1983
MA2111	W L PROGRAM DESIGN	D11	12/1/1982
MA2112	W L ROBOT DESIGN	D11	12/1/1982
MA2113	W L PROD CONT PROGS	D11	12/1/1982
OP1010	OPERATION	E11	2/1/1983



© 2006 IBM Corporation

ISC screen shot



orporation

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.

PROJECT (starting point table)

PROJNO	PROJNAME	DEPTNO	PRENDATE
IF2000	USER EDUCATION	C01	2/1/1983
MA2110	W L PROGRAMMING	D11	2/1/1983
MA2111	W L PROGRAM DESIGN	D11	12/1/1982
MA2112	W L ROBOT DESIGN	D11	12/1/1982
MA2113	W L PROD CONT PROGS	D11	12/1/1982
OP1010	OPERATION	E11	2/1/1983

department.deptno = project.deptno

DEPARTMENT (junction table)

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

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 whose data is not to be archived.

EMPLOYEE

department.deptno = employee.workdept

EMPNO	LASTNAME	WORKDEPT
60	STERN	D11
130	QUINTANA	C01
150	ADAMSON	D11
160	PIANKA	D11
140	NICHOLLS	C01
170	YOSHIMURA	D11
310	SETRIGHT	E11

© 2006 IBM Corporation

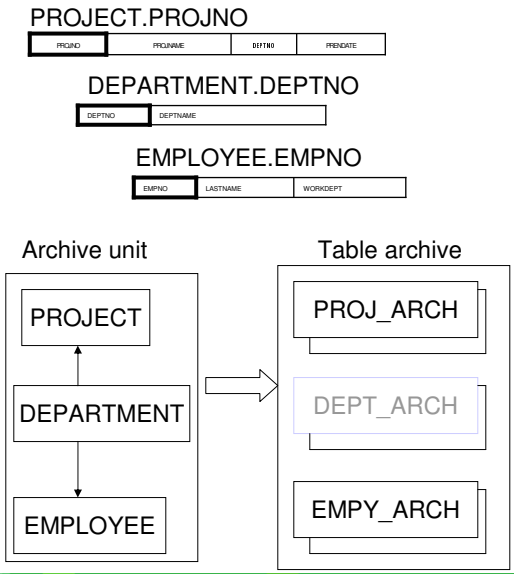
Steps in Archiving...

5. Specify row uniqueness

In most cases DAE requires row uniqueness. DAE can locate any primary keys or unique columns that exist. If they do not exist, the user must then designate unique columns.

6. Specify table archive targets

Specify one or more table spaces in which to place archived tables, or choose to use the DAE default archive tables.



© 2006 IBM Corporation

Steps in Archiving...

7. Specify a row filter

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

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

PROJECT

PROJNO	PROJNAME	DEPTNO	PRENDATE
IF2000	USER EDUCATION	C01	2/1/1983
MA2100	WELD LINE AUTOMATION	D01	2/1/1983
MA2110	W L PROGRAMMING	D11	2/1/1983
MA2111	W L PROGRAM DESIGN	D11	12/1/1982
MA2112	W L ROBOT DESIGN	D11	12/1/1982
MA2113	W L PROD CONT PROGS	D11	12/1/1982
OP1000	OPERATION SUPPORT	E01	2/1/1983
OP1010	OPERATION	E11	2/1/1983

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
160	PIANKA	D11
140	NICHOLLS	C01
170	YOSHIMURA	D11
310	SETRIGHT	E11

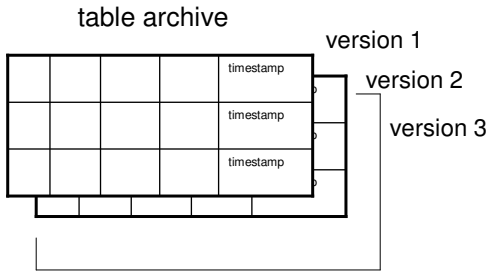
© 2006 IBM Corporation

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.

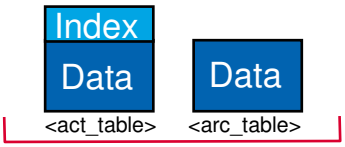


© 2006 IBM Corporation

Retrieving data from the program's point of view

- By allowing data to be archived to tables, applications can access the archive

Application still reads from <table_name>



- Application reads need to supply a date_col predicate to avoid access to <arc_table>
- Updates need to point to the act_table
- Prevent updates to arc _table

```
CREATE VIEW <table_name> AS
 SELECT * FROM <act_table>
  WHERE DATE_COL >= (CURRENT_DATE-365 DAYS)
UNION ALL
 SELECT * FROM <arc_table>
  WHERE DATE_COL < (CURRENT_DATE-365 DAYS)
```


© 2006 IBM Corporation

Benefits of Using DB2 Data Archive Expert

- Archive data (rows) across a set of related tables
  - Application perspective
- Choice of archiving targets (table, file)
- Allows you to discover related tables (Grouper) using system RI and application RI (UMRI)
- Allows scheduling archiving activity
- Choose when to remove the archived data from the source
- Captures meta-data about the archive
- External callable interface using stored procedures
- Archives X-platform

© 2006 IBM Corporation

Séminaire technique Data Archive Expert

- **Séminaire Technique - Proof of Technology (POT) - sur les outils DB2 multi-plateforme (environnement distribué).**

  Qui est concerné : le séminaire s'adresse aux DBA et autres personnes qui souhaitent évaluer les logiciels. Il n'est pas demandé de connaissances particulières en programmation.

  Coût : Ce séminaire technique est gratuit. Les participants prennent à leur charge les frais de voyage et d'hôtel.

  Horaire : de 9:00 heures à 17:30 heures.

  Durée : 2 jours

  Lieu : TEC (Technical Exploration Center - IBM Noisy Le Grand)

  Instructeur : Sylvain Bacquet, Information Management Consultant.

  Inscriptions : par email à sylvain_bacquet@fr.ibm.com ; tel +33(0) 4 72 18 42 61

- **Les outils concernés :**

  DB2 Performance Expert : Les performances de vos bases de données, passées au crible

  DB2 Data Archive Expert : Outil d'archivage des données complet permettant de déplacer des données rarement utilisées vers les archives

  DB2 Recovery Expert : Pour restaurer complètement et automatiquement vos ressources de base de données

  DB2 Change Management Expert : Réduction du taux d'erreurs humaines grâce à l'automatisation et la gestion

© 2006 IBM Corporation

Evénement de référence pour la gestion de l'information

5000 participants

- Ecouter les entreprises en pointe
- Apprendre auprès des experts et architectes
- Planifier son approche 'Information On Demand'
- Faire son marché de solutions
- Echanger, se former, se certifier

Contenu

- **2,5 jours pour IT Managers**
  - 180 sessions au choix
  - Témoignages utilisateurs
  - Stratégie et vision
  - Entretiens individuels
- **5 jours pour architectes**
  - 650 sessions techniques
  - Sessions "meet with architects"
  - 'Hands-on' labs
  - Tests de certification gratuits
- Inscription:
  - www.ibm.com/events/informationondemand
  - Conditions spéciales via IBM France:
    - Contacter Claude HAIK
    - Tel: 01-49-05-81-22
    - claude_haik@fr.ibm.com

Ticket d'entrée: $1895

© 2006 IBM Corporation