Session: I09

Hitchhikers Guide to Data Replication The Story Continues ...

Michael Paris
Trans Union LLC

May 9, 2007 9:50 a.m. – 10:50 a.m.
Platform: Cross-Platform

May 6-10, 2007
San Jose Convention Center
San Jose, California, USA

IDUG® 2007
North America

GoFurther

TransUnion is a global leader in credit and information management.

For more than 30 years, we have worked with businesses and consumers to gather, analyze and deliver the critical information needed to build strong economies throughout the world.

The result?

Businesses can better manage risk and customer relationships. And consumers can better understand and manage credit so they can achieve their financial goals.

Our dedicated associates support more than 50,000 customers on six continents and more than 500 million consumers worldwide.

The Hitchhiker's Guide to IBM Data Replication

- Rules for successful hitchhiking

  - •DON'T PANIC

    - •Know where your towel is
    - •There is more than meets the eye
    - •Have this guide and supplemental IBM materials at your disposal

GoFurther

2

Here are some additional items to keep in mind besides not smoking (we are forced to put you out before the sprinklers kick in) and silencing your electronic umbilical devices (there is something to be said for the good old days of land lines only and Ma Bell).

But first a definition

Hitchhike:

Pronunciation: 'hich-"hIk

Function: verb

intransitive senses

1 : to travel by securing free rides from passing vehicles

2 : to be carried or transported by chance or unintentionally <destructive insects hitchhiking on ships>

transitive senses : to solicit and obtain (a free ride) especially in a passing vehicle

-hitch·hik·er noun – person that does the above

Opening with the words "Don't Panic" will instill a level of trust that at least by the end of this presentation you will be able to engage in meaningful conversations with your peers and management on the subject of replication. Please do not engage in this topic by the water cooler.

By no means definitive and somewhat humorous, this guide will provide the initiated and uninitiated alike with capabilities and implementations of the products offered by IBM. Further knowledge can be gained from reading the supplements at the end of this guide.

As in "The Hitchhiker's Guide to the Galaxy", the knowledge of 'where one's towel is' is mandatory. For our purposes the use of a towel is two fold, to impress management as in 'That person really knows where their towel is' and but most important to wipe the sweat from the brow when designing an implementation or troubleshooting.

My apologies to Douglas Adams.

The Hitchhiker's Guide to IBM Data Replication

- Topics of Discussion
  - Replication Design in a Federated Environment
  - Factors for replication product usage
  - Implementing WII
  - Monitoring and problems
  - My brain hurts, I'm ready for a Pan Galactic Gargle Blaster

GoFurther

3

This page intentionally left blank.

- IBM Definition:

IBM DB2 Replication (called DataPropagator on some platforms) is a powerful, flexible facility for copying DB2 and/or Informix data from one place to another. The IBM replication solution includes transformation, joining, and the filtering of data. You can move data between different platforms. You can distribute data to many different places or consolidate data in one place from many different places. You can exchange data between systems.

4

GoFurther

This definition was obtained from an IBM website in researching the guide. A lot of words but you get the idea of what the product is all about.

I would have liked it better if the word 'robust' had been included.

Robust

Pronunciation: rO-'bust, 'rO-(")bust

Function: adjective

Etymology: Latin robustus oaken, strong, from robor-, robur oak, strength

1 a : having or exhibiting strength or vigorous health

  b : having or showing vigor, strength, or firmness <a robust debate> <a robust faith>

  c : strongly formed or constructed : STURDY <a robust plastic>

2 : ROUGH, RUDE <stories... laden with robust, down-home imagery -- Playboy>

3 : requiring strength or vigor <robust work>

4 : FULL-BODIED <robust coffee>; also : HEARTY <a robust dinner>

5 : relating to, resembling, or being any of the primitive, relatively large, heavyset hominids (genus Australopithecus and especially A. robustus and A. boisei) characterized especially by heavy molars and small incisors adapted to a vegetarian diet –

Management really likes this word, now I know why.

Replication Design in a Federated Environment

- Hitchhiker's Guide Definition of Replication:
  - It's a bypass
    - Everyone needs a bypass to get from A to B
  - There are 2 major types
    - SQL
    - Q
  - A Bridge to Non DB2
    - WII Replication Edition (aka Data Joiner)
    - Replication Edition NR (non database files)
  - Loosely couple shared data needs
  - This answer is better than 42

GoFurther

5

Replication gives us some things to think about when designing and implementing applications.

•Service Level Agreements – Not being tied to remote calls to other databases that may not be available

•Processing requirements – Data received as the application needs it

•Data Consolidation – One source multiple targets

•Data Transformation – If you can do it an SQL statement you should be able to do it replicator

•Data Subsetting – Vertical, horizontal and views

•Load Balancing – replica or peer to peer

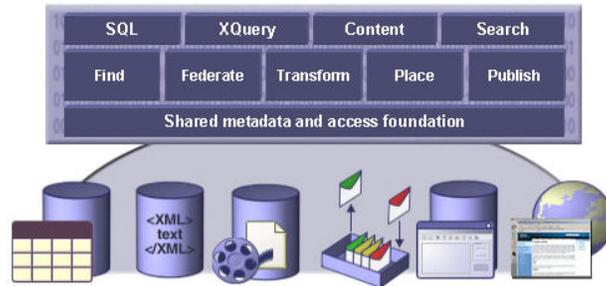•Possible alternative to HADR depending on application data tolerance

The bridge to non-DB2 this presentation will introduce is WII Replication Edition. Our replication bypass must have this bridge in order to transfer data to/from foreign DBMS's.

I've asked around and it really is cool on the resume assuming you know where your towel is during the interview.

For those of you who are not aware, 42 was the answer given by the computer Deep Thought when set on the task to answer 'What is the meaning of life, the universe and everything'. It really does not have anything to do with this presentation.

Replication is part of the SQL, Federate, Transform, Place and Publish methods of Information Integration Portfolio of products.

Wondering why this is part of the publishing aspect? The capture component of Q replication with a couple of internal definition changes can publish data to a queue for consumption by application programs.

IDUG® 2007 North America

Replication Design in a Federated Environment

- **Main Components**
  - •Source Server
    - •Point of origin (A)
  - •Target Server
    - •Point(s) of destination (B)
  - •Control Server
    - •How to get from A to B
  - •Federated Server
    - •The bridge to non-DB2 data

GoFurther

7

This and some following foils will introduce us to some replication speak.

The terms can be a little confusing but are necessary when speaking to a group of UNIX or LAN administrators since they took the word server first and believe it to be a physical machine. These terms should only be used in the presence of other replication administrators after ensuring the meeting room is not bugged, all members have their towels and the secret administrator greeting is given.

Remember that where you find capture running would be a replication change data source. Where I want to put the data through apply, that's a target. The how and when piece (control) could reside either on the source or target server depending on the approach taken to transfer data.

In a federated environment we get a bit fuzzier because capture can only reads db2 logs so a bit of magic is invoked, the Information Integrator server.

Don't panic, the visuals found later in the guide help to sort this out.

Replication Design in a Federated Environment

- Registrations
- Subscriptions
- Subscription Members
- Member Columns
- Nicknames
- Triggers
- Sequence Objects

GoFurther

8

Registrations identify the source tables to the replication product and whether change data scenarios are involved.

Subscriptions are the paths taken by our data from their origination to destination(s).

For a subscription to be built, the administrator needs to register to replication the tables that will be part of our source server. With DB2 replication, the minimum this process entails is turning DATA CAPTURE on for the source table and letting replicator know through GUI generated SQL what objects the capture component should be looking at. There is a bit of a twist to the registration process when dealing with Foreign DBMS's.

Subscriptions can contain a single table source/target, single members, or multiple table source/targets, multiple members, normally tied at the source database with RI or by source application transaction boundary. Mapping of columns and special functions is stored internally by subscription member. As an added bonus, IBM gave the subscriptions the ability to execute pre SQL or post SQL for any special processing you wish to include.

When dealing with foreign databases, Information Integrator is used to identify replication tables through the use of nicknames with the change data tables being populated through triggers and coordinated with the use of sequence objects.

Replication Design in a Federated Environment
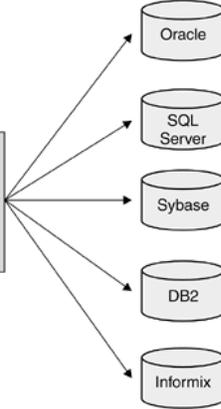
This diagram illustrates the basic use of Information Integrator, single point of federated access to DB2 and non-DB2 databases. This is accomplished through an access method known as wrappers with the tables identified as DB2 nicknames.

Information Integrator takes the SQL in this example and decomposes the query into individual SQL statements for each of the foreign DBMS's. WII will retrieve the targeted table data through the nicknames utilizing the appropriate wrapper. The individual result sets are then combined back at the WII server to resolve the initial query.
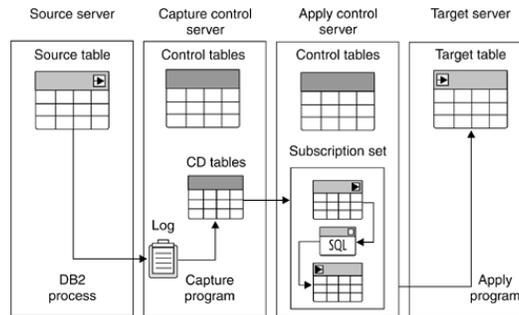
Replication Design in a Federated Environment
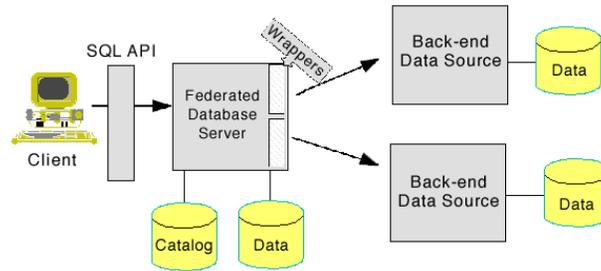
This is a slightly deeper under the cover look at Information integrator.

Wrappers are provided for the databases illustrated in the previous slide with the ability to write custom wrappers for other types of data access. The writing of custom wrappers is outside the scope of this presentation.

It should be noted that runstats should be run for each nickname established for efficient access to foreign data sources whether as a source or target of replication.

Replication Design in a Federated Environment

Source server	Capture control server	Apply control server	Target server
Source table	Control tables	Control tables	Target table

CD tables

Subscription set

SQL

Log

| DB2 process | Capture program | | Apply program |

11

GoFurther

This is a conceptual look of SQL Data replication.

As you can see in the diagram the replication product at the lowest level is comprised of four components. These components are logical concepts not necessarily represented in any real world implementation. The capture server is part of the physical implementation of the source server and the apply server is normally part of the implementation of the target server.

This high level diagram illustrates the process of capturing change data of a source, the basic packaging of a source/target subscription and the communication/control of replication flow.

Replication Design in a Federated Environment

12
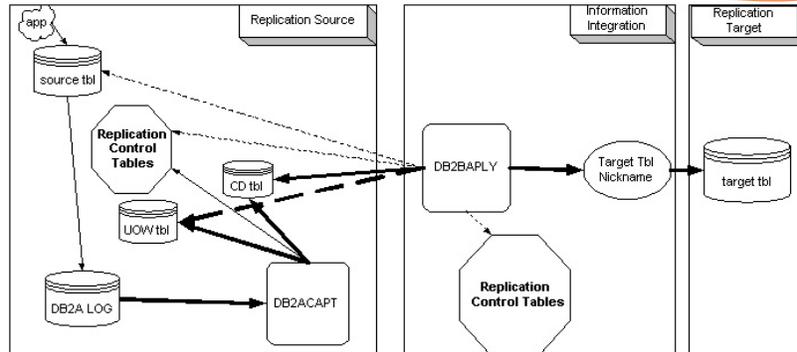
SQL or Q replication can be leveraged onto the WII framework to port data between DB2 and foreign data sources.

Subscriptions are the paths taken by our data from their origination to destination(s).

For a subscription to be built, the administrator needs to register to replication the tables that will be part of our source server. The minimum this process entails is turning DATA CAPTURE on for the source table and letting replicator know through GUI generated SQL what objects it should be looking at.

Subscriptions can be a single table source/target or multiple table source/targets tied by source database RI or by source application transaction boundary. As an added bonus, IBM through in the ability to execute pre SQL or post SQL for any special processing you wish to include.
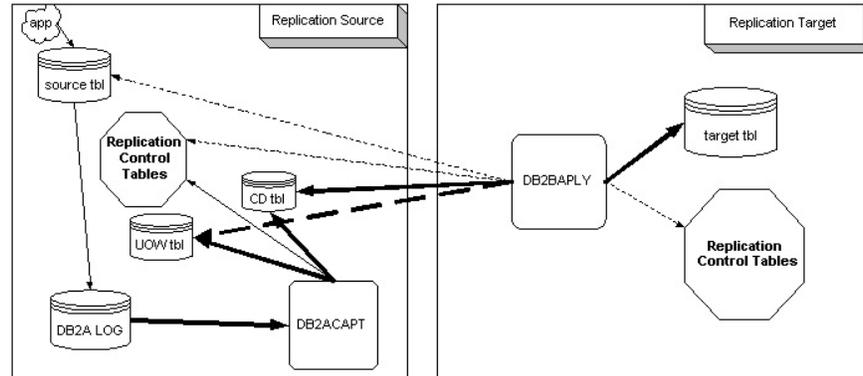
Here we go with some quick definitions found in the dictionary and the guide used to compare the 2 major types of replication products.

Any diametrically opposed critters involving speed would have sufficed but in putting the presentation together the fable of the 'Tortoise and the Hare' came to mind. Might have had something to do with the 'killer rabbit' scene coming on at the same time I hit this foil.

Factors for replication product usage

- Tortoise - SQL

app
source tbl
Replication Source
Replication Control Tables
CD tbl
UOW tbl
DB2A LOG
DB2ACAPT
DB2BAPLY
Replication Target
target tbl
Replication Control Tables

14

GoFurther

The older of the two products, SQL replication relies entirely on the DB2 engine for change data storage and communications.

The capture component reads the DB2 log for changes to registered source tables then writes all changes to the source table's change data table (CD) and also writes rows to the Unit Of Work table (UOW) to indicate committed changes.

Apply cycles at 5 minute intervals checking subscription timing for ripe subscriptions. Once one is found, apply makes a DDF call to the source server and joins the CD and UOW and applies committed transactions to the target. Apply also indicates to what point changes have been applied so CAPTURE can perform CD table maintenance called pruning.

The sequence of events to get data from source to target:

1. Register the source table (data capture on and Change Data table is created)
2. Activate source table registration
3. Start CAPTURE
4. Grant select privileges for the source and CD tables to APPLY application
5. Define CDB communications at target server
6. Create subscription for the target table, columns and the APPLY program to be used
7. Start APPLY
8. Activate subscription
    1. Initial source table full refresh of target table
        1. delete/insert SQL in single UOW or asnload utility or manual refresh
    2. Changes processed at intervals determined by machine clock, event triggering or both
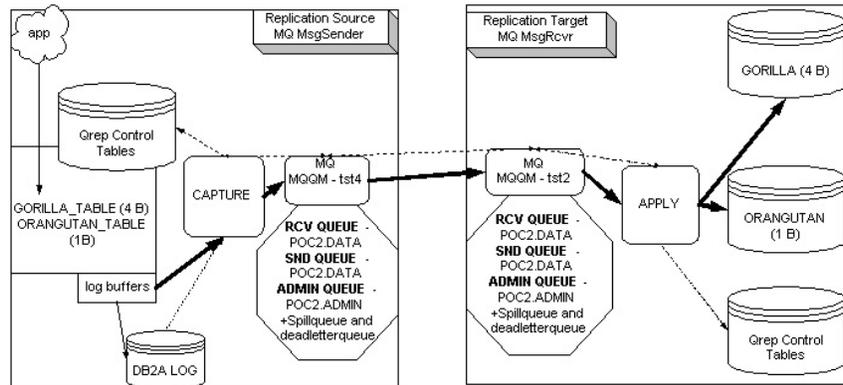
Though not possessing long ears, Q replication answers the need for low latency. This product strips down the DB2 overhead costs by replacing the communications infrastructure with the MQ guaranteed message delivery system. IBM has also incorporated the ability of CAPTURE to read directly from the log buffer when the product is in nominal operation and enhancements to APPLY for subtask parallelism.

Grab your towel – this one is a bit more interesting.

1. Create MQ environment
    1. This step is normally one performed by the MQ administrators. There are steps in tutorials and in the manuals to perform these functions in the non-ZOS environments. This hitchhiker recommends that replication queues not be mixed in with non replication messaging applications.
2. Create the QUEUE_MAP
3. Register the source table/queue_map definition (data capture on and internal definition set through GUI SQL)
4. Create subscription/queue_map definition
    1. At this point you determine whether to auto, manual or ignore refresh the target environment
5. Turn on CAPTURE
6. Turn on APPLY
7. Watch it hum or use your towel
    1. Auto refresh occurs through the spill queue then switches to data queues

Factors for replication product usage

- Mouse – not mentioned but needed
    - any of numerous small rodents (as of the genus *Mus*) with pointed snout, rather small ears, elongated body, and slender tail
    - Hitchhikers guide:
        - The most intelligent creature found on the planet Earth followed by dolphins and humans.
        - **Example: FORT SUMNER, New Mexico (AP) -- A mouse got its revenge against a homeowner who tried to dispose of it in a pile of burning leaves. The blazing creature ran back to the man's house and set it on fire.**

16

GoFurther

Thought we were done with the animals? Not yet.

The mouse in this case is the Websphere Information Integrator (WII) offering for the federated view of databases or information stored outside of DB2. In this framework, communications to or from these sources is done through a concept called wrappers. This concept allows communications to non-DB2 databases or non database data sources.

WII framework enables data propagator to be leveraged into the federated environment.

The following three mice will show you how they run.
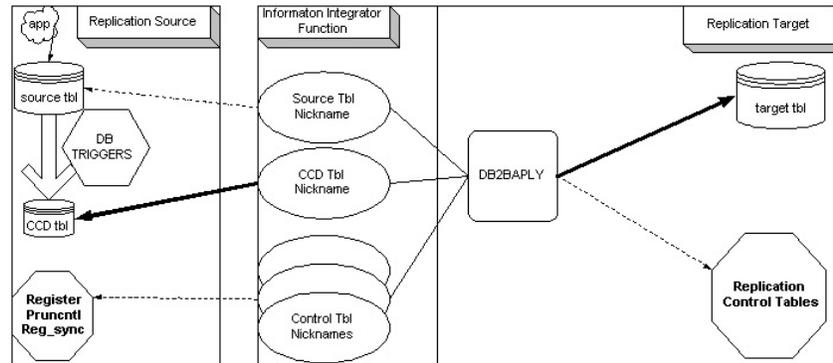
Replication Supported DBMS

INFORMIX

ORACLE

SYBASE

SQL_SERVER

Factors for replication product usage

- Mouse – WII Replication Edition

This is how the setup would appear if we were SQL replicating from an non-DB2 DBMS.

Notice there is no CAPTURE program, APPLY is what makes things happen.

Note the use of database triggers to capture insert/delete/update processing. In this scenario an External incomplete CCD table is used to capture changes and nicknames are used by APPLY to get at the source data, change data and foreign control table information.

Factors for replication product usage
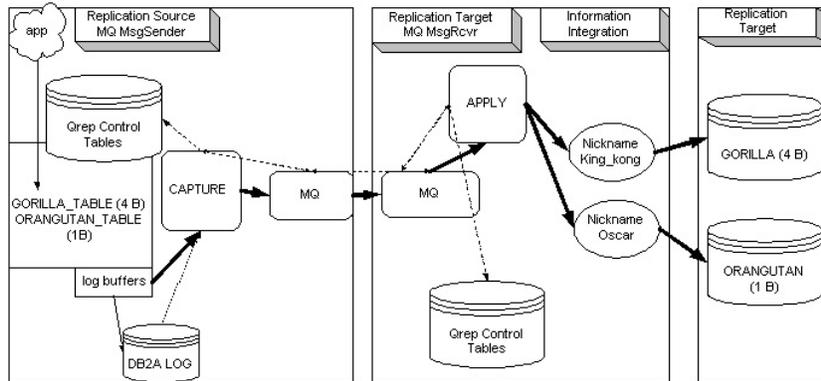
- Mouse – WII Replication Edition

18
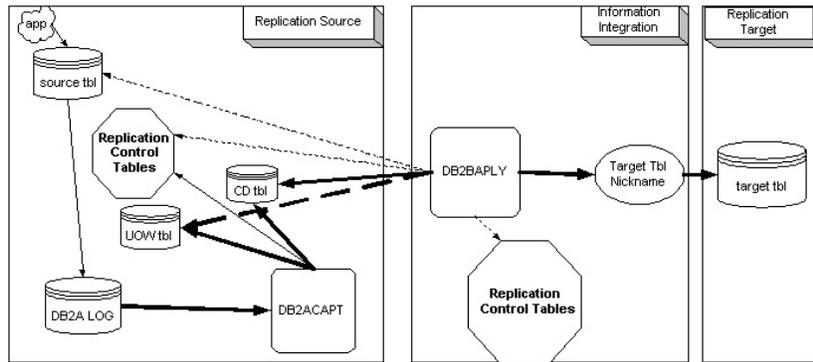
Here is SQL replication in reverse, applying DB2 data to another DBMS

Factors for replication product usage

- Mouse – WII Replication Edition

19

Q Replication for ORACLE and SYBASE targets

Factors for replication product usage

- Data concerns
  - Source Data
    - RI
    - Primary Key Updates
    - Table Changes
  - Database
    - Locking Strategies
    - Recovery Coordination
    - Load Utilities
    - Log Size

20

These are considerations and awareness items that need to be touched on when deciding how to implement and maintain a replication environment.

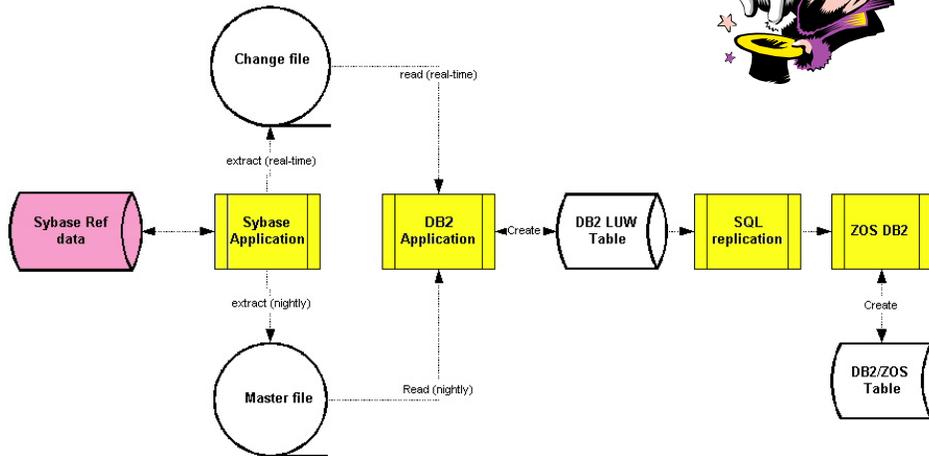Factors for replication product usage

- WII Replication Edition
    - Database trigger tolerance of non-DB2 source database application
    - Nickname maintenance
    - CCD maintenance and availability at the source
    - Security coordination

GoFurther

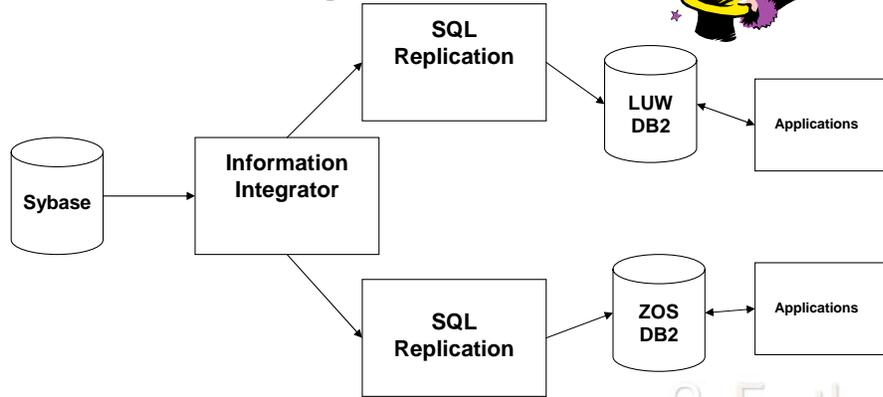**21**

Since SQL and Q rep are carried along by the mouse, the same concerns are carried forward with a few more specifically for running federated.

Implementing WII

Change file

read (real-time)

extract (real-time)

Sybase Ref data

Sybase Application

DB2 Application

Create

DB2 LUW Table

SQL replication

ZOS DB2

extract (nightly)

Create

Master file

Read (nightly)

DB2/ZOS Table

22

Implementing WII

- **Information Integrator**

SQL
Replication

Sybase → Information Integrator

LUW
DB2 → Applications

SQL
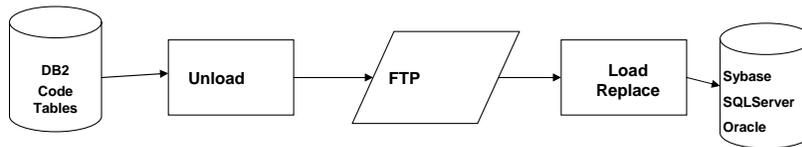Replication → ZOS
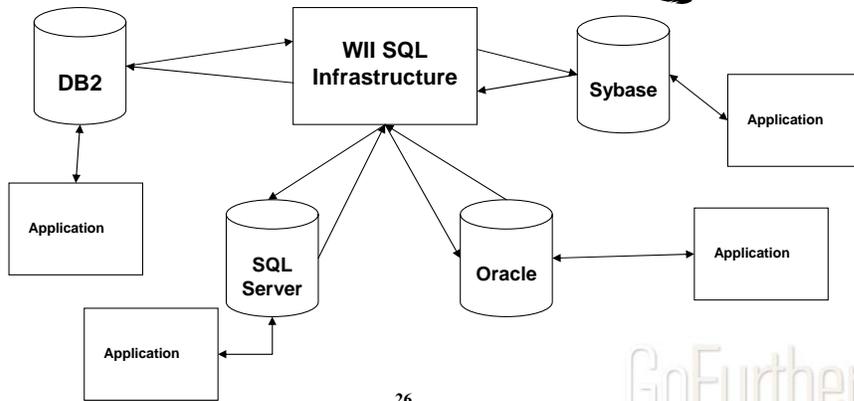DB2 → Applications

GoFurther

23

Implementing WII

Benefits:
•Change data updates via SQL Replication instead of compare process and pre-staging
   •Reduced total application development effort by elimination of application processes
   •Flexible synchronization across DBMS platforms and applications

24

GoFurther

Implementing WII

```
┌─────────┐      ┌──────────┐      ╱─────────╱      ┌──────────┐    ┌─────────────┐
│  DB2    │─────▶│ Unload   │─────▶╱  FTP    ╱─────▶│  Load    │───▶│ Sybase      │
│  Code   │      │          │     ╱         ╱       │ Replace  │    │ SQLServer   │
│  Tables │      └──────────┘    ╱─────────╱        └──────────┘    │ Oracle      │
└─────────┘                                                         └─────────────┘
```

25

GoFurther

Implementing WII

- Information Integrator

26

GoFurther

Implementing WII
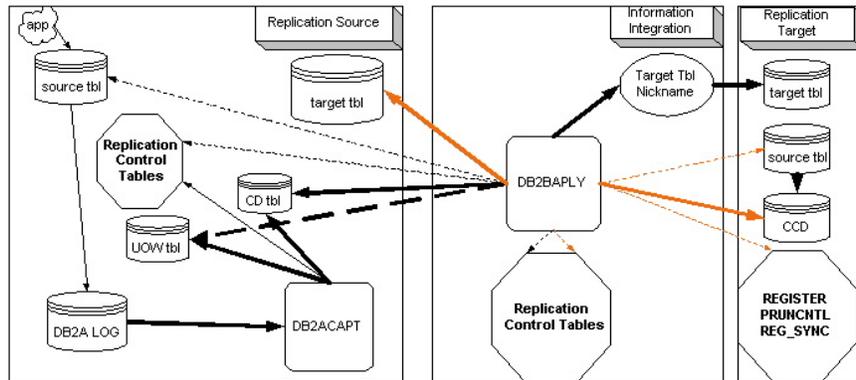
Benefits:

•Change data updates via SQL Replication instead of Unload/FTP/Load

   •No downtime to application, near real time data transfer

   •Flexible synchronization across DBMS platforms and applications

27

GoFurther

Implementing WII

- ## WII replication

28

Break out your towels, we are replicating cross platform, cross database.

Monitoring WII

- Replication Alert Monitor
  - Replication Center GUI
  - ASNMON commands
  - Multiple monitors can be created
    - Capture program (SQL replication)
    - Apply program (SQL replication)
    - Q Capture program (Q replication or event publishing)
    - Q Apply program (Q replication)

GoFurther

29

Break out your towels, we are replicating cross platform, cross database.

Monitoring WII

- Replication Alert Monitor
  - Centralized management
  - Automatic alerts and notifications
  - **Non-DB2 relational databases:** The Replication Alert Monitor does not monitor triggers that are associated with non-DB2 relational databases that are used as sources in a federated database system.

GoFurther

30

Break out your towels, we are replicating cross platform, cross database.

My brain hurts, I'm ready for a Pan Galactic Gargle Blaster

- Summary:
  - Q replication
    - Low Latency
    - Minimal Data transformation
    - One way
  - SQL replication
    - Low volume update transaction/reference tables
    - Control
    - Bidirectional

GoFurther

31

Now that we have traverse the universe of relational replication and have seen some examples of implementations, the summary really speaks for itself.

My brain hurts, I'm ready for a Pan Galactic Gargle Blaster

- Summary:
  - Admin id needs to be able to have create object capabilities in the foreign database
  - Trigger tolerance
  - WII Replication
    - Pass through
    - Replication server
  - Availability of DBMS expertise

GoFurther

32

Now that we have traverse the universe of relational replication and have seen some examples of implementations, the summary really speaks for itself.

My brain hurts, I'm ready for a Pan Galactic Gargle Blaster

- Summary:
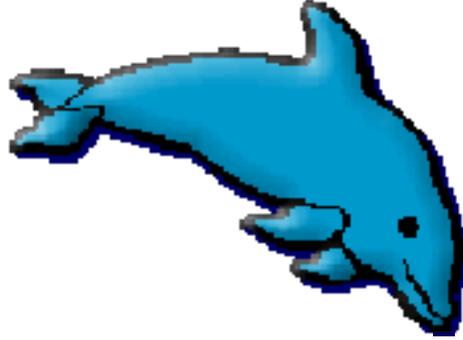 - To be a successful replication hitchhiker:

    ·DON'T PANIC
      - You now have the guide
  - Above all
    - Know where your towel is

33

GoFurther

IDUG® 2007 North America

Questions

- So long and thanks for all the fish

GoFurther

Related guides

- IBM Manuals
  - IBM DB2 Information Integrator SQL Replication Guide and Reference Version 8.2 (SC27-1121-02)
  - IBM DB2 Information Integrator Replication and Event Publishing Guide and Reference (SC18-7568-00)
  - IBM DB2 Information Integrator Federated Systems Guide (SC18-7364-01)
  - IBM DB2 Information Integrator Application Developer's Guide ((SC18-7359-01)
- Redbooks
  - Data Federation with IBM DB2 Information Integrator V8.1
  - A Practical Guide to DB2 UDB Data Replication V8
  - My Mother Thinks I'm a DBA!

GoFurther

35

The Hitchhiker's
Guide to IBM
Data Replication

Michael Paris

- Trans Union LLC
- mparis@transunion.com

36

GoFurther