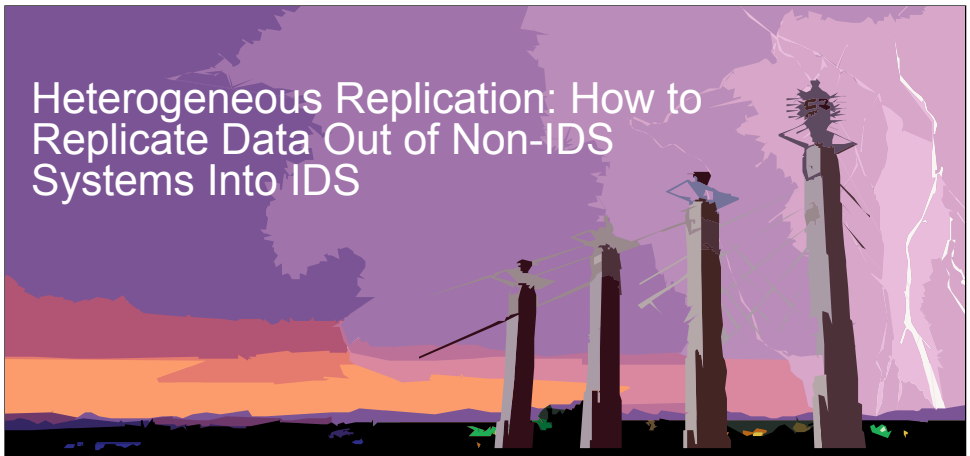


Heterogeneous Replication: How to Replicate Data Out of Non-IDS Systems Into IDS

Stuart Litel, Kazer
Daniel Farrell, Relational Database Systems

Session Code: C05
Monday April 28th 3:30-4:30 PM

2008 IIUG Informix Conference

www.iiug.org

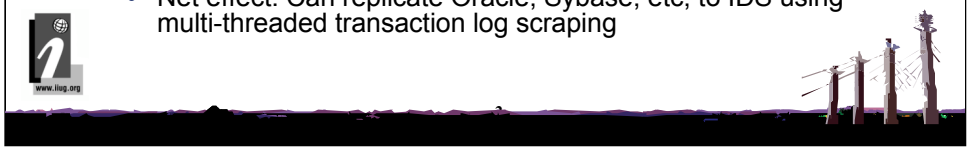
What is IBM Information Server,

(DataMirror Transformation Server Component) ?

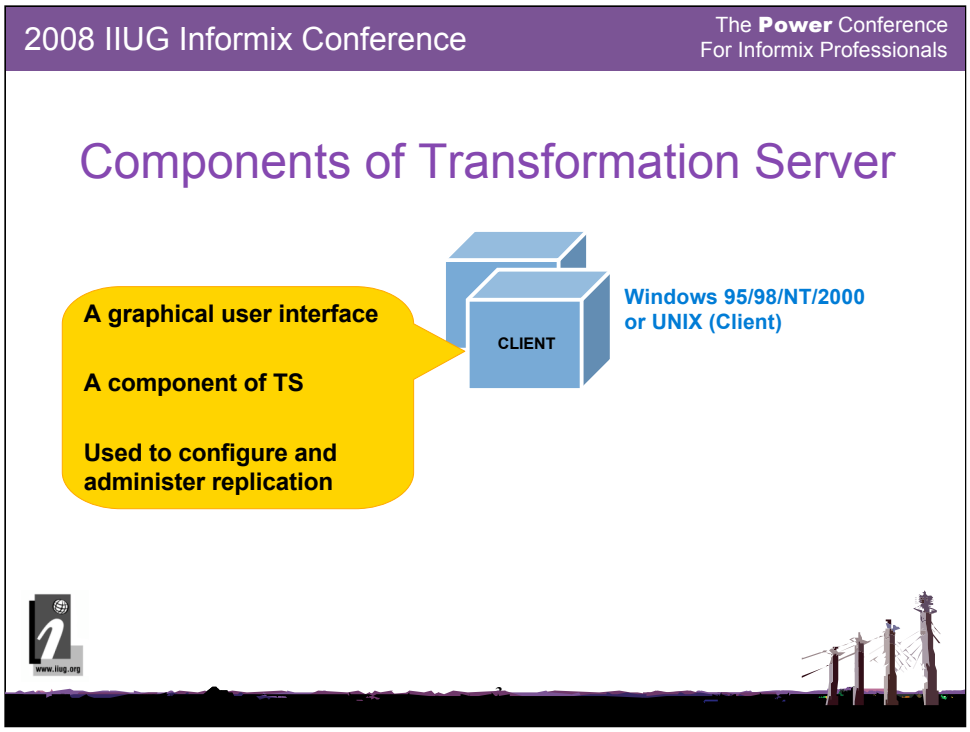
Transformation Server (TS) is a software solution that:

- Connects two or more databases together
- Works on a variety of systems (Windows NT, iSeries, etc.)
- Works with a variety of databases (Oracle, DB2, etc.)
- Captures, transforms and flows the data in real-time
  - Capture: Grab/copy changes to data as the change occurs
  - Transform: Modify the data using filters, calculations or functions
  - Flow: Send the data to another database
  - Real-time: Without any delay, changes are immediately sent
- Can be used in any industry

- Net effect: Can replicate Oracle, Sybase, etc, to IDS using multi-threaded transaction log scraping



Transformation Server is a point to point, peer to peer connection. It can be used in any industry from retail to manufacturing and so on.



This is where you will do most of your work

Configuration of your system

Configured using access managers

Set up agents and users

Components of Transformation Server

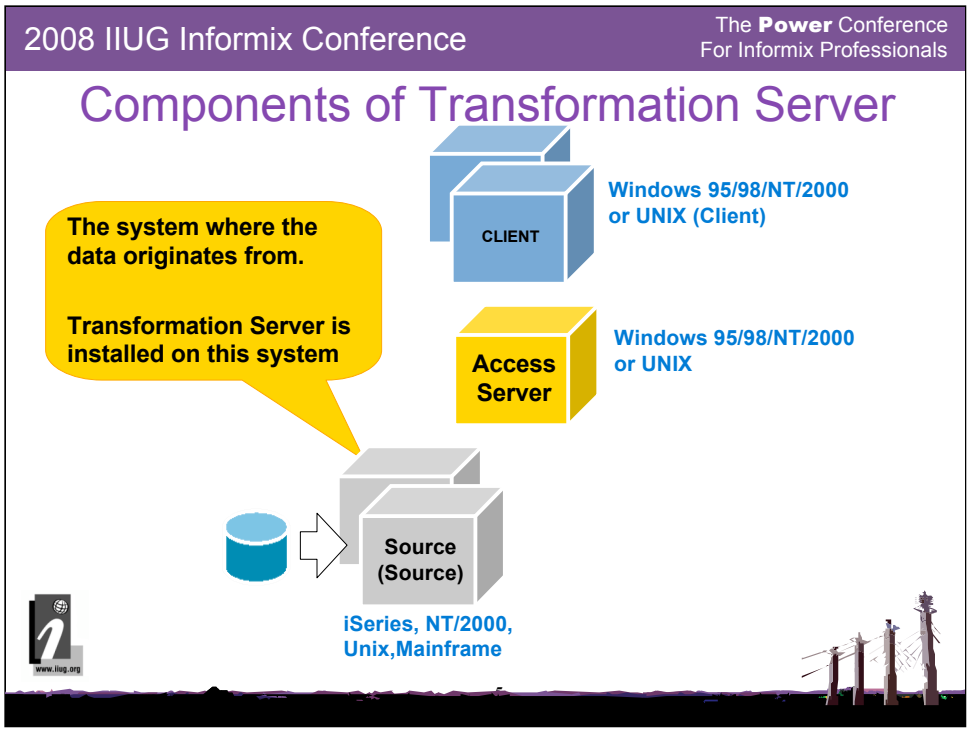
A component of Transformation Server

Enables the user to set up security

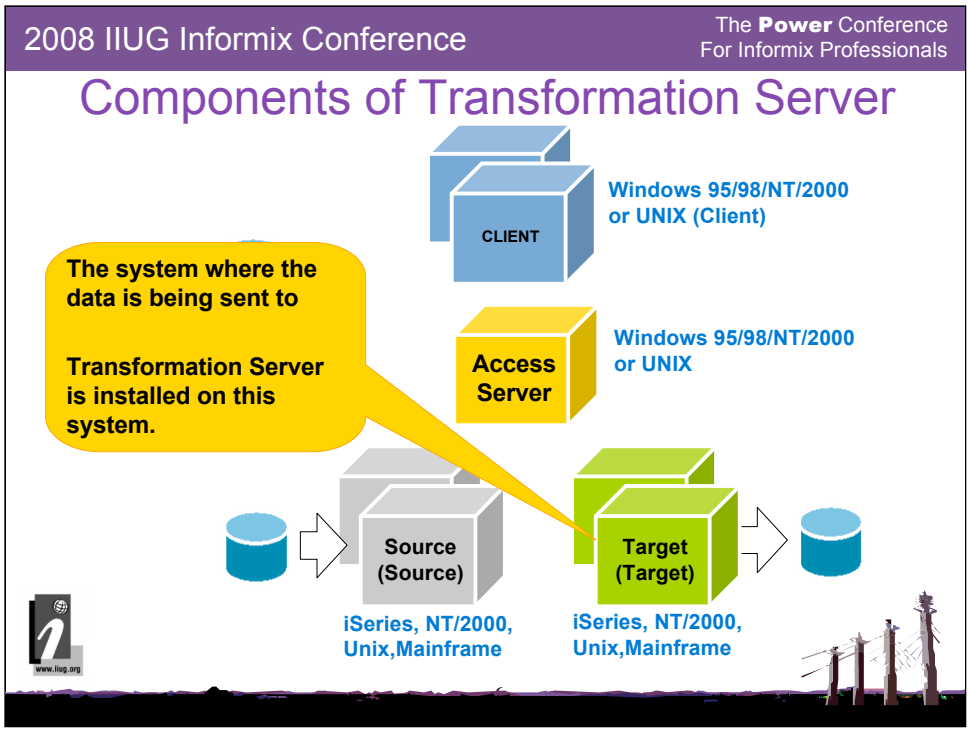
Configured using a GUI (graphical user interface) called the Access Manager

Windows 95/98/NT/2000 or UNIX (Client)

Windows 95/98/NT/2000 or UNIX



Source installation



Target installation

Data replication terminology

- Source
- Target
- Datastore
- Subscription
- “Replication Log”
- MetaData
- Table/File
- Row/Record
- Column/Field

Replication Terminology

Source is known as the source, where the data will come from.

Target is known as the target, where the data will be sent to.

Replication log is a generic term for inserts, updates, deletes for temp storage, journal logs, each platform does it differently but does the same thing.

MetaData stores the entire replication environment

Catalog focuses on what files are being replicated. It’s like making a list.

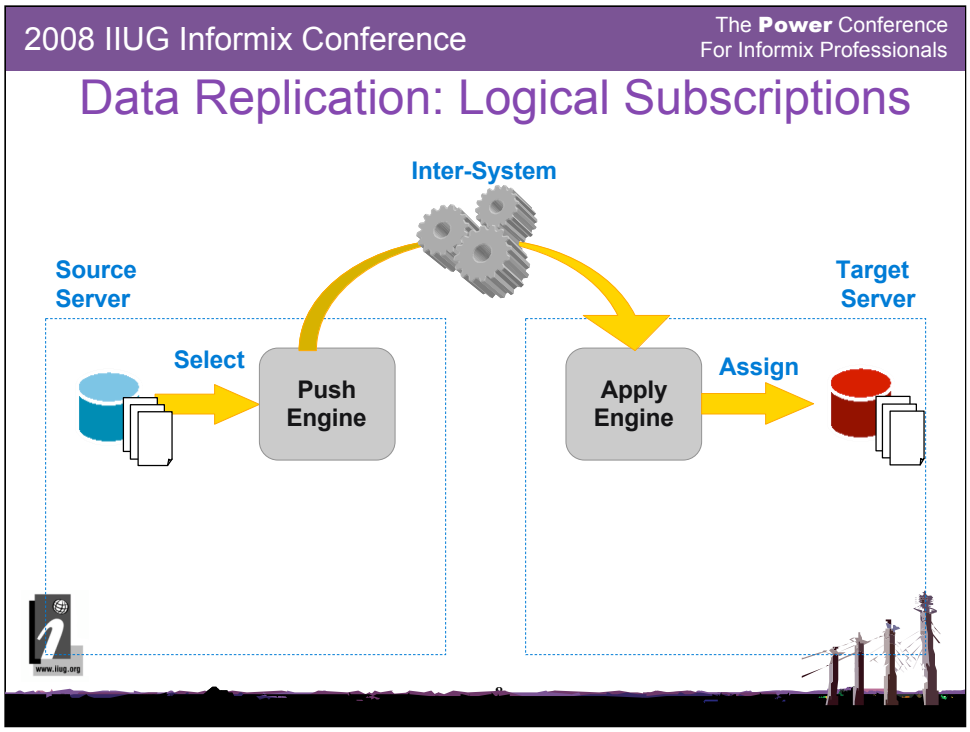
Subscription is how we get data from one system to another.

Table/file

Replication Terminology:

The computing industry has introduced a number of new metaphors.

- The word Source replaces Source.
- Target replaces target.
- The replication log is where the inserts, updates and deletes are stored. This is different on each platform. iSeries uses journals, SQL Server uses the distribution database and Oracle uses a proprietary replication log.
- Metadata is “data about data”, or data about the replication environment. It is a set of tables that will contain all of the details of the replication environment.
- Catalog is where we store all of the details about the files/tables that we will be using for replication (column names, types, whether or not it’s nullable).
- Subscription is what you create in order to point transformation server to another system. By providing information about things such as the hostname, port number, platform type, TS version, we are able to send data from the Source to the Target.
- We replicate at the table level (called a file on iSeries). The term table is used throughout the training materials.

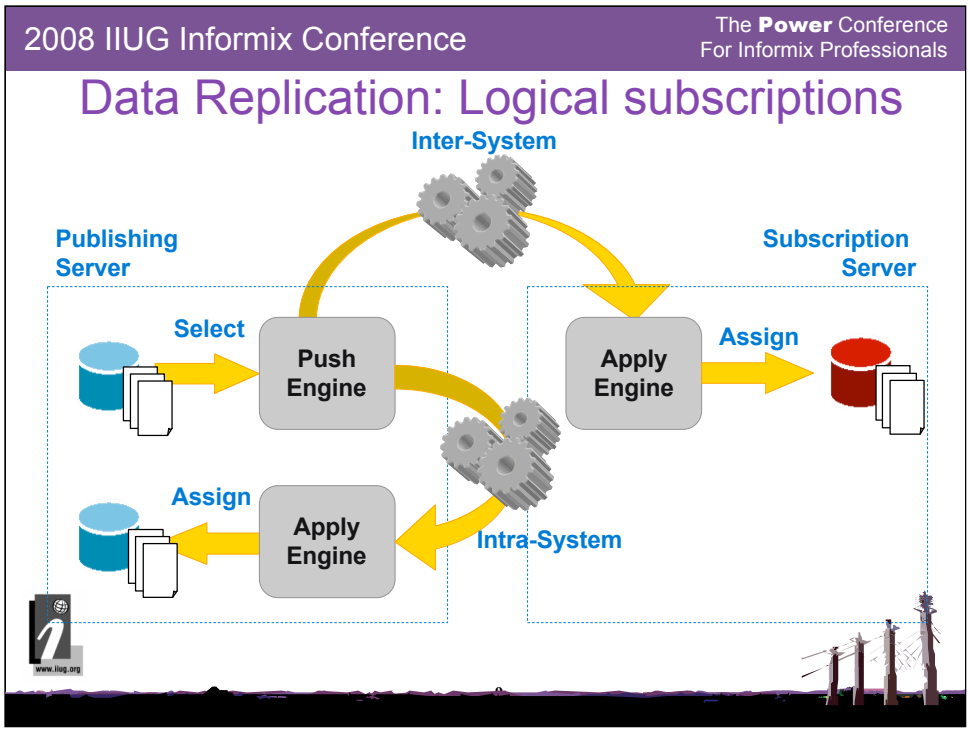


Explain the flow of how TS works.

-Logical subscriptions are used to replicated data.

-TS installed on both the Source and Target

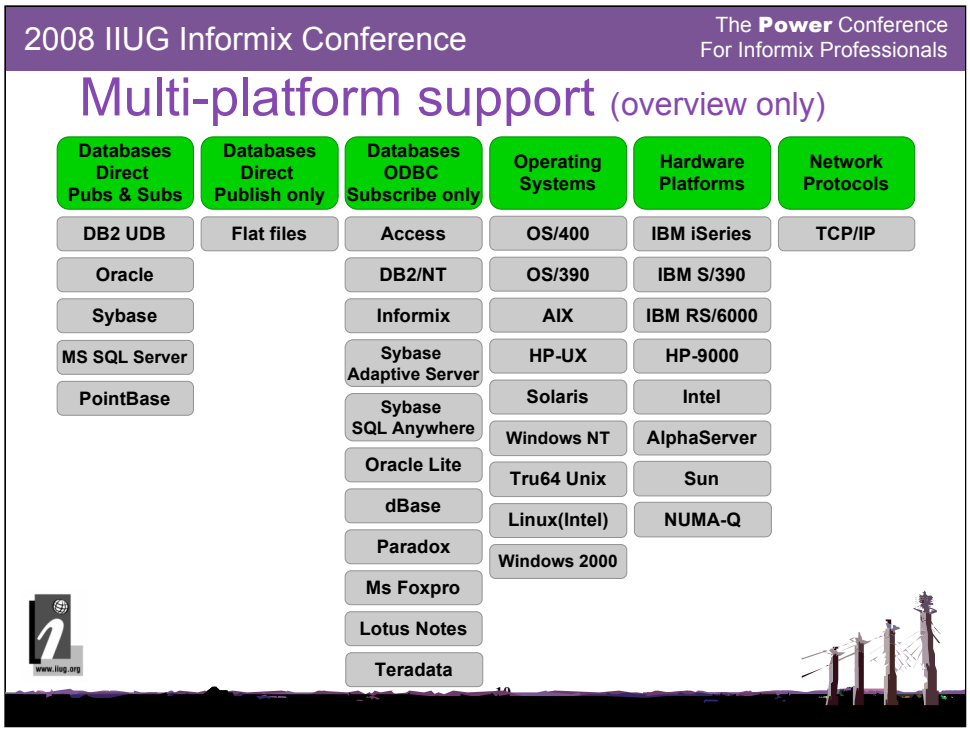
-Data is selected from the source table, pushed to the Target by the push (Source) engine using the information in the subscription and is received by the target apply engine on the Target where the data is placed in the table which we have indicated during the assign.



You can also send data between the Source and Target on the same system.

Potential Questions:

Why would a company want to replicate their data intrasystem? (testing, developers, data marts, data cleansing, back-ups, EAI, etc.)

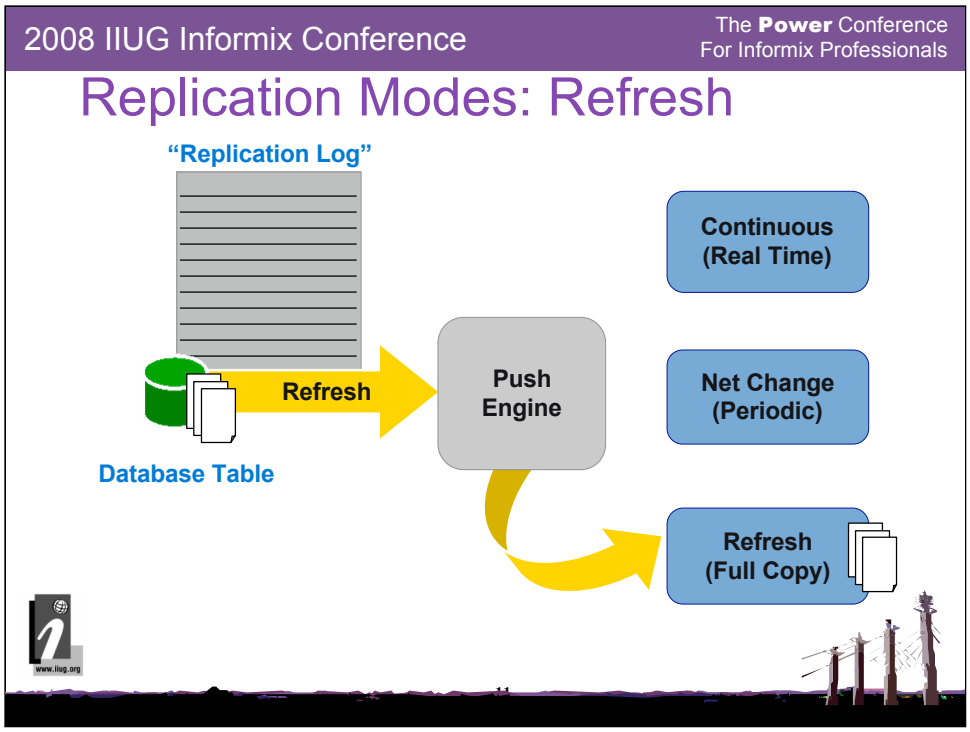


Database Direct replication is a direct connection from one database to another.

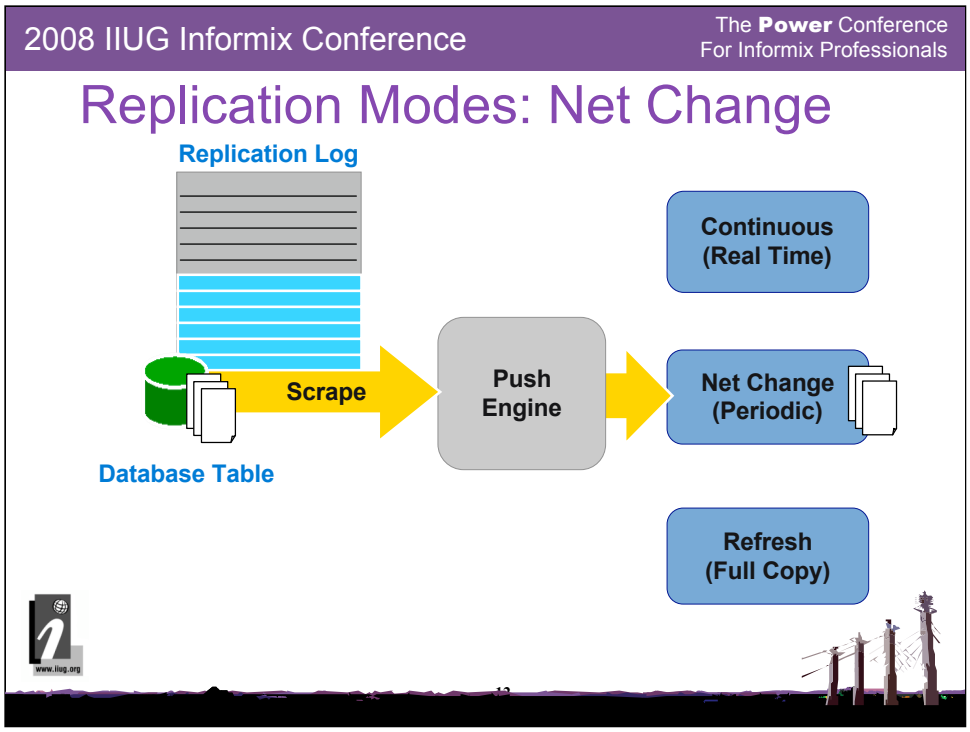
Go from flat files to any type of supported system

Can go to a number of different databases via ODBC, although this is somewhat slower then a direct replication due to the added middle layer.

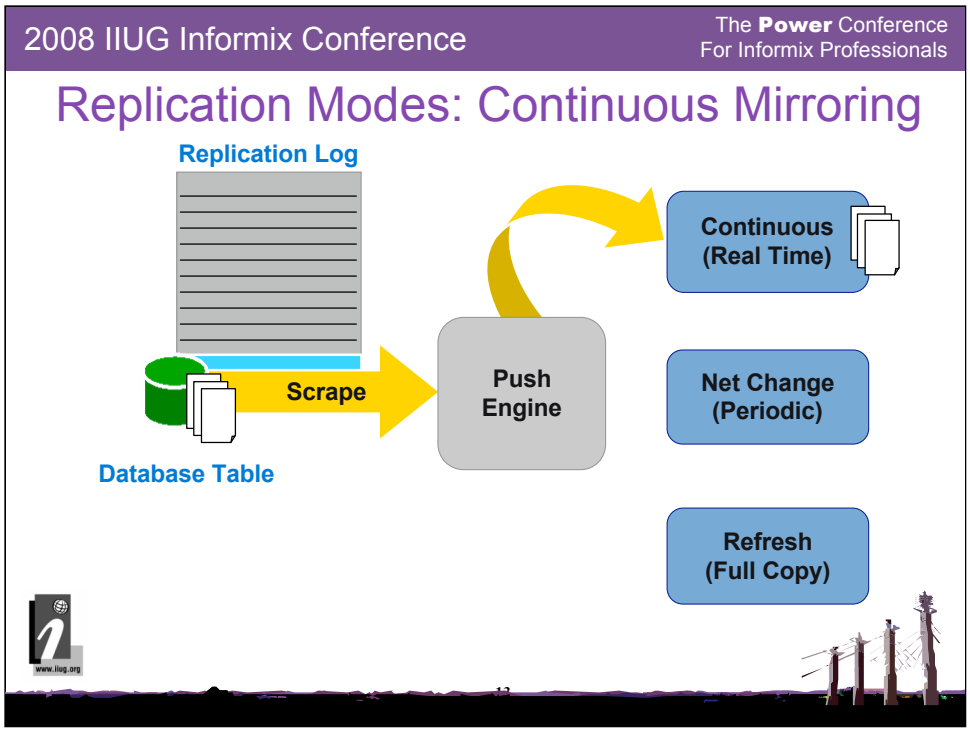
Point out the highlights. I usually just read the headings (e.g. we support a number of operating systems, hardware platforms and network protocols).



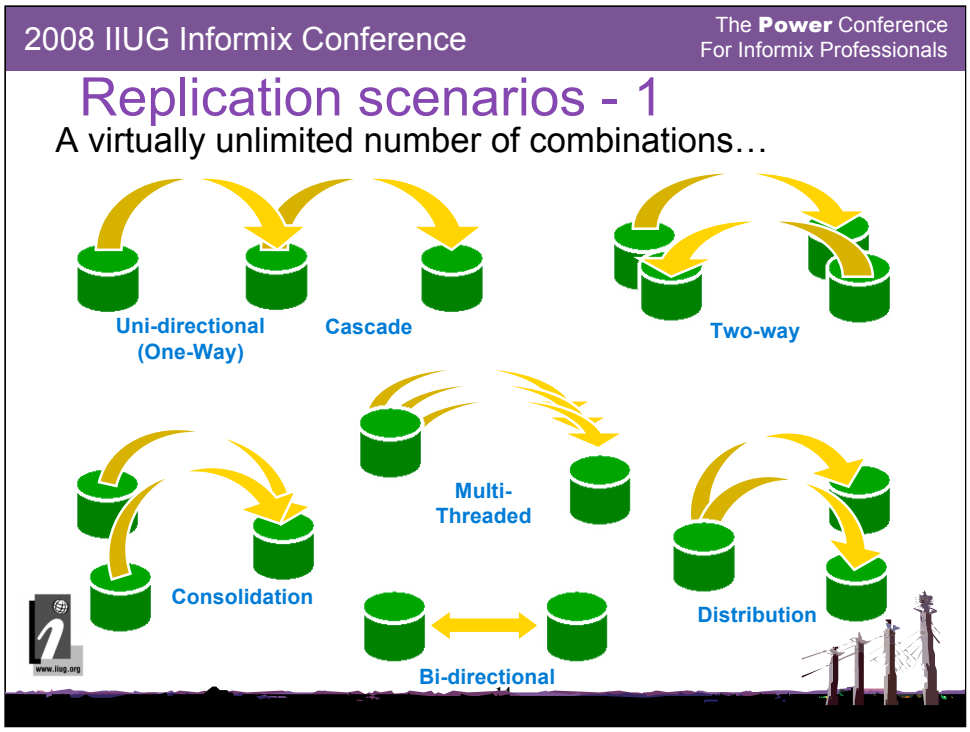
With refresh, you are basically taking a copy of the database on the source and placing it on the Target.



When net change mirroring, the replication log (AS/400 journal, NT distribution database, Oracle proprietary replication log) stores the changes that take place throughout the time period (hour, day, other). At a point in time, replication is started and a scraper scrapes the log in order and pushes it across to the Target installation where the data is applied.



When continuous mirroring, the replication log (AS/400 journal, NT distribution database, Oracle proprietary replication log) stores the changes that take place. A scraper continuously scrapes the log and pushes it across to the Target installation where the data is applied.



There are several ways to set up Transformation Server.

Uni-one direction- is one direction

Cascade- flowing from one system and then continuing to another system

Distributed- take the same table and move it to two different locations

Consolidated- from multiple sources to one source branch offices to head office etc.

Replication Scenarios

There are a number of replication scenarios that can be handled by Transformation Server.

Uni-directional is the movement of data in one direction from the Source to the Target tables.

A **Cascade** occurs when the movement of data continues from the table on the Target to another table.

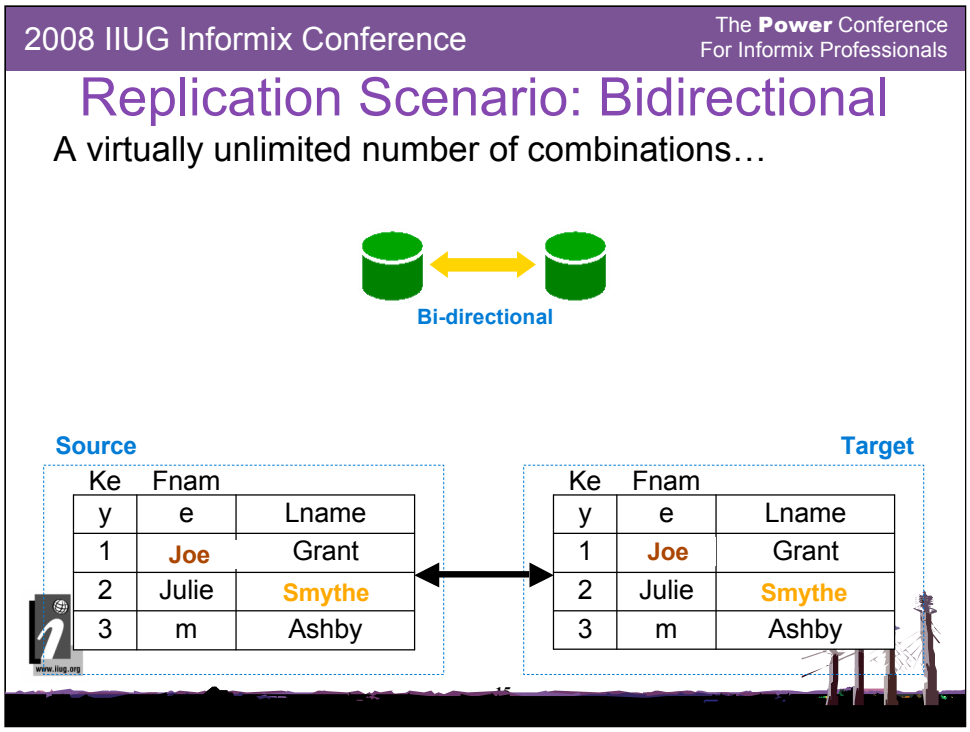
When data is moved from multiple tables into a single table, the data is being **consolidated**.

The opposite of a consolidated replication scenario is a **distributed** scenario. This is the movement of data from a single table into multiple tables.

A **uni-bi-directional** replication scenario is the movement of two (or more) distinct tables between two systems. While one table may move from the Source to the Target, a different table is being replicated from the Target back to the Source.

A **bi-directional** replication scenario moves the same table between a Source and a Target (and back).

Multi-threaded refers the creation of multiple logical subscriptions from the Source to the Target. By doing this, it is possible to improve on data latency.



Uni-bi-directional - has different data going in two different directions

BI-directional - same table going back and forth

Multi-threaded – has multiple subscriptions coming across. Breaks up the processes

Replication Scenarios

There are a number of replication scenarios that can be handled by Transformation Server.

Uni-directional is the movement of data in one direction from the Source to the Target tables.

A **Cascade** occurs when the movement of data continues from the table on the Target to another table.

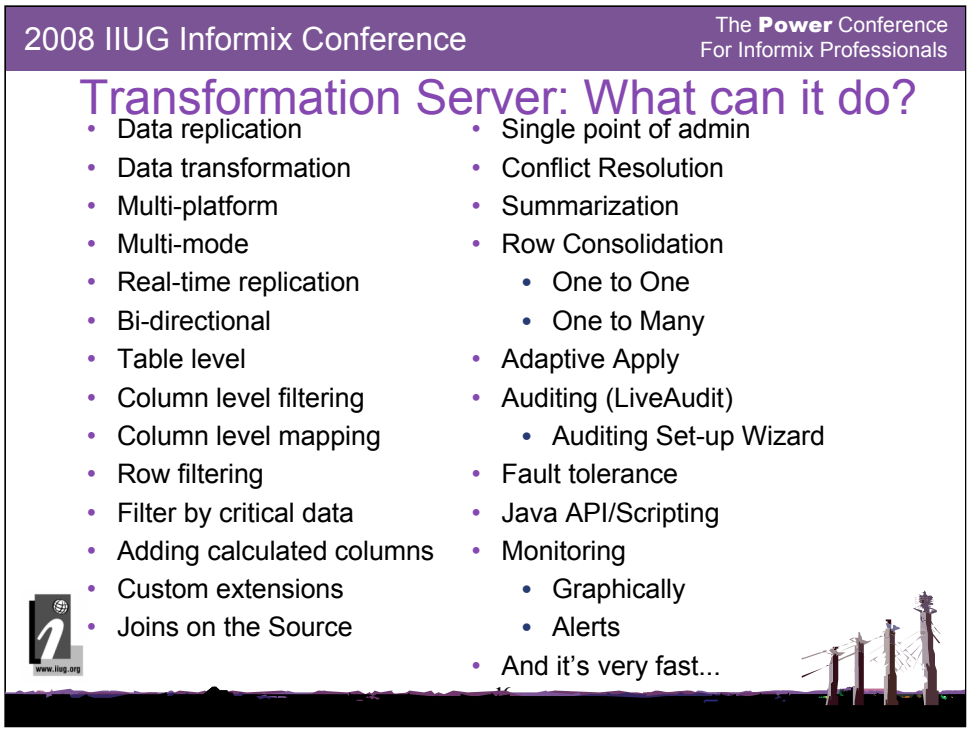
When data is moved from multiple tables into a single table, the data is being **consolidated**.

The opposite of a consolidated replication scenario is a **distributed** scenario. This is the movement of data from a single table into multiple tables.

A **uni-bi-directional** replication scenario is the movement of two (or more) distinct tables between two systems. While one table may move from the Source to the Target, a different table is being replicated from the Target back to the Source.

A **bi-directional** replication scenario moves the same table between a Source and a Target (and back).

Multi-threaded refers the creation of multiple logical subscriptions from the Source to the Target. By doing this, it is possible to improve on data latency.



What can Transformation Server do?

Transformation Server is able to replicate data between databases. **Replication** is the process of creating and managing duplicate versions of a database.

Transformation Server works in **real-time**. There is a sub-one second delay in the movement of data from your publishing database to your subscribing database .

Transformation Server is **multi-platform**. It is able to move data between dissimilar operating systems and databases. For example, you can move data from an DB2 database on an iSeries running OS/400 to a SQL Server database on an server running Windows NT.

Replication can occur using **multiple modes**. Transformation Server can use four different modes to move data:

- Continuous Mirroring, Net Change Mirroring
- Refresh while active
- Refresh while locked

Data can be moved **bi-directionally**. Data from the same table can be moved between two systems, with updates occurring at both ends. For example, the head office may mirror its order table down to the warehouse where the shipping details are input when the order is scanned for shipping purposes. This data is then sent back to head office for use by the sales department to assist with inquiries. When moving data bi-directionally, Transformation Server enforces recursive prevention, ensuring that data will not move in a continuous loop.

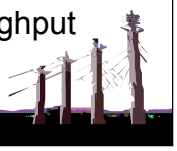
You can **summarize** more than one numeric fields into one numeric fields.

More than one row can be **consolidate** into one as long as row can hold the date (Size)

If row is not exists TS can insert a row and if it is already exits then automatically can update the same row through **Adaptive Apply** functionality. Same behaviour for Update process, if row exits it will perform update operation otherwise it will insert a row.

6.0 Key Features

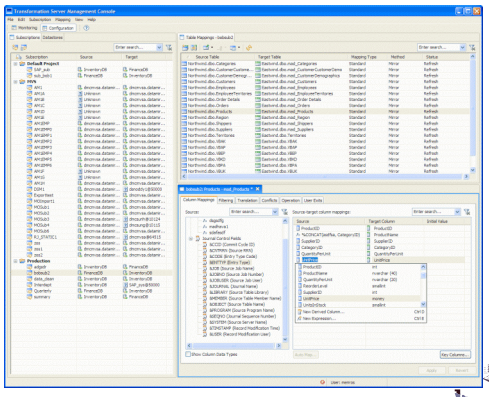
- **Usability – new user interface with:**
  - Monitoring Enhancements
  - Wizards
  - Auto-mapping
  - Change Management
  - Context Sensitive Help
  - Drag and Drop Transformations
- **Character Set Translation**
  - Automatic translations for integrating data in any character set
- **Performance**
  - Reduced source overhead and increased throughput



6.0 Management Console

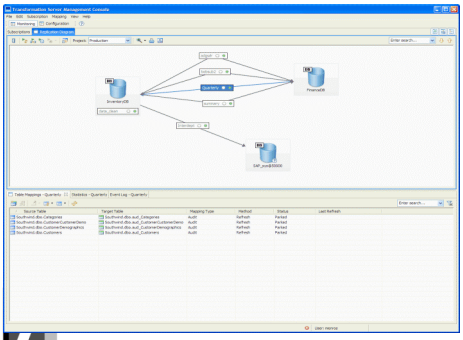
Lower Total Cost of Ownership through Ease-of-Use

- Reduced training costs
- Increased ROI
- Upgrade from Enterprise Administrator with no cost

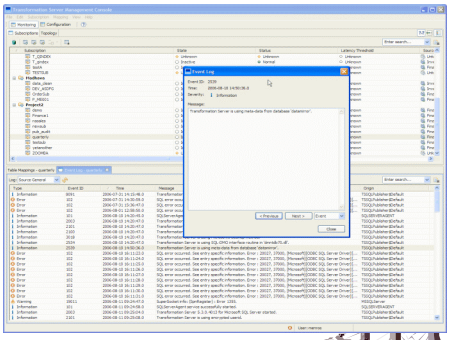


6.0 Monitoring Enhancements

- **Graphical visualization of data integration environment**
  - Detailed monitoring panel provides immediate system statuses
  - Event logs provide in-depth records of all integration processes



Diagrammatic Monitoring View

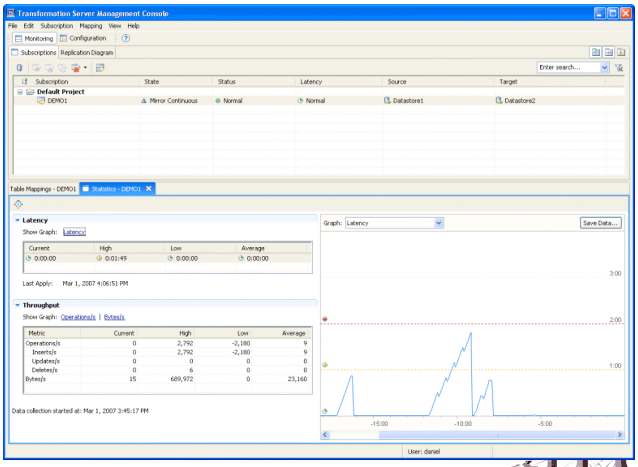


Event Log

6.0 Real-Time Latency Monitoring

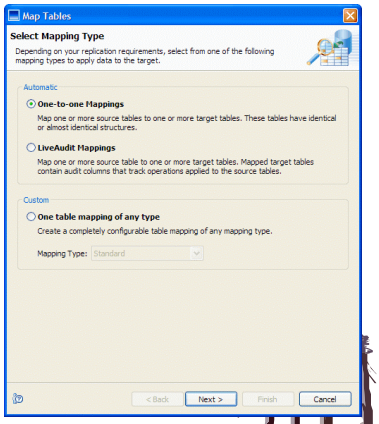
- **Performance reports and records of data integration statistics**

  - Real-time latency reporting and statistics
  - Throughput statistics of both operations and data volume



6.0 Wizards

- **Step-by-step instructions provide fast and easy configuration**
  - Common procedures automated
  - Reduces chances for human error



Feature:

Wizards for automated, step-by-step guidance for common tasks such as creating and configuring subscriptions reduce the learning curve.

Need:

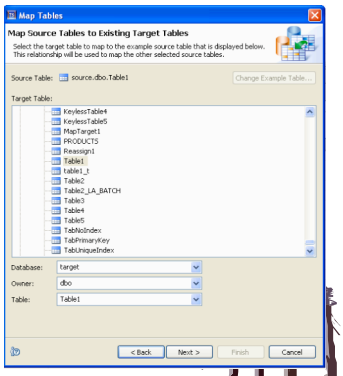
Regardless of the complexity of a customer's replication environment, there will always be integration configuration tasks that are common to all of them. For tasks that are frequently needed, making them easier to execute means that data replication processes can be set up quickly and correctly every time.

Benefit:

Guided procedures minimizes number of user errors and lessens implementation time as users are guided through the steps required for common tasks. The streamlining of the required steps make it faster to create and define subscriptions. This automation of common tasks not only leads to faster deployment time, but also reduces the learning curve – there's no need to have to remember what steps are required to create a subscription, and what the correct order is, or where to do it in the interface – the wizard handles these details so minimal effort is required on the user's behalf. The reduced learning curve results in an increased return on investment – essentially return is maximized while user effort is minimized.

6.0 Auto-Mapping

- **Similar tables and columns are mapped automatically**
  - **Reduces time to create and configure subscriptions**
  - **Mitigates chance of human error**

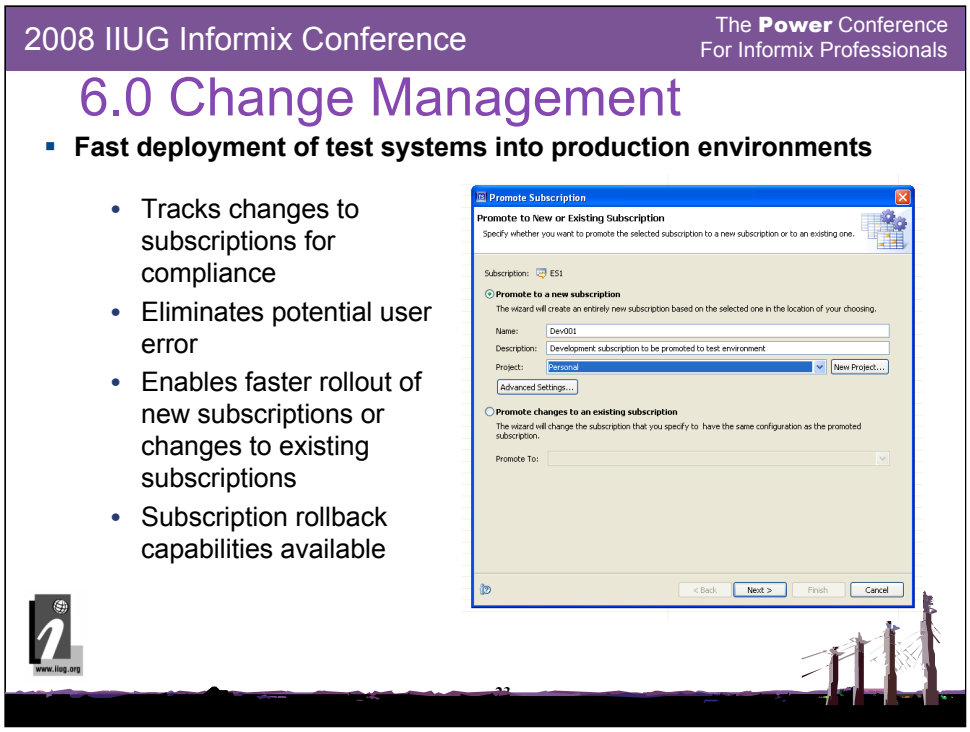


Feature: Auto-mapping is a feature within the table mapping wizard where table and columns are automatically mapped between source and target systems.

Need: So for example, **i**f a customer has 100 tables that need to be mapped from the source system to the target system, only one target table has to be specified and the interface is intelligent enough to use that example to automatically configure the table and column mappings for the remaining 99 tables.

Benefit:

This significantly reduces the amount of time required in configuring subscriptions which increases return on investment and decreases total cost of ownership. The auto-mapping capability also reduces the chance for human error which ensures that your tables are mapped correctly the first time, leading to faster time in achieving data integration to meet the business requirements.



Feature:

Enhanced change management results in faster deployment time since subscriptions, and changes to subscriptions, can be easily promoted from development to test to production environments. Subscriptions can also be exported to XML for integration with source code control systems.

Need:

Data integration configuration details for a replication environment are an important source of information – this is essentially stored in a subscription. Making it easier to manage subscription changes is valuable because you don't have to spend unnecessary time in the maintenance efforts behind manually importing and exporting subscriptions from one environment to another, making copies or backups, or figuring out how to safeguard your subscriptions in a central repository. These are all practical concerns and needs that are met with improved change management.

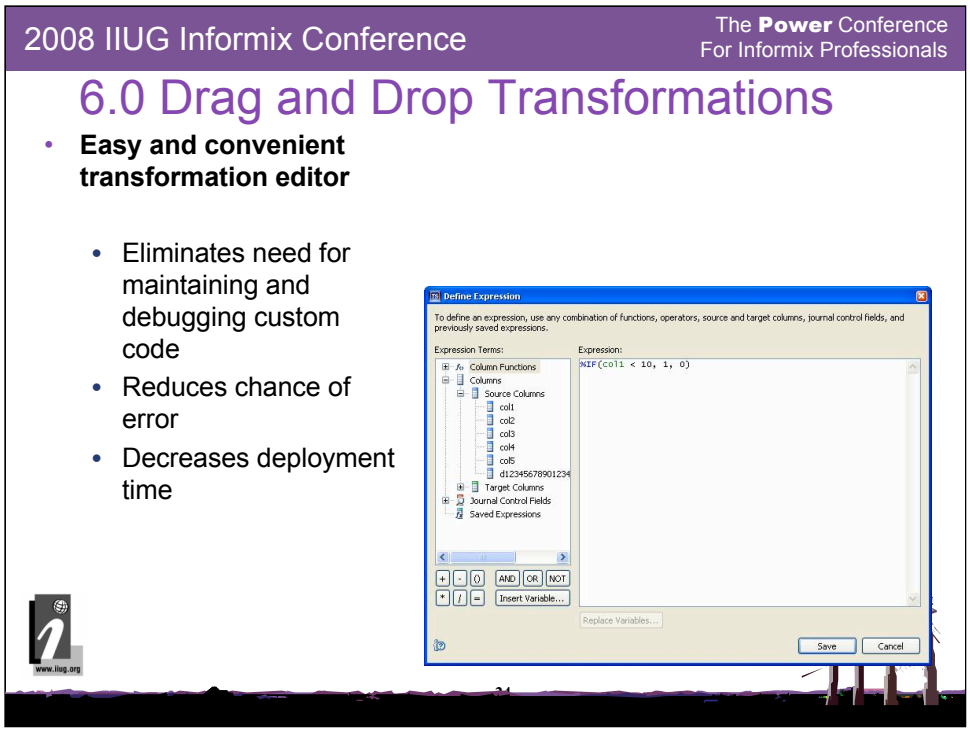
Benefit:

The new user interface ensures all subscription changes can be tracked for compliance purposes (i.e. keeping backup and records of current and past configurations on file)

as well as eliminating potential user error when migrating subscriptions into production environments and supports full test cycles which should always be followed.

Being able to easily migrate subscriptions from development, to test to production reduces both environmental complexity and deployment time, resulting in the faster development, test and production rollout of new subscriptions or changes to existing subscriptions.

Enhanced change management also allows for the immediate roll-back of changes to subscriptions if necessary.



Feature:

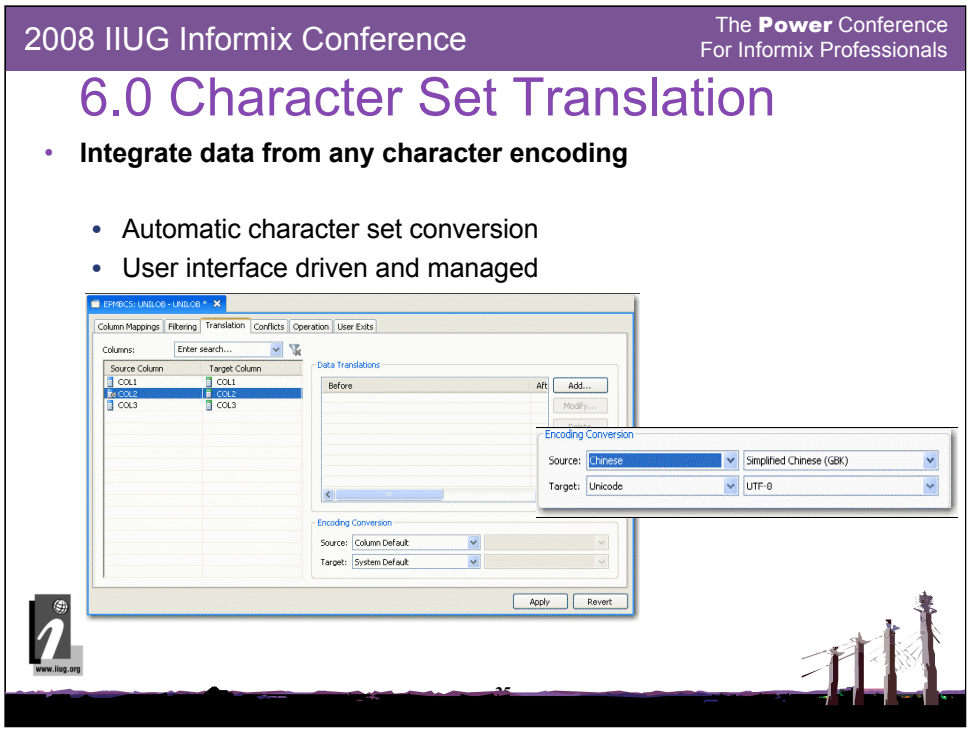
The ability to drag and drop when configuring transformations make it easier and quicker to specify the necessary operations for the data to meet the business need.

Need:

Transforming data is almost always a requirement when integrating data. Transformations such as derived expressions, row and column filtering, summarization, row consolidation or value translation are used to ensure the data in the target systems meets the business requirement.

Benefit:

Drag and drop transformations reduce or eliminates the need to write custom code that needs to be maintained and debugged, this reduces errors and quickens implementation time. Having transformations configured through a drag and drop mechanism makes it easier to use and understand how to perform the transformations the business need requires. This allows users to concentrate on core job functions, not writing code to transform data to meet business requirements. Also means meeting the business requirement does not require custom development that must be maintained and managed.



Feature:

With enhanced character set translation, all necessary data conversions between different character encodings are handled directly within the product.

The available character sets and encodings for mapping between source and target systems are accessible through the user interface.

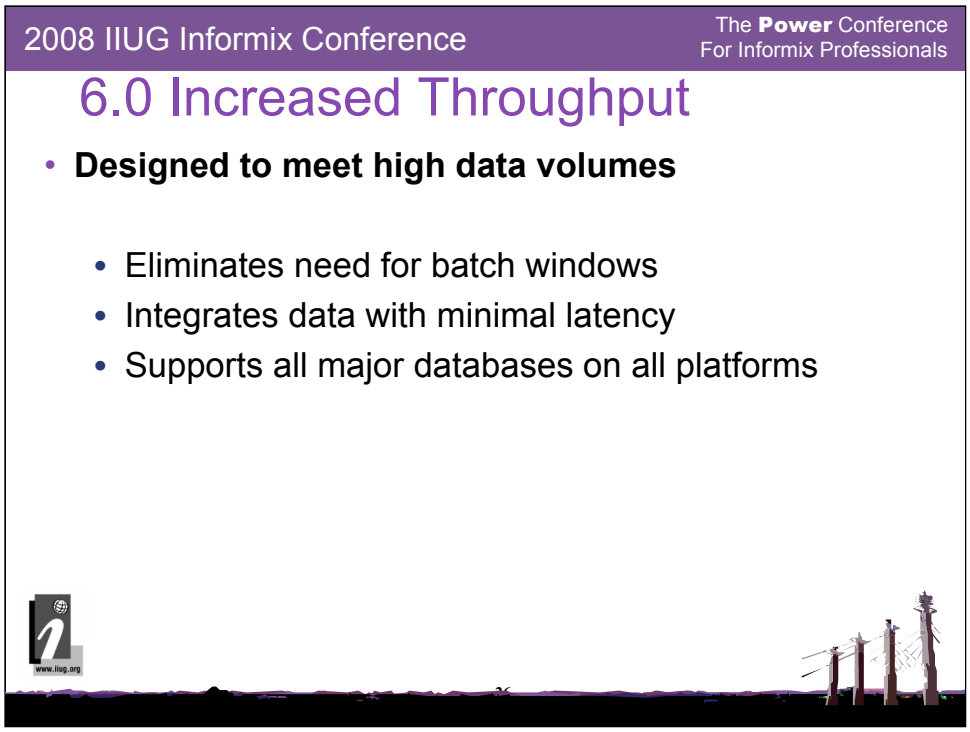
Need:

Companies that operate globally deal with data in multiple languages, which means that their data is encoded in different character sets; this data needs to be integrated just like any other data, however, there are subtle, but important technical intricacies that need to be handled during character set conversion.

Benefit:

Enhanced character set translation allows for the automatic conversion of data between character sets. There's no need for writing custom procedures for handling data conversion intricacies which eliminates the chance for error and reduces the amount of custom code to maintain. This simplifies the data conversion process to provide seamless transformation of data in multiple character sets.

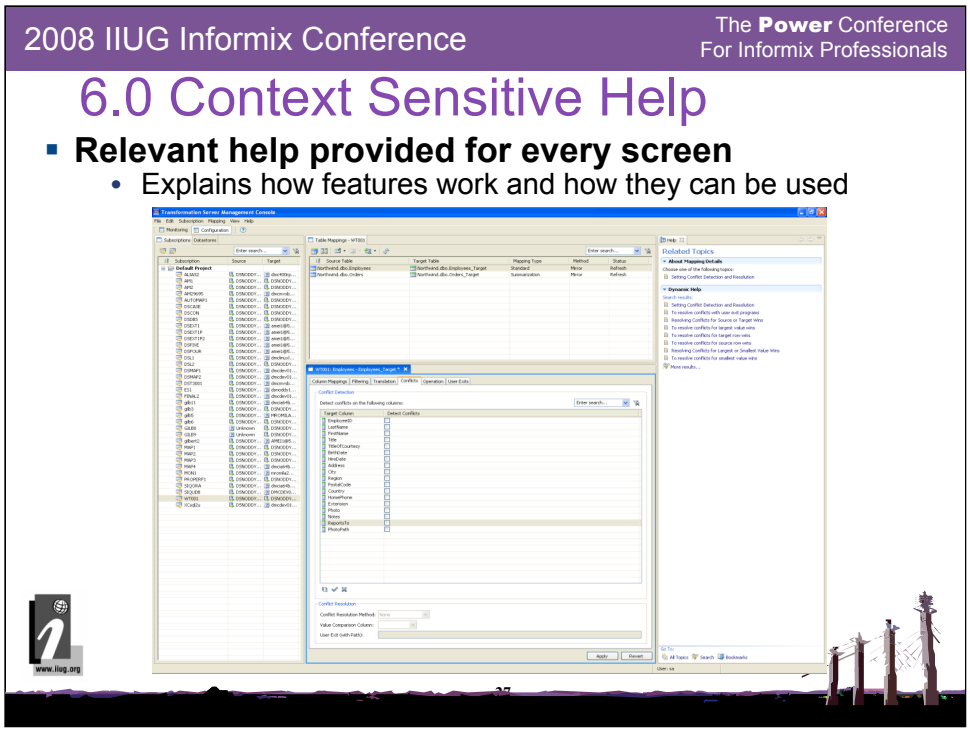
Now that we've covered usability and highlighted the key features of our new interface, let's move to the next key feature – performance.



Need: The need for data is only increasing, which means it must be readily delivered to the business in order for it to sustain or ideally surpass its position in today's competitive landscape. Businesses also have the flexibility to choose technology that meets their unique needs which often, if not always, results in a heterogeneous environment which adds an additional element of complexity to the replication technology since data must be integrated between a wide range of disparate systems.

Benefit:

Transformation Server has been tailored for heterogeneous environments - with its broad support for databases and platforms, it can meet expanding business needs and growing data volumes. As the business demands more fresh data, Transformation Server's increased throughput can meet the demands regardless of the complexity associated with heterogeneous environments. This eliminates the need for a batch window which is unacceptable for enterprise that rely on the availability of up-to-date data for your business needs.



Feature: Online help contents are context sensitive, so information that is relevant to the current task in the console is presented, helping users resolve their problem more quickly.

Need: When users are faced with a question on how to use the tool to do a certain task, they need a way to find the answer – this is where the usual online help comes in since it provides additional information to resolve it, but searching through the knowledge base to and pick out what's relevant to the problem is still required. **Context-sensitive** help goes the extra mile by pre-filtering the masses of information to present you with relevant information to help users with their current task.

Benefit:

Immediate presentation of relevant help information for each screen or feature provides the right information to help users understand how a feature works and how it can be used. Since users can learn about features relevant to their task and apply their new-found knowledge immediately, this not only increases productivity, but also helps in knowledge retention.

Demonstration ..

Session ####
Session Title

Stuart Litel, Kazer
Daniel Farrell, Relational Database Systems