Problem Statement

Users can not visualize complex data integration (replication/propagation or manipulation of data to/from heterogeneous database sources and targets) networks without aids such as diagrams. A simple single direction, two-node network can be easily understood without explicit diagrams or displays. An explicit diagram is required to aid user comprehension once the complexity of the network goes beyond a certain trivial level. The following attributes quickly surpass a normal user's ability to comprehend a data integration network:

a) Two way replication – source ↔ target
b) Cascade replication – source ↔ target 1 ↔ target 2
c) One to many – source ↔ targets $_{(1\rightarrow n)}$
d) Many to one – sources $_{(1\rightarrow n)}$ ↔ target
e) Many nodes (more than 5)
f) Replication schemes which involve a subset of the nodes
g) Different types of nodes

A set of related diagrams is required that will allow the user to break the network up into manageable portions. The diagrams must accommodate all of the attributes listed above. This report explores what would be the requirements (solutions) to fulfill them.

Discussion/Solution

- **Diagram Usability**

Automatically generated diagrams of complex networks normally are not directly useable by humans. Customization of the diagram is required in order to make it comprehensible. Since we are considering non-trivial networks, this requires the ability to generate user-defined diagrams.

User generated diagrams imply the following requirements:

a) The ability to select nodes to be included in a diagram. Based upon the diagram being created, a node may appear more than once on a diagram. NOTE: The definition of what a diagram node represents will depend upon the type of diagram being created.

b) The ability to place nodes at a specific location on a diagram. For example: Users often arrange telecommunication diagrams to correspond to the geographical location of the network nodes.

c) The ability to create, retrieve, modify and delete a diagram definition from a centralized repository.

d) "Auto draw" functionality to help the user create a diagram using "Replication Domain Knowledge". Actions such as connecting nodes automatically, providing pick lists of entities (computer names) to add to a diagram and other mechanisms are required.

e) Ability to assign access rights to diagrams and have access rights associated with what nodes a user has access to build views

- **Navigation**

A complex network must be broken down into a number of simple diagrams. A method of inter-diagram navigation is required in order to allow the various portions of the network to be placed into context with each other.

The user must be able to selectively navigate from one diagram to another. The following navigation abilities are required:

a) Drill down/up: The ability to go from a high level diagram to a more detailed low level diagram and return back.

b) Drill across: The ability to navigate to a "peer" diagram. This implies that the navigation hierarchy is not strictly a top down tree based scheme. It may contain loops and arbitrary connections.

c) User defined diagrams implies the ability to have user defined navigation links.

d) Multiple links: The ability to have multiple different navigation links on any diagram.

e) The ability to save defined navigational links and have them associated with saved diagrams. The navigational links are an attribute of the diagrams.

- **Operational Status Monitoring**

The user needs the ability to quickly determine the current operational status of a data integration network in order to verify that the product is operating. This may be the complete data integration network or a user defined subset of the data integration network.

The monitor must indicate to the user if there is any operation going on that takes a lengthy amount of time (anything that may look as a stall situation, but is not), and indicate if there is a problem, thereby prompting the user to look into a log.

As such, the following operational statuses are of interest to the user:

a) Inactive: Replication is inactive and was ended normally.

b) Active: Replication is operating with in normal parameters. Represents that there are no (detected!) problems with data movement and transformation.

c) Idle: replication is active, but no data is being moved between tables.

d) Error: Represents a fatal problem with data movement and transformation.

e) Starting / Stopping: Replication between hosts is either starting or stopping.
This is only of interest if starting and stopping take a long period of time. Otherwise it is a transient state that the user will not normally see.

- **Graphical Displays**

Querying a source table and target table is not an activity that is sufficiently 'dazzling' for training purposes. All graphical displays need to address the following issues:

Color Independence: Graphics must be designed such that they can be displayed on monochrome CRTs, and viewed by persons with colour-blindness, while still conveying full configuration and status information.

Status Legible from a distance: Graphics must be designed such that a person standing within a reasonable distance away (5-8 meters, on a 20 inch screen) can see the status without undue effort. It is not expected that detailed configuration information will be viewable at a distance.

Multi Screen Capability: Must be able to display multiple graphics displays within one CRT, or different graphics displays on multiple CRTs. An implicit requirement is to allow the graphics in a view that can be to re-sized to scale to the current physical size of the display window. This automatic scaling will allow the user to select the size of window appropriate to view multiple displays on a single CRT.

Product Configuration: A desirable capability is to be able to graphically configure replication. By extension, the user could also operate the product in a graphical manner.

Critical Events: A critical event for a monitor is an event that the user wishes to be notified about, and would visually change on the monitor screens. There is a requirement for the user to be able to define critical events, and to be notified of critical events. The following are the requirements for critical events:

a) Notification of critical events should go through alert and alarm mechanisms.
b) There needs to be a mechanism to be able to define threshold values.
c) They should be customer configurable via a list of all events that can be sent to the GUI monitor.

d) Need a mechanism to be able to see historical critical events

Conclusions:

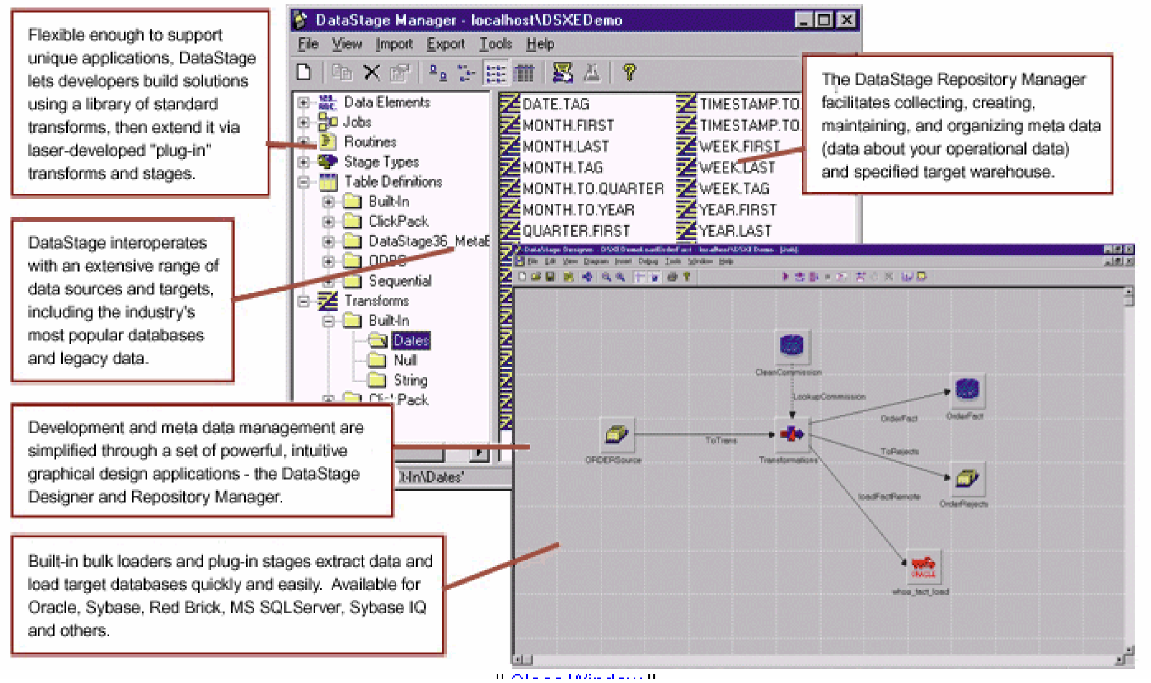
The implementation of the above suggestions would allow that:

a) Users can better visualize complex data integration networks
b) Users can better determine the current replication status of replication
c) Users can better isolate and troubleshoot data integration networks
d) Users can better control non-trivial data integration networks

Appended are two pictorial views for comparison between what is currently in use and what is suggested is this report. The first one resembles the existing MS-Windows-Explorer-like graphical user interface. The second is a simplified impression of how the proposed graphical user interface will look like.

References:

Software Engineering: A Practitioner's Approach, Roger Pressman
Handbook of Usability Testing, Jefferey Rubin
Company's Glossary



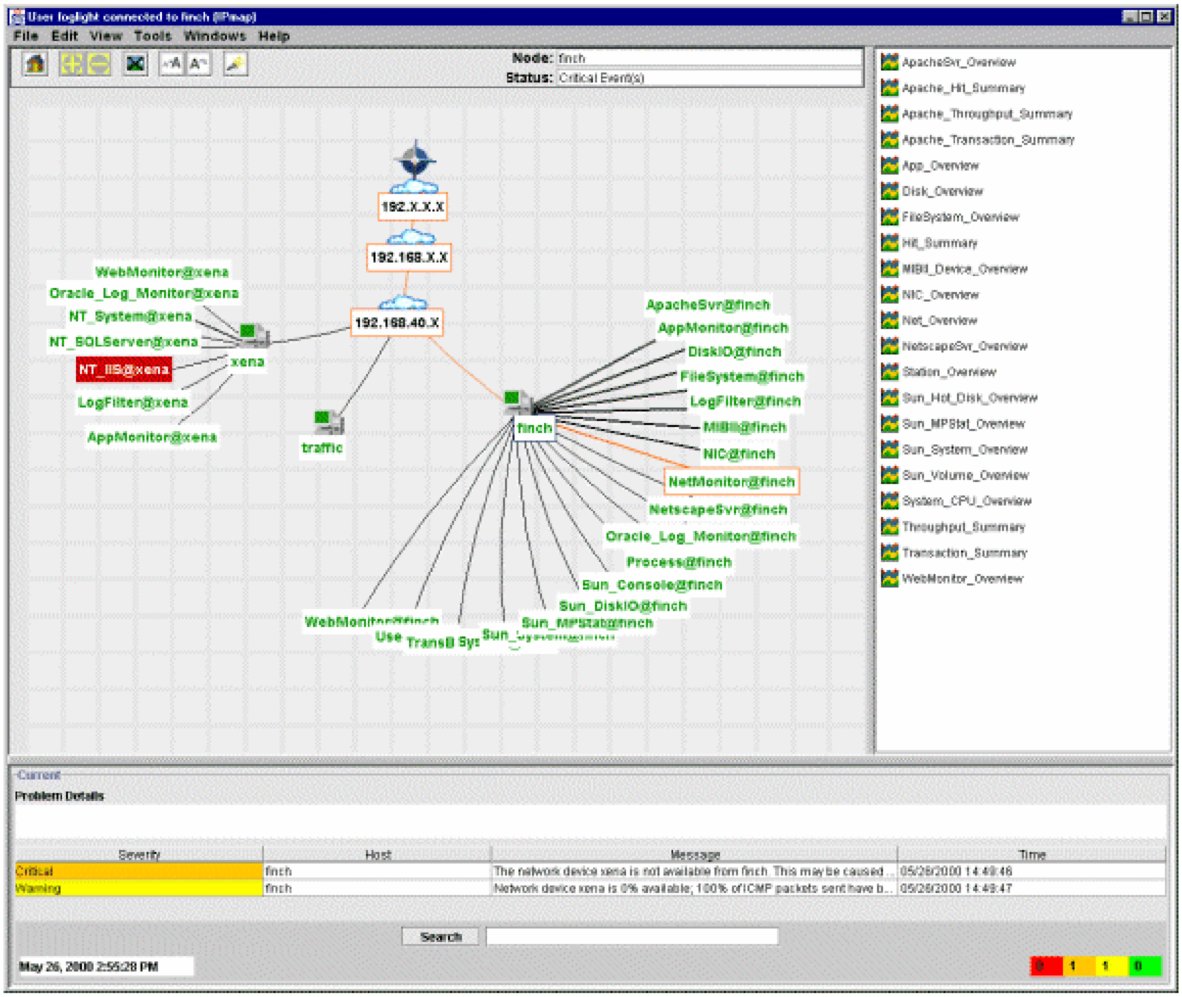
Flexible enough to support unique applications, DataStage lets developers build solutions using a library of standard transforms, then extend it via laser-developed "plug-in" transforms and stages.

The DataStage Repository Manager facilitates collecting, creating, maintaining, and organizing meta data (data about your operational data) and specified target warehouse.

DataStage interoperates with an extensive range of data sources and targets, including the industry's most popular databases and legacy data.

Development and meta data management are simplified through a set of powerful, intuitive graphical design applications - the DataStage Designer and Repository Manager.

Built-in bulk loaders and plug-in stages extract data and load target databases quickly and easily. Available for Oracle, Sybase, Red Brick, MS SQLServer, Sybase IQ and others.



Node: finch
Status: Critical Event(s)

Problem Details

Severity	Host	Message	Time
Critical	finch	The network device xena is not available from finch. This may be caused ...	05/26/2000 14:49:46
Warning	finch	Network device xena is 0% available; 100% of ICMP packets sent have b...	05/26/2000 14:49:47

May 26, 2000 2:55:28 PM