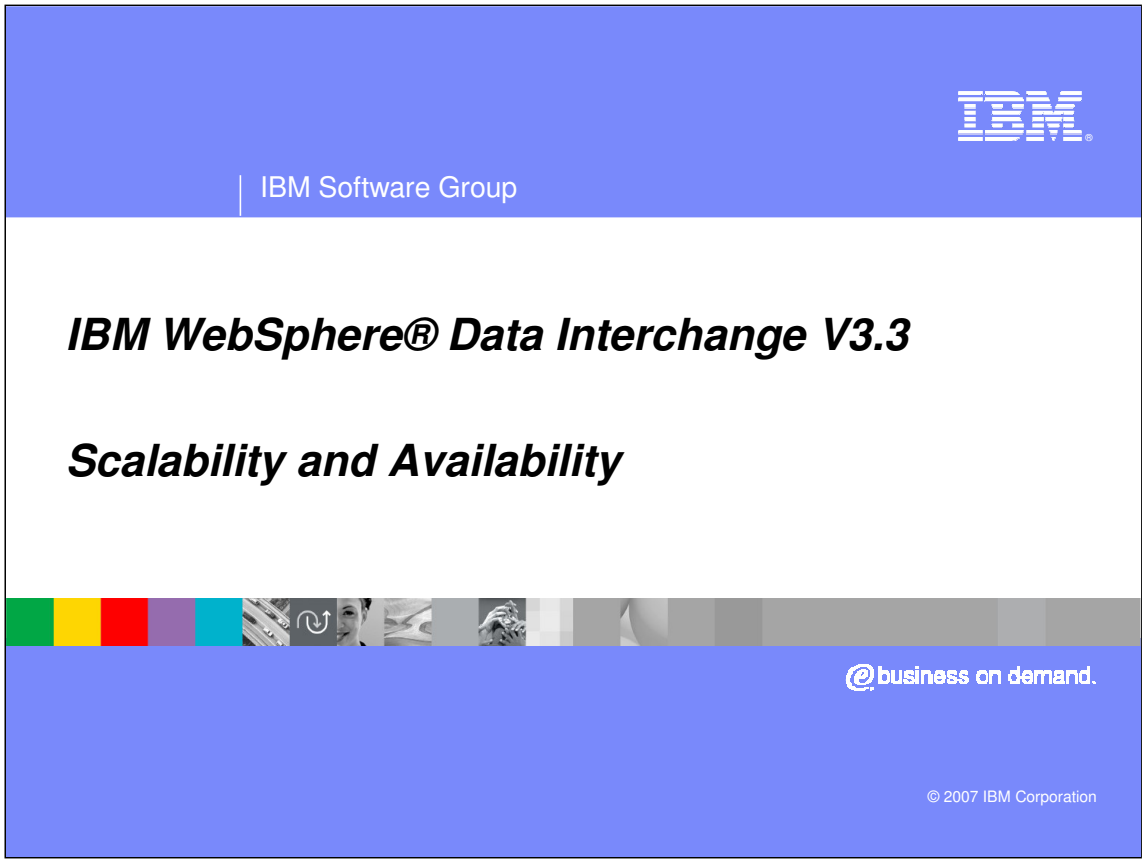


This presentation will describe EDI scalability and High Availability.

Agenda

- What is HACMP?
- EDI HA Architecture
- Summary
- Reference

The presentation will give an overview EDI work load management and high availability. System administrators, system engineers, and other information systems professionals who want to learn about features and functionality provided by the HACMP software should be aware of scalability and high availability concepts.

What is HACMP?

- High Availability Cluster Multiprocessing (HACMP)
  - high availability (HA)
  - cluster multi-processing (CMP)
    - Parallel access
    - Nodes
    - Resources

HACMP is a collection of utilities and practices that is part of the AIX operating system. A "cluster" is a collection of nodes and resources (such as disks and networks) which cooperate to provide high availability of services running within the cluster. Clustering servers enables parallel access to data, which can help provide the redundancy and fault resilience required for business-critical applications

With HACMP software, critical resources remain available. For example, an HACMP cluster could run a database server program that services client applications. The clients send queries to the server program that responds to their requests by accessing a database, stored on a shared external disk.

What is HACMP?

- Custom software
- Industry-standard hardware
- Minimize downtime
- Restore services

This high availability system combines custom software with industry-standard hardware to minimize downtime by quickly restoring services when a system, component, or application fails. Although *not* instantaneous, the restoration of service is rapid, usually within 30 to 300 seconds. In an HACMP cluster, to ensure the availability of these applications, the applications are put under HACMP control. HACMP takes measures to ensure that the applications *remain available* to client processes even if a component in a cluster fails. To ensure availability, in case of a component failure, HACMP moves the application (along with resources that ensure access to the application) to another node in the cluster.

What is HACMP?

- NOT the same as hardware availability
- Complex application access requirements
  - Nodes (CPU, memory)
  - Network interfaces (including external devices in the network topology)
  - Disk or storage devices.
- Causes for downtime
  - Operator errors
  - Environmental problems
  - Application and operating system errors

High availability is sometimes confused with simple hardware availability. Fault tolerant, redundant systems (such as RAID) and dynamic switching technologies (such as DLPAR) provide recovery of certain hardware failures, but do *not* provide the full scope of error detection and recovery required to keep a complex application highly available.

Recent surveys of the causes of downtime show that actual hardware failures account for only a small percentage of unplanned outages.

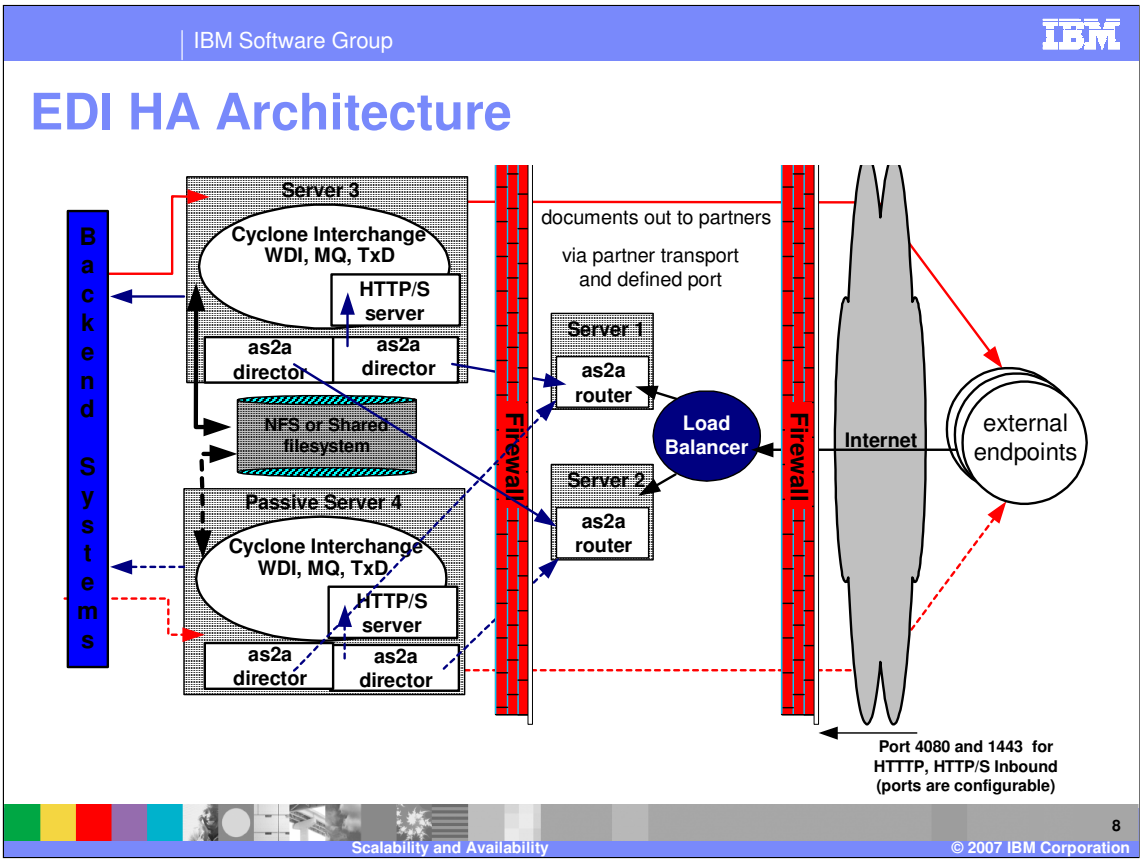
Section

EDI HA Architecture

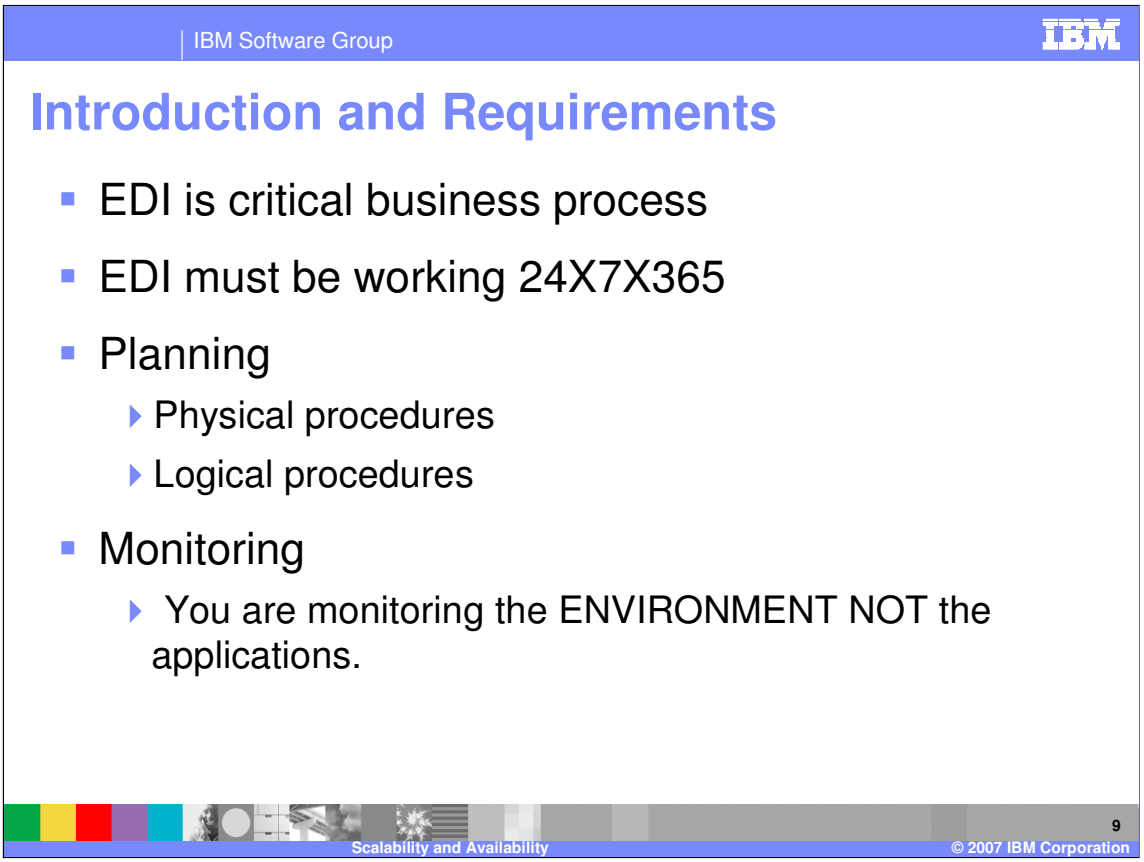
EDI HA Architecture

- Introduction/Requirements
- Review the HACMP Environment
- Concept of an EDI Server
- DB2 Choices
- MQ Adapter Choices
- Setup of the Environment
- Testing of the Environment
- Starting and Stopping WDI on AIX
- Lessons Learned

This section will apply the HACMP to an EDI environment.



What happens when one Server goes down? Trading partners are sending data all time. Keeping systems highly available should be the top goal of every system administrator or corporation. In this example we have two servers Server 3 and Server 4 with HACMP compatibility. Both have a Cyclone Interchange, WDI, MQ, TXD and two AS2A directors are running. The two AS2A directors are in each server has a one to one relation with each of the AS2A router. Both the servers are using common SAN (Shared system) storage. One server is active and another is passive, so when one server fails (hardware or Network) then second server, Server 4, will start automatically with same configuration. The second (passive) server is started and it clones itself as the primary server.



In 2001, more than 2 trillion U.S. dollars in transactions were traded via traditional EDI architectures. EDI is established in 95% of Fortune 500 companies, and many of these enterprises have been reluctant to extend this EDI solution into XML-based implementations. As a result, there is a requirement for a robust EDI interchange solution.

What every business needs is high-availability (HA) solutions that keep a company's server running 24x7x365, allow end users to never experience any system outages, and let system maintenance occur without causing downtime. This requires thorough and complete planning of the physical and logical procedures for access and operation of the resources on which the application depends.

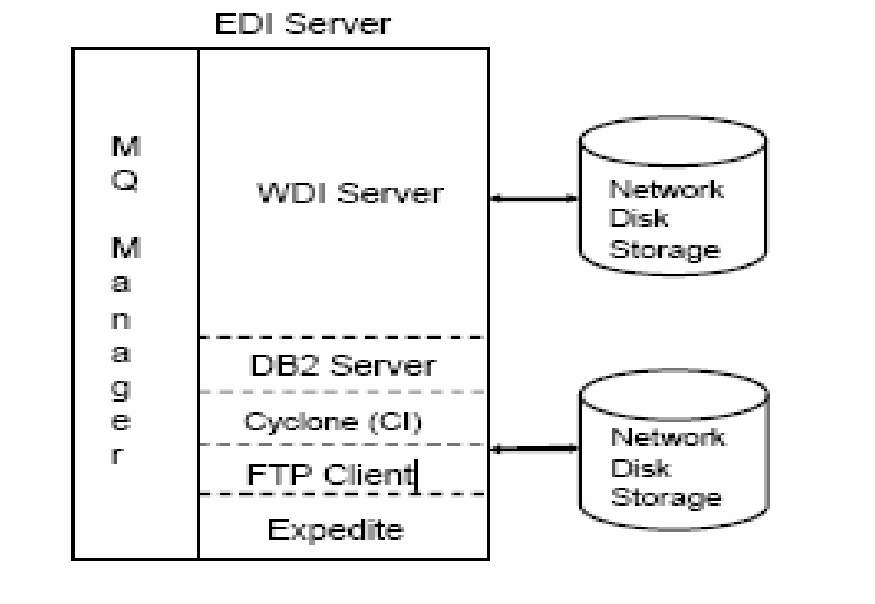
High availability requires a monitoring and recovery package that automates the detection and recovery from errors and a well-controlled process for maintaining the hardware and software aspects of the cluster configuration while keeping the application available.

HACMP Environment

- Eliminate single points of failure
  - Failure Detection and Automated Recovery to Redundant System
  - Switching Back to Primary System is Automatic or Manual (Ping-Pong Effect)
- A budget for hardware and software is required
- A Support Team to build and test AIX on your HACMP Environments

Your major goal throughout the planning process is to eliminate single points of failure. A *single point of failure* exists when a critical cluster function is provided by a single component. If that component fails, the cluster has no other way of providing that function, and the application or service dependent on that component becomes unavailable. Realize that, while your goal is to eliminate all single points of failure, you may have to make some compromises. There is usually a cost associated with eliminating a single point of failure. For example, purchasing an additional hardware device to serve as backup for the primary device increases cost. The cost of eliminating a single point of failure should be compared against the cost of losing services should that component fail. Again, the purpose of the HACMP is to provide a cost-effective, highly available computing platform that can grow to meet future processing demands.

Concept of an EDI Server



For each critical application, be mindful of the resources required by the application, including its processing and data storage requirements. For example, when you plan the size of your cluster, include enough nodes to handle the processing requirements of your application after a node fails.

WebSphere Data Interchange can run in a HA cluster environment to ensure fault tolerance and recoverability. It can also run in an MQ Clustered environment to ensure fault tolerance, workload balancing and an infinitely scalable architecture.

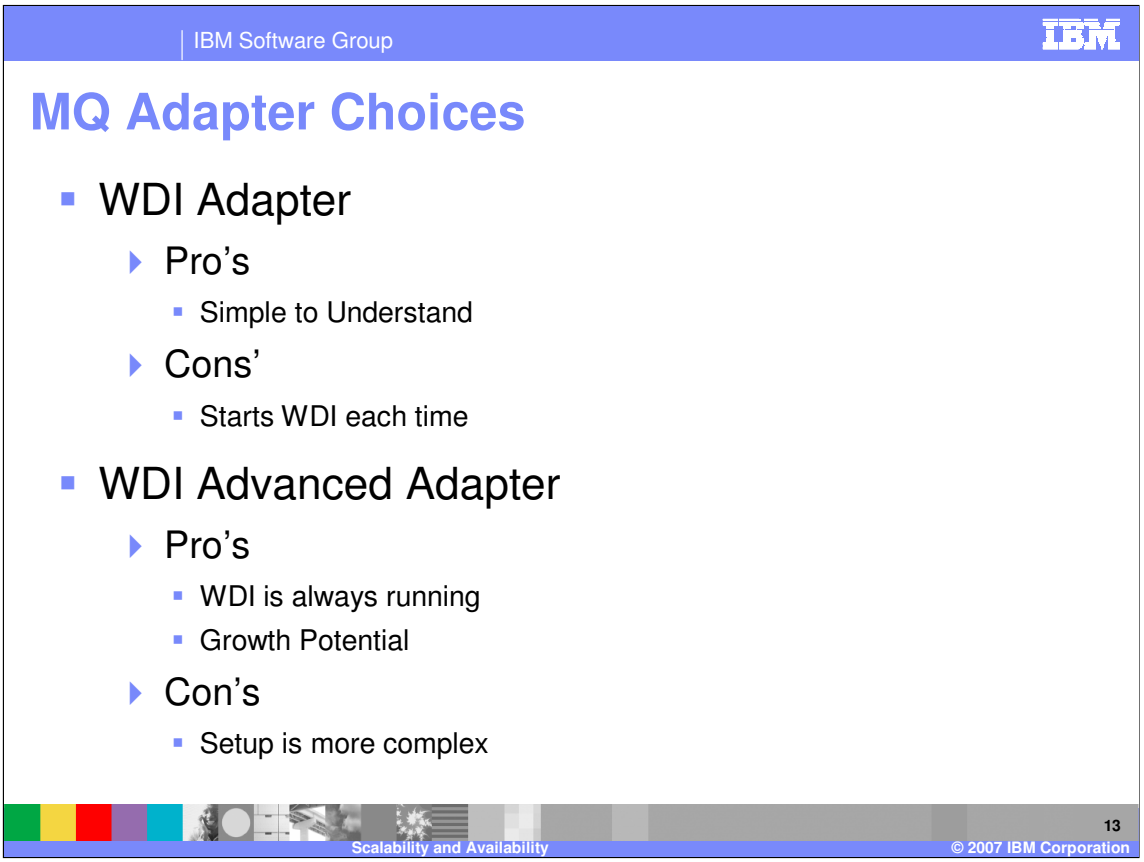
Clustered services, such as WMQ queue managers, are configured to use virtual IP addresses which are under cluster control. When a clustered service moves from one cluster node to the other, it takes its virtual IP address with it. The virtual IP address is different to the stationary physical IP address that is assigned to a cluster node. Remote clients and servers which need to communicate with clustered services must be configured to connect to the virtual IP address and must be written such that they can tolerate a broken connection by repeatedly trying to reconnect.

DB2 Choices

- DB2 Server on Same Platform as EDI Server
  - Pro's
    - Better Performance
  - Con's
    - Down when WDI Server is down
    - Whom to monitor/admin DB2?
- DB2 Server on Separate Platform from EDI Server
  - Pro's
    - Up Even while EDI Server is down
    - DB2 Group to Admin/Monitor?
  - Con's
    - Slower Performance

The WDI server database might be executed in a database instance on the same node as the server, in which case the database instance and its lower level dependencies must be failed over with the server. Alternatively the server database might be executed in a remote instance accessed using a remote ODBC connection, in which case it is necessary is to ensure that the database is accessible from either cluster node, so that the server can operate correctly on either node.

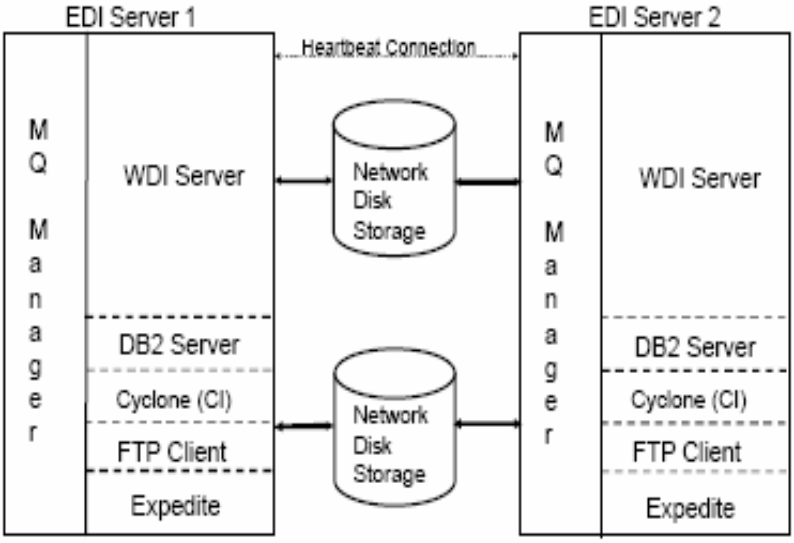
The use of a shared database ensures a simple single configuration point for all of the WebSphere Data Interchange servers running in the MQ cluster.



The WebSphere Data Interchange adapter program is installed as part of WebSphere Data Interchange for Multiplatforms Version 3.3. The configuration scripts provided set up the necessary queues and definition objects. The adapter uses WebSphere MQ Triggering to know when messages need processing. When a message is put to an application queue, a trigger message is created. The WMQ trigger monitor receives the message and executes the adapter. The adapter then passes the information needed to process the application message to the WebSphere Data Interchange server/translator. Application messages are committed, rolled back, or moved to a failure queue depending on the return codes from the WebSphere Data Interchange Server. The adapter will wait the user-configured time interval for any successive messages, and then terminate. The trigger monitor then restarts WebSphere Data Interchange adapter upon receipt of another trigger message.

The advanced adapter is designed to allow the user to create a pool of translators when the WDIServer program is started. When data arrives on an WMQ queue one or more of these translators can be assigned to the queue to process data. This means that each queue can be independently configured to request any number of translators (up to the number of translators in the initial pool).

Setup Environment



HACMP uses *heartbeating* to monitor the activity of its network interfaces, devices and IP labels. A *heartbeat* is a type of a communication packet that is sent between nodes. Heartbeats are used to monitor the health of the nodes, networks and network interfaces, and to prevent cluster partitioning. Heartbeating connections between cluster nodes are necessary because they enable HACMP to recognize the difference between a network failure and a node failure. For instance, if connectivity on the HACMP network (this network's IP labels are used in a resource group) is lost, and you have another TCP/IP based network and a non-IP network configured between the nodes, HACMP recognizes the failure of its cluster network and takes recovery actions that prevent the cluster from becoming partitioned.

Testing the HACMP Environment

- Testing
  - Verify the cluster configuration
    - Initial configuration
    - Changes to the cluster
  - Test the HACMP cluster
    - Test Plan
    - Test Procedure
      - Automated test procedure
      - Custom component testing

Verifying the cluster configuration assures you that all resources used by HACMP are validly configured, and that ownership and takeover of those resources are defined and are in agreement across all nodes. You should verify and synchronize your cluster configuration after making *any* change within a cluster (for example, any change to the hardware operating system, node configuration, or cluster configuration).

To test the HACMP cluster you should develop a test plan and procedures. Make sure that you are familiar with the HACMP clusters on which you plan to run the test. List the components in your cluster and have this list available when setting up a test plan.

Automated testing may be available but most likely executes a series of predefined sets of tests on the cluster. Custom testing will be necessary for system components, for example WebSphere Data Interchange Server.

Testing the HACMP Environment

Sample Custom Test Procedures

1. Stop EDI Server 1 Normally – EDI Server 2 takes over
2. System Crash on EDI Server 1 – EDI Server 2 takes over
3. With Inbound data flowing, stop EDI Server 1 – Let EDI Server 2 take over
4. With Inbound data flowing, crash EDI Server 1 – Let EDI Server 2 take over
5. With Outbound data flowing, stop EDI Server 1 – Let EDI Server 2 take over
6. With Outbound data flowing, crash EDI Server 1 – Let EDI Server 2 take over
7. With Inbound AND Outbound data flowing, stop EDI Server 1 – Let EDI Server 2 take over
8. With Inbound AND Outbound data flowing, crash EDI Server 1 – Let EDI Server 2 take over

Your test procedure should bring each component offline then online, or cause a resource group fallover, to ensure that the cluster recovers from each failure. Listed on this slide are sample EDI server test scenarios. Start Up and Stop scripts have to be written in order to start/stop the applications.

WebSphere Data Interchange Startup

- Sample WDI Start Up Script

```
export WDISERVER_PROPERTIES =
  /usr/wdi/run/wdi.properties
cd /usr/wdi/run
rm nohup.out
nohup /usr/wdi/DIv32/bin/WDIServer &
```


This is a sample WebSphere Data Interchange start up script.

WebSphere Data Interchange Stop

- Sample WDI Stop Script

```
#! /bin/ksh
#
# Stop WDI
WDIPID=`ps -fu wdi|grep WDIServer|grep -v grep|awk '{print $2}'`
cd /usr/wdi/run
/usr/wdi/DIv32/bin/WDIShutdown
while [ ${WDIPID}X != X ]
do
    echo "WDI Server is running (PID $WDIPID)"
    echo "Please wait while WDI is closed down. This may take a while..."
    sleep 5
    WDIPID=`ps -fu wdi|grep WDIServer|grep -v grep|awk '{print $2}'`
done
#
```


This is a sample WebSphere Data Interchange stop script.

Lessons Learned

- Users Ids between the two boxes are separate
- Anything on the Local Hard Disk doesn't Switch
  - /usr/home/userid – Data didn't switch
  - /etc
- First Test Case Takes the longest (Start and Stop Scripts)
- DB2 has to be cataloged on each system (if not local)
- WDI needs time to shutdown. Issuing the command is only the start of the process. Write script to maintain control.
- WDI, when crashed, will leave behind directories and files when using the Advanced Adapter
- Make sure MQ has Persistent set on. Either at the data or the queue level.
- Use Queues as much as possible. File system, even networked, lead to problems
- Upon start up of the backup system, messages waited over two hours before WDI started processing the files. Why? DB2 timeout value too large. Changed to 10 minutes.
  - Changed tcp_keepidle to 540 with command
    no -o tcp_keepidle=540
- Switching between systems (EDI 1 to EDI 2) takes several minutes.

Remember your major goal throughout the process is to eliminate single points of failure. Listed on this slide are a few observations noted with the execution of the sample EDI Server test scenarios mentioned in this document.

The basic measure of success for a test is availability. The following are some examples of criteria that can be used to determine the success or failure of cluster tests:

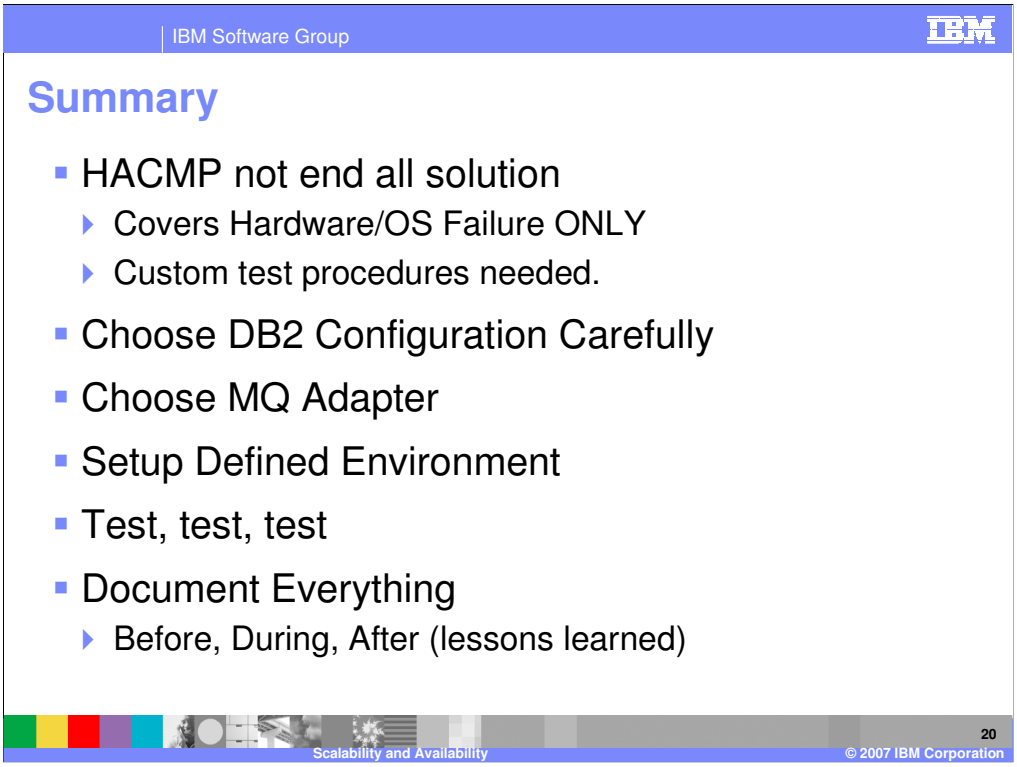
• Did the cluster stabilize? Nodes that should be online are online. If a node is stopped and that node is the last node in the cluster, the cluster is considered stable.

• Has an appropriate recovery event for the test executed?

• Is a specific node online or offline as specified?

• Are all expected resource groups still online within the cluster?

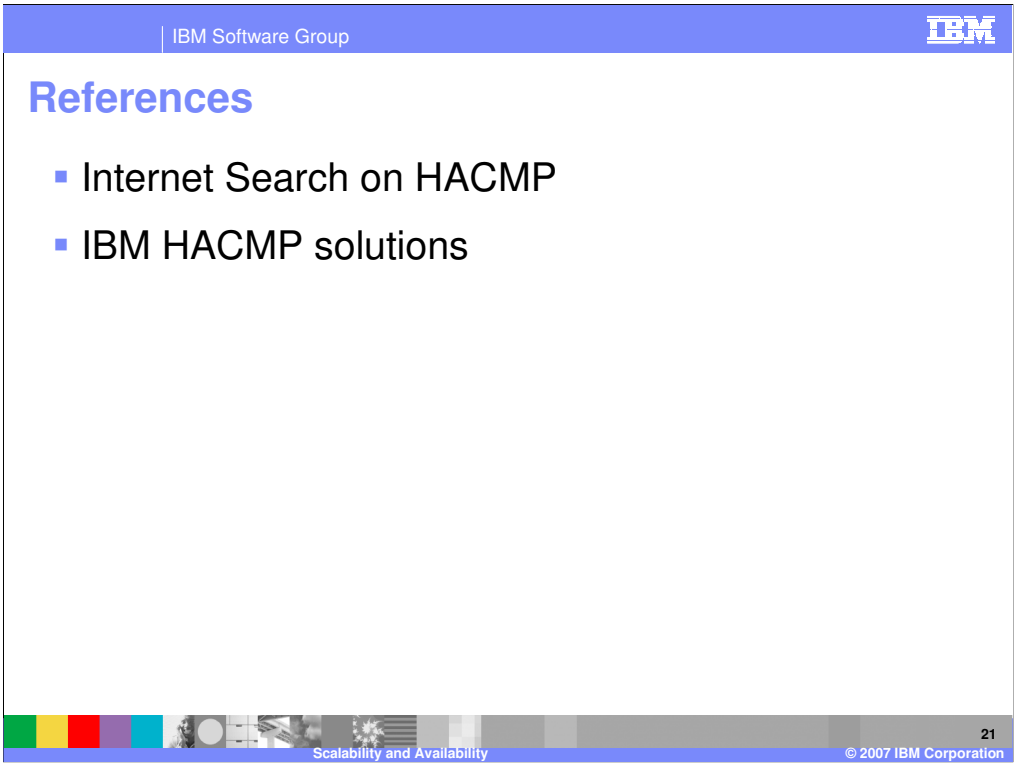
• Did a test that was expected to execute actually execute?



You will need to create a custom test plan and procedures, to meet requirements specific to your environment.

Spend considerable time in the planning stage. This is where the bulk of the documentation will be produced and will lay the foundation for a successful production environment. Start by building a detailed requirements document and focus on what you want and need it to do Next, build a technical detailed design document. Details should include a thorough description of the Storage / Network / Application / Cluster environment and the Cluster Behavior. Finally, make certain the cluster undergoes comprehensive and thorough testing before going live and document the results.

It is your responsibility to document all aspects of the HACMP system unique to your environment. This responsibility includes documenting procedures concerning the highly available applications, recording changes that you make to the configuration scripts distributed with HACMP, documenting any custom scripts you write, recording the status of backups, maintaining a log of user problems, and maintaining records of all hardware. This documentation, along with the output of various display commands and cluster snapshots, will be useful for you and for product support, to help resolve problems.



A simple Internet search on HACMP will give a listing of products and documentation related to HACMP. IBM has an extensive set of software products and documented procedures to implement HACMP. IBM may not offer all the products, services, or features in countries other than the US. Consult your local IBM representative for information on the products and services currently available in your area.

Trademarks, copyrights, and disclaimers

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM	CICS	IMS	WebSphere MQ	Tivoli
IBM(logo)	Cloudscape	Informix	OS/390	WebSphere
e(logo)business	DB2	iSeries	OS/400	xSeries
AIX	DB2 Universal Database	Lotus	pSeries	zSeries

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds.

Other company, product and service names may be trademarks or service marks of others.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

Information is provided "AS IS" without warranty of any kind. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.