IBM

ExperienceReport

Road Map to iSeries WebSphere High-Availability Topologies

Darin Scherer
Steve Simonson

October 2002

Road Map to iSeries WebSphere High-Availability Topologies

IBM @servers

Darin Scherer and Steve Simonson

Copyright IBM Corporation, 2001. All Rights Reserved.
This publication may refer to products that are not currently available in your country. IBM makes no commitment to make available any products referred to herein.

Trademarks and Disclaimers

IBM @server

© IBM Corporation 1994-2002. All rights reserved.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

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

AS/400	IBM(logo)	DB2
AS/400e	iSeries	
e (logo) business	OS/400	
IBM	WebSphere	

Lotus, Freelance Graphics, and Word Pro are registered trademarks of Lotus Development Corporation and/or IBM Corporation.
Domino is a trademark of Lotus Development Corporation and/or IBM Corporation.

C-bus is a trademark of Corollary, Inc. in the United States, other countries, or both.
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 trademarks of Microsoft Corporation in the United States, other countries, or both.
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.
SET and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.
Other company, product and service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information in this presentation concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.

Some information in this presentation addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. 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.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

General Comments

IBM @server

- This presentation documents current status of an initiative to test and document HA e-commerce topologies on iSeries.
- Substantial quality (HA e-commerce) documentation from IBM exists today for WebSphere. This presentation uses this documentation as a base and points out iSeries considerations as necessary. The testing effort and results are work in progress. The following Redbook and other documents are recommended reading for additional information:
  - IBM WebSphere V4.0 Advanced Edition Scalability
    - SG24-6192-00
  - White paper - Failover and Recovery in WebSphere
    - http://www7.software.ibm.com/vadd-bin/ftpdl?1/vadc/wsdd/pdf/modjeski.pdf
  - WebSphere Application Server 4.0 (online documentation)
    - http://publib.boulder.ibm.com/was400/40/AE/english/docs/pvindex10.html
  - WebSphere Application Server 3.5 (online documentation)
    - **http://www-1.ibm.com/servers/eserver/iseries/software/websphere/wsappserver/docs/as400v35/docs/**

General Comments...

IBM @server

- Testing performed with two applications
  - E-Investments - internal IBM test application (topic of this presentation)
  - IBM WebSphere Commerce (performed at a later date)
- Partner with interested parties to disseminate required information to iSeries customers.
  - Custom Technology Center (CTC)
  - iSeries Technology Center (ITC)
  - iSeries Advanced Technical Support
  - PartnerWorld
  - WebSphere Sales and Support Center (WSSC)
  - HABP companies (DataMirror, Lakeview Technology, Vision Solutions)

General Comments...

IBM @server

- Share results with other platforms
- Analyze results
  - Suggest enhancements to WebSphere or iSeries platform
- Recommend coding practices in user applications to enable attainment of high-availability goals

High Availability versus Continuous Availability

IBM @server

- High Availability - Minimal loss of service to meet the availability needs of a particular business, including handling of unplanned and planned outages
- Continuous Availability - 24 X 365 operations, no planned or unplanned outages that affect applications that are deemed critical to be accessible at all times
- Very difficult and expensive to assure Continuous Availability

J2EE - Java 2 Platform Enterprise Edition ...

- Defines a standard that applies to all aspects of architecting, developing and deploying multi-tier server based applications. WebSphere 4.0 is J2EE compliant and utilizes open standards.

J2EE - Java 2 Platform Enterprise Edition ...

- **J2EE further defined**:
  - **Application Components**
    - Application clients
      - Web/Browser
      - Java/Thick client - Deployed in Client container
      - Thin client - Standalone Java program
    - Applets
    - Servlets and JSP
    - EJBs (Enterprise Java Beans)
  - **Containers**
    - All application components run in containers built on J2EE specification
      - Client container - Java clients
      - Web container- Servlets and JSPs
      - EJB container - Enterprise Java beans
      - Applet container

J2EE - Java 2 Platform Enterprise Edition

IBM @server

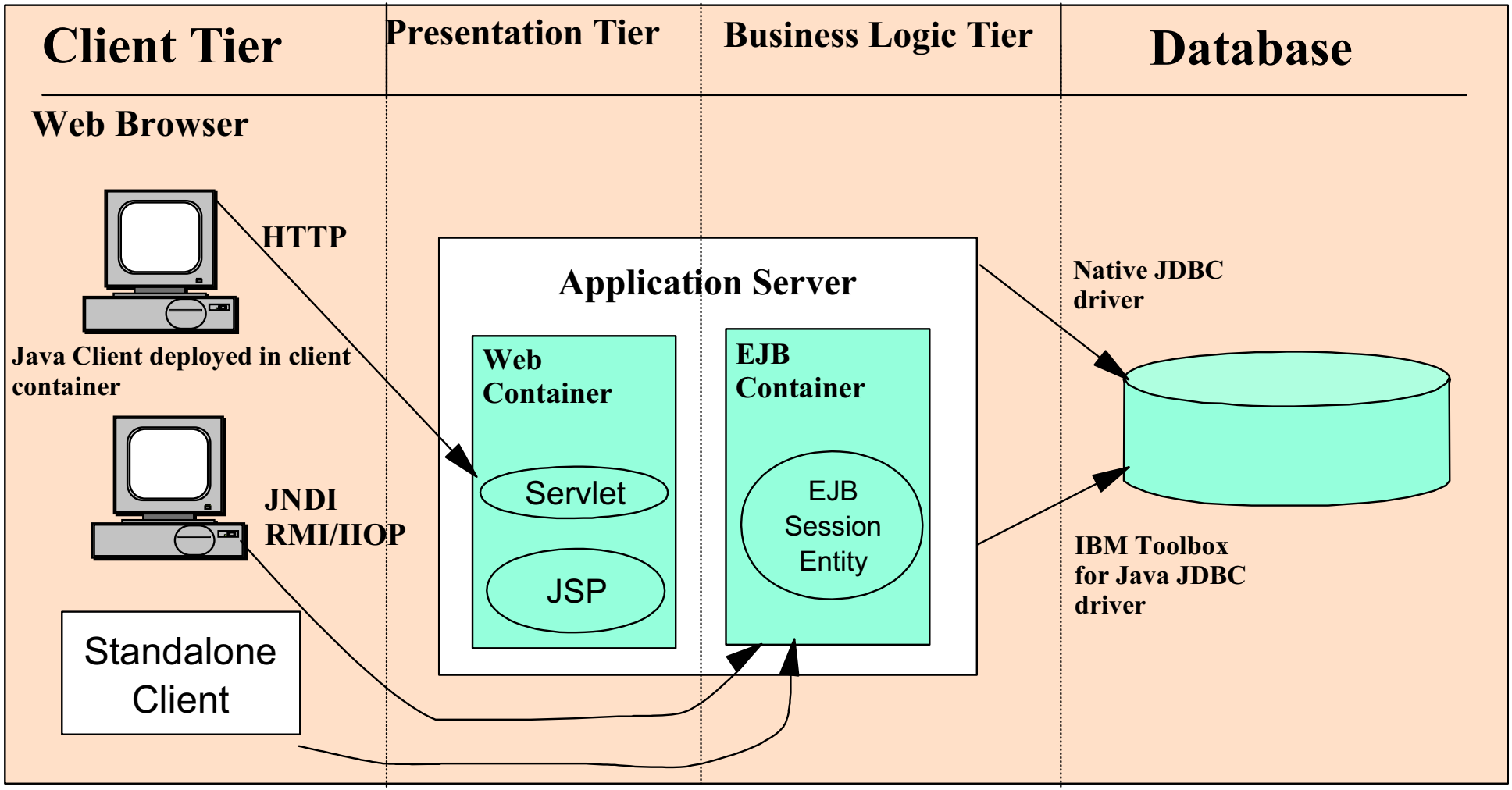
- **J2EE further defined**: (Continued)
  - **Resource drivers**
    - Provide access to a shared resource such as a database. Resource manager drivers are software components that provide network connectivity to an external resource manager.
    - JDBC drivers
      - Native JDBC driver - uses DRDA
      - IBM Toolbox for Java JDBC driver - preferred network driver
- **J2EE Services**
  - J2EE platform provides components with a set of standard services that can be used to communicate with one another
    - HTTP and HTTPS - Web container
    - JNDI - Java Naming and Directory Interface - allows J2EE components to locate other objects needed

J2EE - Java 2 Platform Enterprise Edition

- **J2EE Services** (Continued)
  - J2EE platform provides components with a set of standard services that can be used to communicate with one another
    - JDBC - Java Database Connectivity - provides connectivity with relational database
    - JMS - Java Messaging Service
    - JTA JTS - distributed transaction management
    - RMI/IIOP - implementation of RMI APIs over IIOP protocol. Allows developers to write remote interfaces in the Java programming language.

Multi-tier distributed application model

IBM @server



Outline of Presentation

IBM @server

- Discuss two eCommerce HA topologies differentiated by database protection methodologies.
  - Database replication
  - Switch disk
- HA components will be discussed in execution order starting at Internet client requests and proceeding to the backend enterprise database.

Components of eCommerce High-Availability

IBM @server

- Dual Edge Server network dispatchers for failover capability
- Dual HTTP servers
- Dual WebSphere domains for J2EE application deployment
  - Dual application deployments (E-Investments internal application)
    - Servlets
    - EJBs
    - Remote OSE/Remote HTTP
    - Persistent session data
- Data source (IP address directed)

Components of eCommerce High-Availability...

IBM @server

- LPAR for both process and data isolation -- logical separation of the IBM HTTP Server from WebSphere -- as well as server consolidation
- HA firewall solution of your choice
- HA DB
  - iSeries clustering
  - Remote journal
  - Database replication software from DataMirror, Lakeview Technology and Vision Solutions
  - IASP and switchable disks (V5R2 requirement)

Components of eCommerce High-Availability

IBM @server

- Other possible components of e-business
  - LDAP
  - Various third-party application software and servers
  - HA Internet routing
  - Messaging servers such as MQSeries
  - Thick and thin Java clients

LoadRunner

IBM @server

- Load testing tool from Mercury Interactive used to simulate our 300 client workload
- Provides real-time feedback on success rate of HTTP requests and responses during outage testing

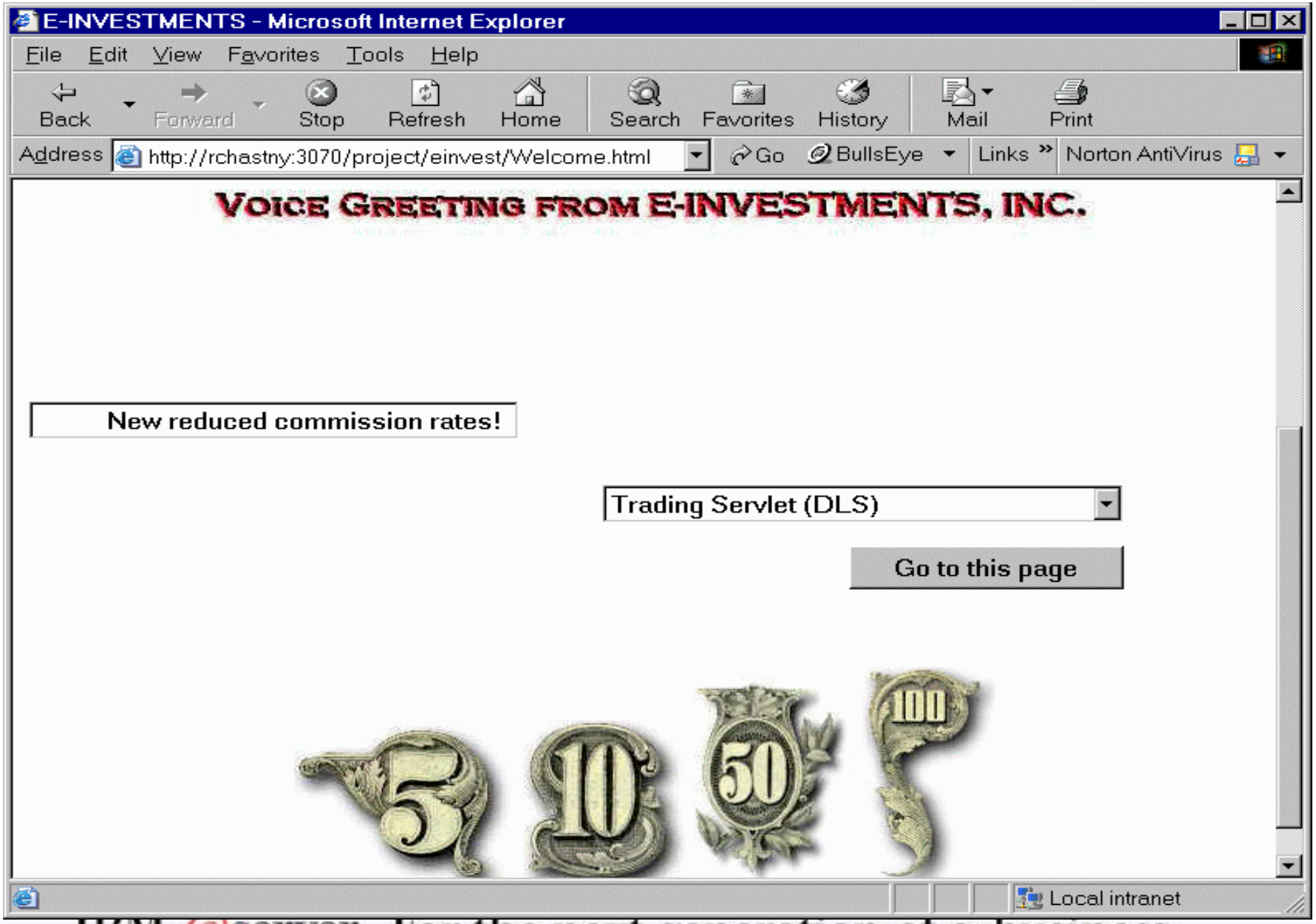
Application - E-Investments

IBM @server

- Servlets
- EJBs - Stateful and Stateless session beans, CMP entity beans
- 5 GB SQL relational database consisting of 18 tables, including UDB LOB fields
- Thick and thin Java client applications will be added for additional testing
- Commitment control used for transaction processing
- The application utilizes stateless session beans which call CMP entity beans. Both the session and entity beans run under required transactional attribute. The stateless session bean is called from a servlet. The database transaction performed by the stateless bean should be atomic - either all files are updated or no files are updated.

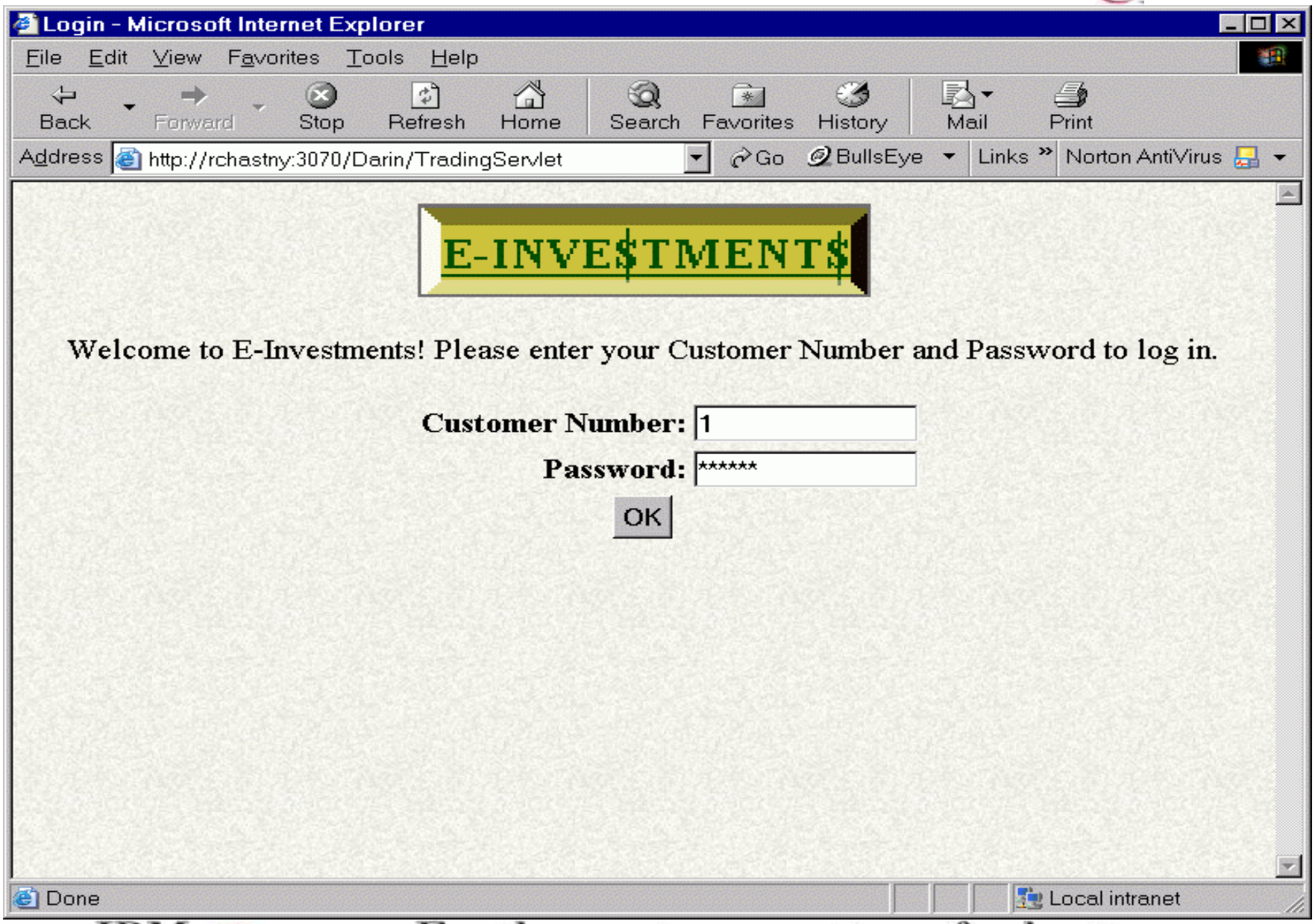
E-Investments Application Screen Shots

IBM @server



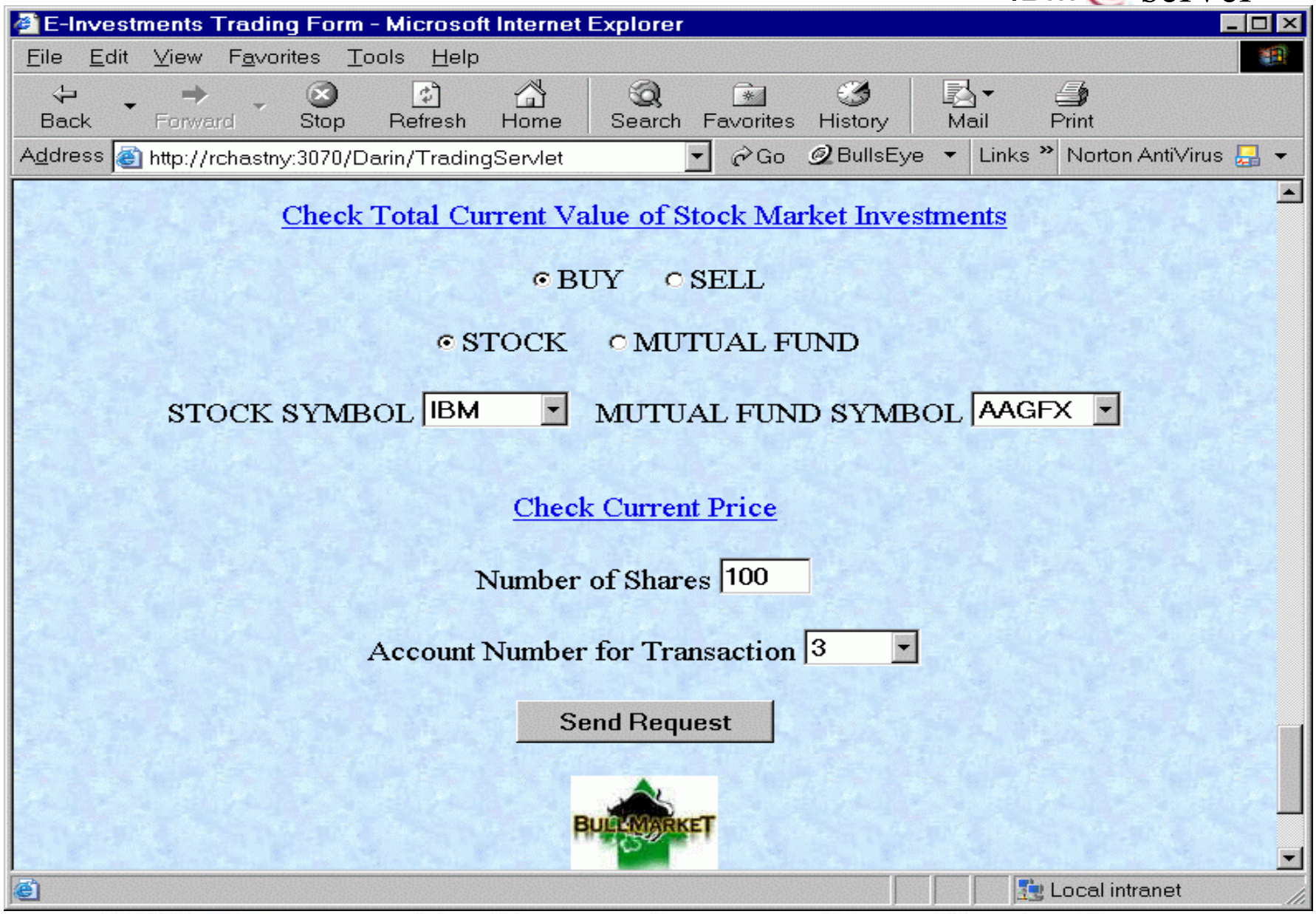
E-Investments Application Screen Shots

IBM @server



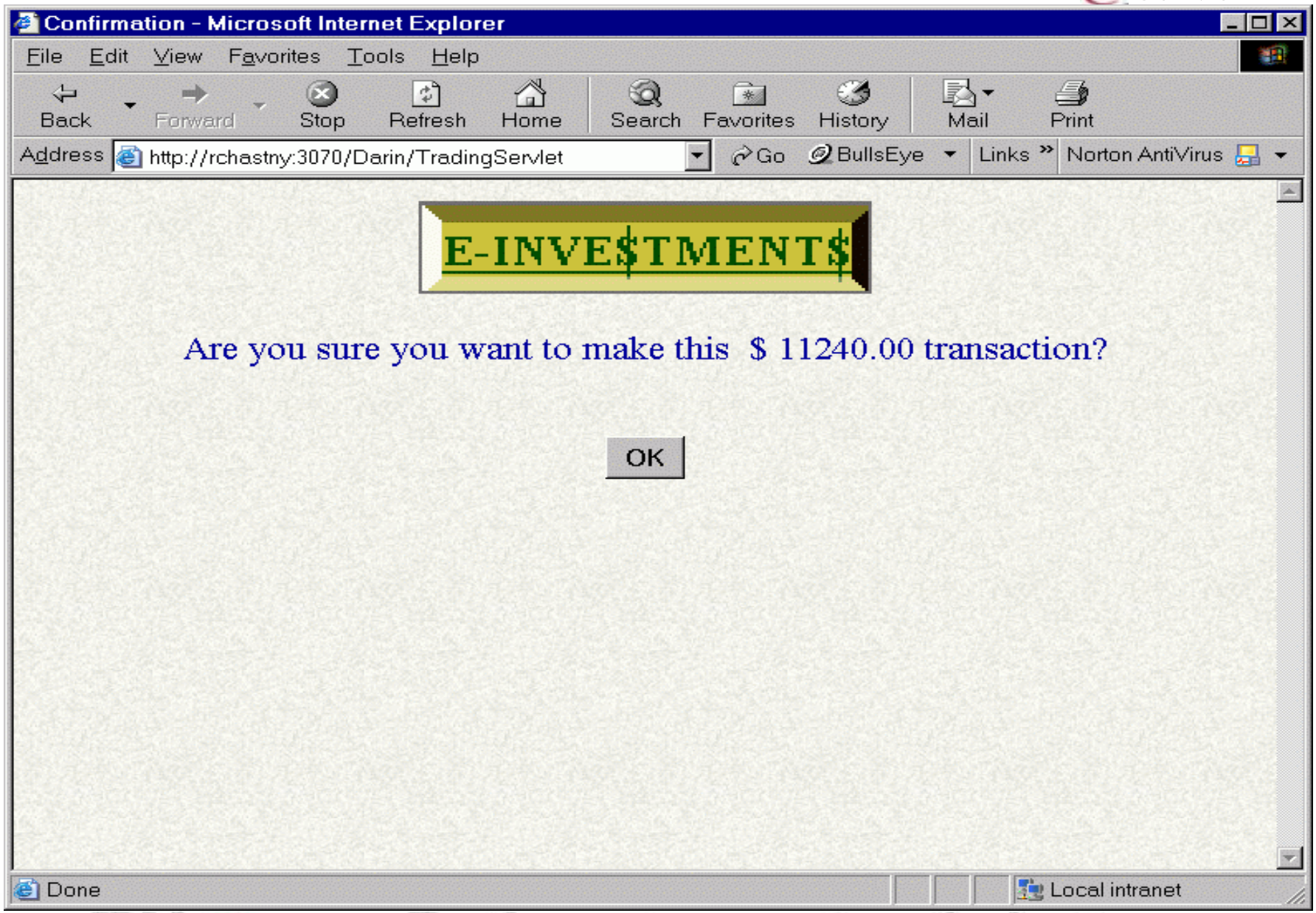
E-Investments Application Screen Shots

IBM @server



E-Investments Application Screen Shots

IBM @server

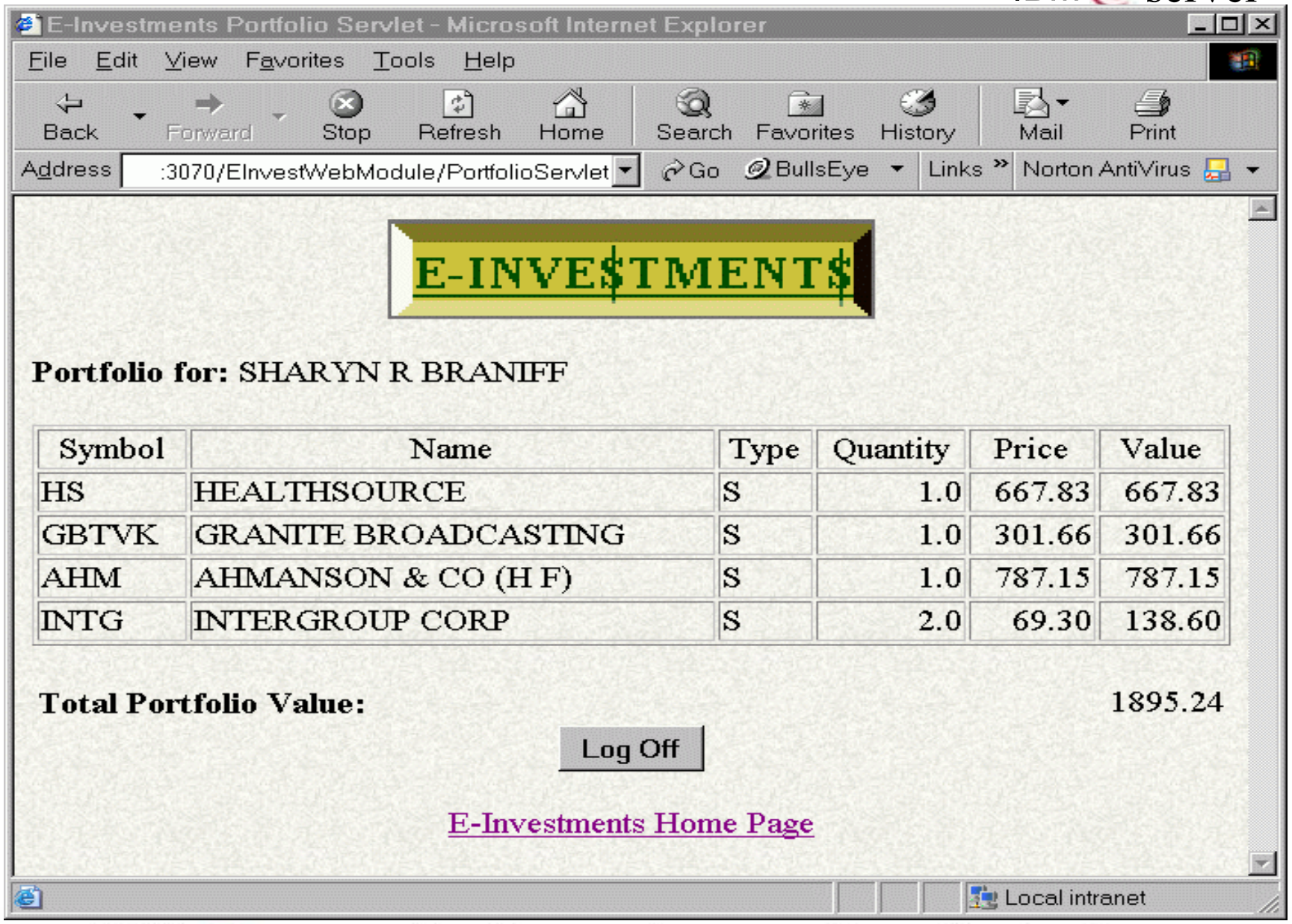


E-Investments Application Screen Shots

IBM @server

E-Investments Application Screen Shots

IBM @server



Recommended Coding Practices for HA Environments...

- Application considerations for high-availability environments
  - Use the servlet session tracking APIs in conjunction with a persistent session data source to preserve important data across requests and to allow access to that data by a different JVM in failover situations
  - Servlets must be thread safe to ensure application reliability
    - Stress test your applications to ensure their functionality and performance capabilities prior to deployment into a HA topology
    - Mercury Interactive's LoadRunner tool was used in our testing
    - Stress test those same applications after deployment into the final topology
    - Limit usage of class or instance variables to avoid threading issues

Recommended Coding Practices for HA Environments...

IBM @server

- Servlets must be thread safe (Continued)
  - Ensure that instance variables are always set properly in both the doGet() and doPost() service methods since a different JVM may be used to handle subsequent requests
  - Beware of code blocks, including synchronized methods or code sections, that perform special processing or access a common external resource since the same servlet will now be running in multiple JVMs
- Deploy a simple servlet (no database processing, etc.) in each JVM to assist with problem determination
- Code a unique, static web page for each web server which identifies the actual server name to aid in debug efforts when problems arise
- Increase the three 'ping' value settings under the Advanced tab in the WebSphere Admin GUI for each application server or Server Group --- i.e. append a zero to the default values

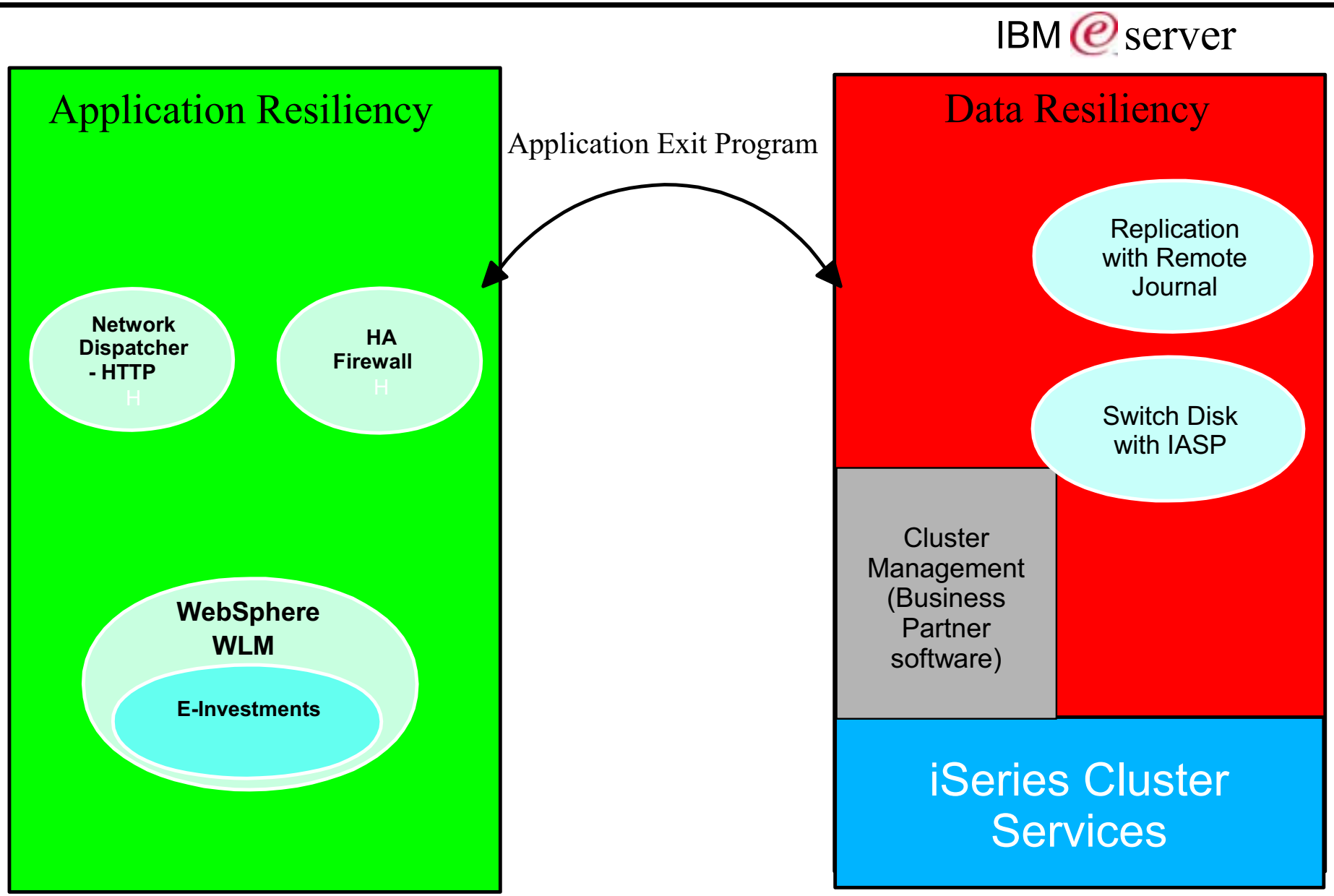
Recommended Coding Practices for HA Environments

- In the exception paths of the servlet, use the log() method to report useful information for debugging purposes
  - Regularly inspect the various WebSphere log files on each WebSphere node to ensure applications are running cleanly
- Use custom advisors component of Network Dispatcher to test the healthiness of your HA topology. Test application components and database. If a problem arises, log adequate information in the advisor logs. Edge Server provides intelligent failover support when built-in or custom advisors are utilized.
- Use a properties file for designating system-specific information (EJB server names, data sources, system names) instead of using servlet initialization parameters or hard-coded references
  - Enables easier deployment of WebSphere applications to multiple domains without recompilation or the need for additional J2EE WAR, EAR, or deployment descriptor files
  - Reduces risk of errors which may affect failover capabilities

Recommended Coding Practices for HA Environments

- Modify two iSeries TCP attribute values via CHGTCPA command on the WebSphere and web server nodes to help prevent total web site unavailability when the backend database server is not available to service requests. Even though the backend database may be unavailable, Web requests not accessing the database should still be serviced.
  - ARP cache timeout value should be 1
    - reason: compensates for iSeries two-minute timeouts for connect requests. WebSphere typically has mutiple retry attempts.
  - TCP R2 retransmission count value should be 9
    - reason: limits retransmission attempts to a total time interval of 2 minutes.

iSeries Cluster Support for HA Database



Failures to test...

IBM @server

- Simulate various failures
  - Firewall failure
  - Network dispatcher failover
  - HTTP server not responding
  - One of the WebSphere nodes crashes
  - Application server fails to respond
  - WebSphere administrative server failure
  - Persistent session data not available
  - Enterprise database not available (outage time less than 5 minutes)
  - Power outage of system -- multiple node failures
  - Disk subsystem failure
  - Network failure

Failures to test

IBM @server

- Simulate various failures (Continued)
  - System upgrade - quiesce one WAS server, load PTFs and then restart without interruption to Internet clients

Common Goals for Our HA Topologies...

IBM @server

- Using WebSphere version 3.5.5/4.03 and V5R1 iSeries release:
  - Create topologies which provide HA (high availability)
    - Some clients may notice interruption of service, others may not
    - Failover time (unscheduled outage) minimized if backend DB fails.
      - iSeries has single system availability rate of 99.9%
    - All open transactions rolled back to last committed boundary after failover complete
      - Assumption that applications incorporate commitment control
    - May perform scheduled maintenance with minimal interruption of service to external users
    - **WebSphere HA mechanisms combined with underlying iSeries HA strengths. Extends current value proposition to e-Commerce implementations for iSeries customers.**

Common Goals for Our HA Topologies

IBM @server

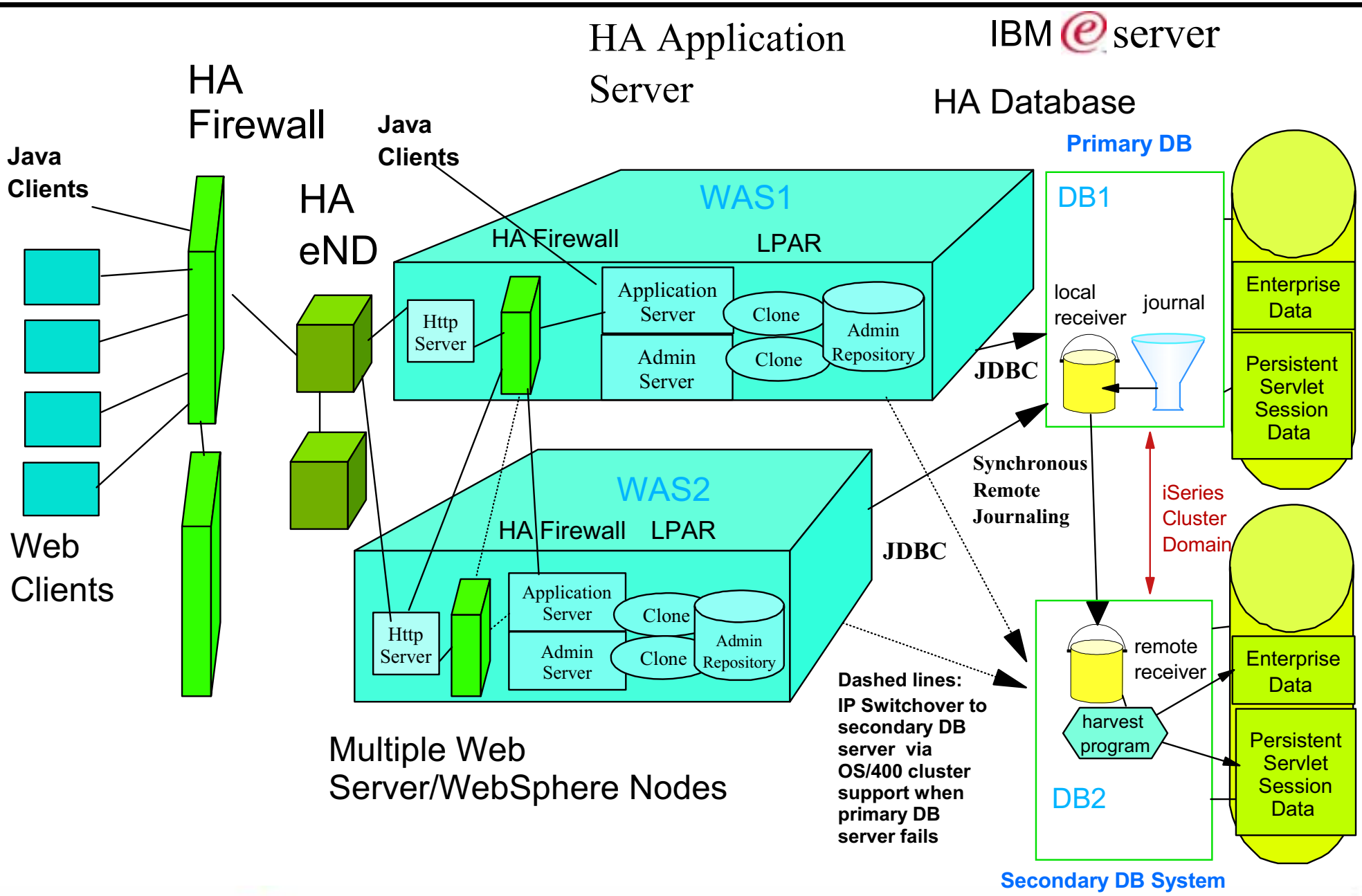
- Remote journal in conjunction with harvesting programs on backup database server keeps DB files current to assure quick failover

Topology I (Data Replication) - Commentary...

IBM @server

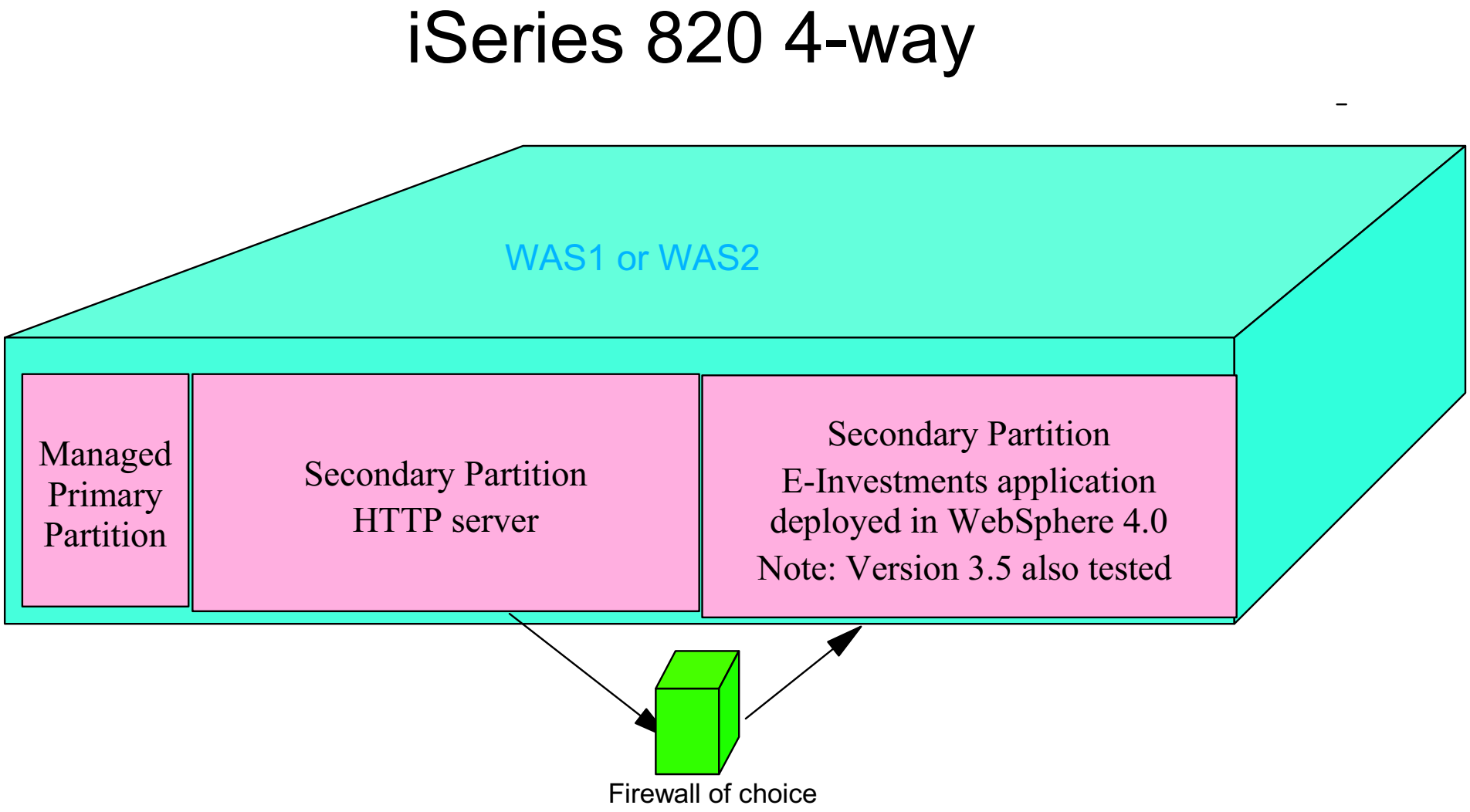
- Four iSeries 820 4-ways (WAS1, WAS2, DB1, DB2)
  - WAS1 and WAS2 - each run identical applications deployed within a WebSphere domain
  - DB1 and DB2 - HA DB (iSeries cluster with remote journal)
- WAS1 and WAS2 use WebSphere vertical cloning for both application servers and admin server
- HTTP Session state persisted to backend DB. Data source directed via IP takeover address
  - Failure of WAS1 or WAS2 will be transparent except for active clients at time of failure
- NetWork Dispatcher provides HTTP failover and load balancing

Topology I - Data Replication (V5R1)



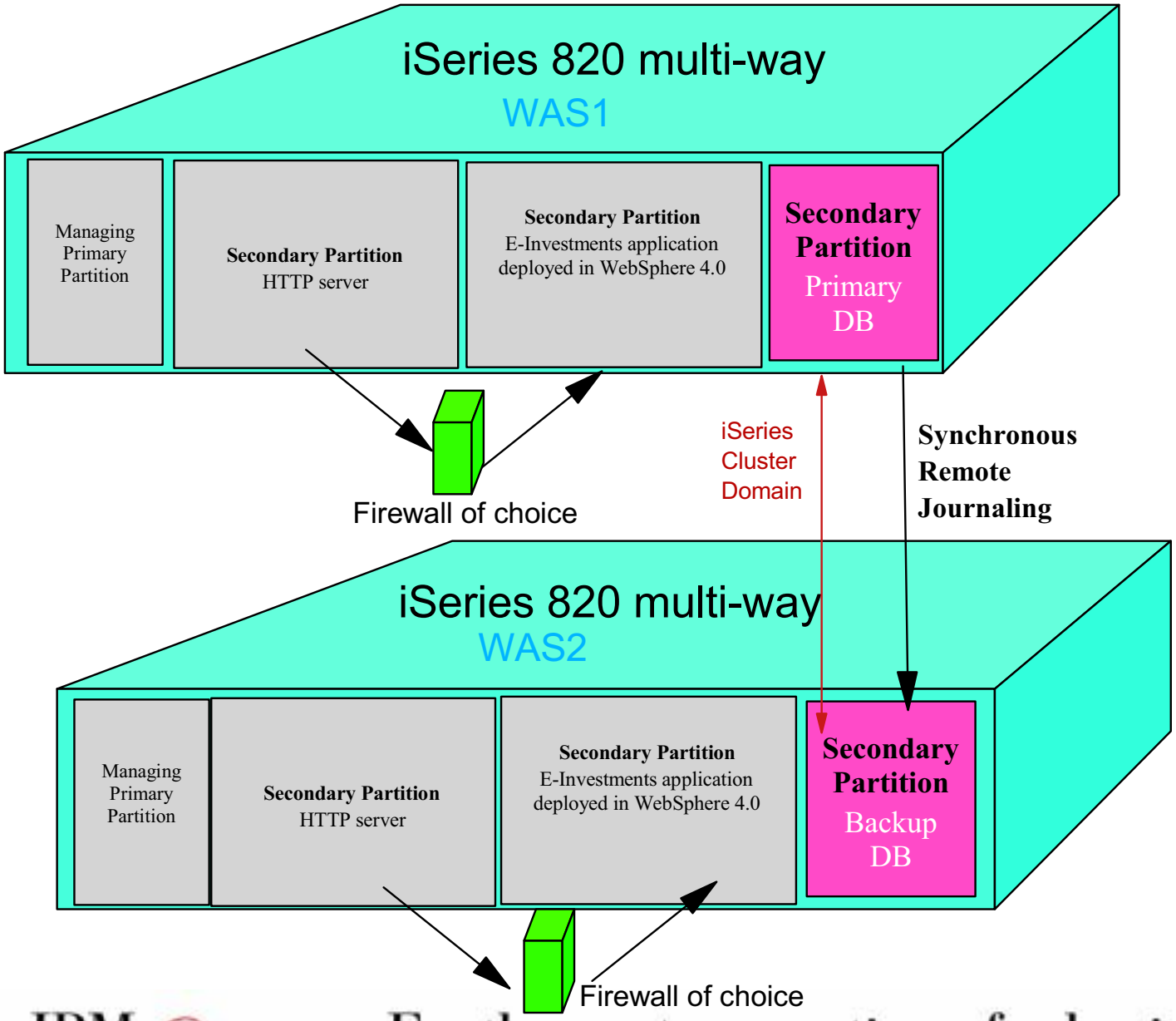
Topology I - LPAR WAS1/WAS2 systems

IBM @server



Topology I - Another LPAR config option (Server Consolidation)

IBM @server



LPAR

IBM @server

- LPAR facilitates server consolidation thereby cutting costs. Resources - processors, memory, buses, I/O processors (IOPs), and I/O adapters (IOAs)
- Managing Primary - all partitions are dependent upon primary. Recommend using a managing primary with minimal amount of hardware resource.
- Note: only the Database partition is in a native iSeries cluster
- Redbook - LPAR Configuration and Management - Working with iSeries Logical Partitions
  - SG24-6251-00

WebSphere Edge Server...

- Edge servers reside at the edge of the network and provide..
  - Caching and filtering
  - Load balancing component, known as **Network Dispatcher**, is a server that is able to dynamically monitor and balance TCP servers and applications in real time. It improves a Web site's availability, scalability and performance by transparently clustering edge, Web and application servers. It enables multiple TCP servers to be dynamically linked in a single entity that appears to the network as a single logical server. Three routing methods:
    - MAC forwarding - Dispatcher receives the client's request packet and translates the destination MAC address of the packet from the Dispatcher MAC address to the load balanced server's MAC address. Very efficient, but Web server must be on same IP subnet as the dispatcher.

WebSphere Edge Server...

- Routing methods continued:
  - MAC forwarding (Continued) - Server affinity is assured via sticky bit time configured via the dispatcher. If set to 30 minutes then any request from a unique IP address will be routed to that the same server for that time duration.
  - NAT/NATP forwarding - Methods allow dispatcher to send client requests not only to servers in the same IP subnet, but also to servers in a different IP subnet or IP network. Also a firewall can be inserted between the dispatcher and Web server.
    - NAT (Network Address Translation)
    - NAPT (Network Address Port Translation) - changes not only destination IP address, but also port. Works well with multiple Web server daemon on the same machine.
    - Affinity - "Stickiness" to source IP

WebSphere Edge Server...

IBM @server

Routing methods continued:

- Kernel content-based routing (cbr).
  - The information to make load balancing decisions may be in the application-defined data that flows between the client and the back-end server. NAT used in Kernel mode.
  - Dispatcher content-based routing forwarding method
  - Server partitioning
  - Enhanced Advisor Request/Response (URL) enhancement
- Server affinity - allows load balancing for those applications that need to preserve state across distinct connections from a client.
  - Affinity features
    - "Stickiness" to source IP address (MAC and NAT/NAPT routing)
    - SSL session ID (CBR)
    - URI (CBR)

WebSphere Edge Server...

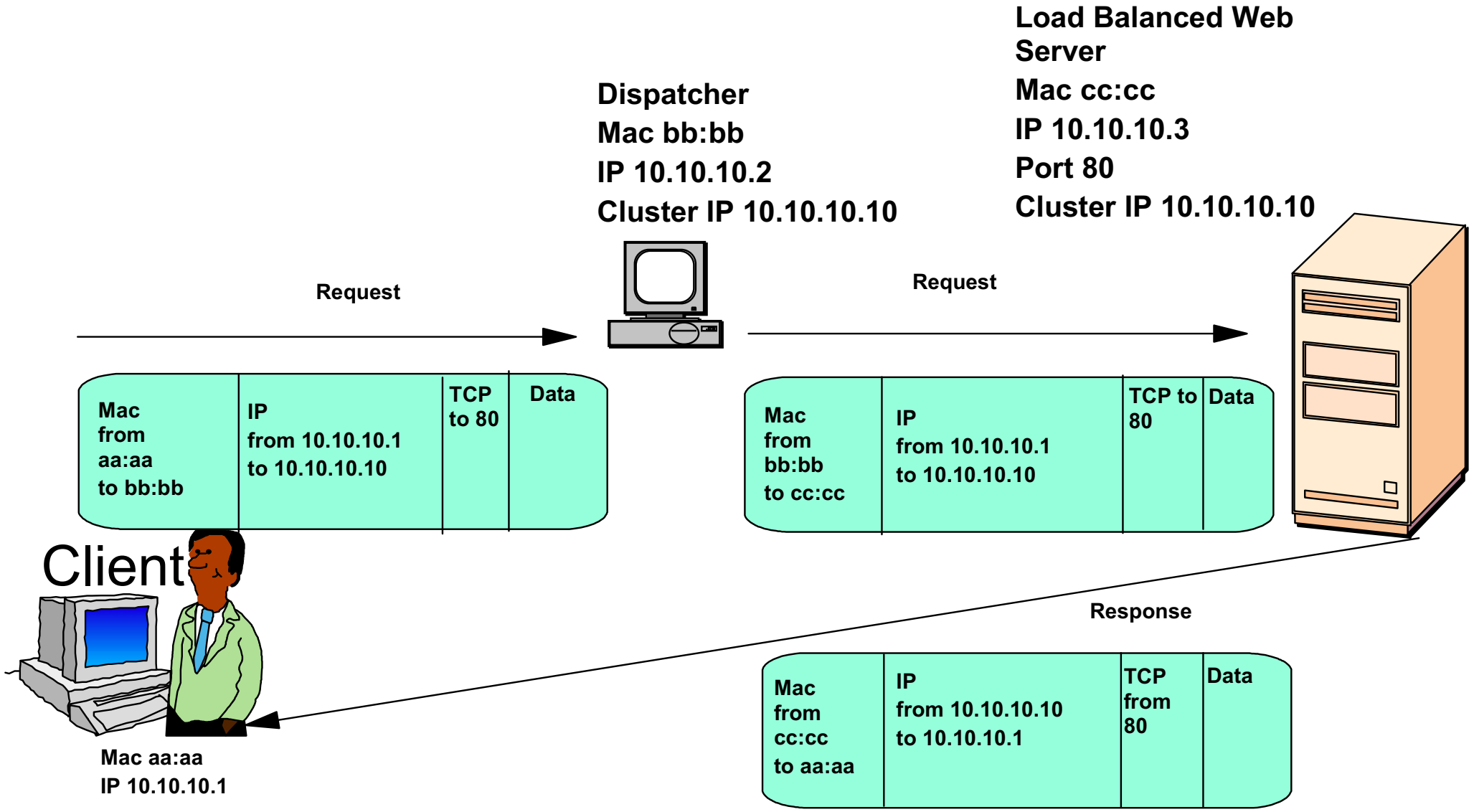
IBM @server

- Server affinity (Continued)
  - Passive cookie (CBR)
  - Active cookie (CBR)
- The Dispatcher's high-availability feature involves the use of a secondary machine that monitors the main, or primary, machine and stands by to take over the task of load balancing should the primary machine fail at any time. It has content distribution and server monitoring capabilities.

WebSphere Edge Server...

Packet Flow for Dispatcher MAC address translation

IBM @server

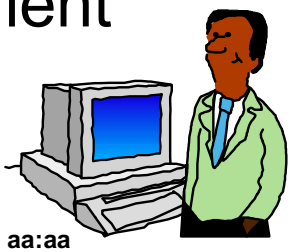


WebSphere Edge Server...

IBM @server

Packet Flow for NAT routing method

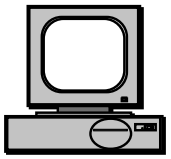
Client

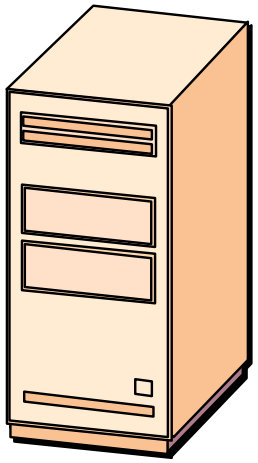


Mac aa:aa
IP 10.10.10.1

Dispatcher
Mac bb:bb
IP 10.10.10.2
Cluster IP 10.10.10.10
Return IP 10.10.10.20

Load Balanced Server
Mac cc:cc
IP 10.10.10.3
port 80

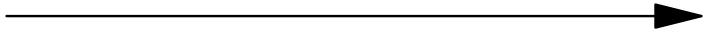




Request

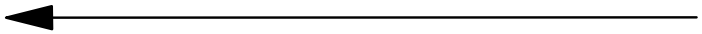
Mac	IP	TCP	Data
from aa:aa to bb:bb	from 10.10.10.1 to 10.10.10.10	to 80	

Request



Mac	IP	TCP	Data
from bb:bb to cc:cc	from 10.10.10.20 to 10.10.10.3	to 80	

Response



Mac	IP	TCP	Data
from bb:bb to aa:aa	from 10.10.10.10 to 10.10.10.1	from 80	

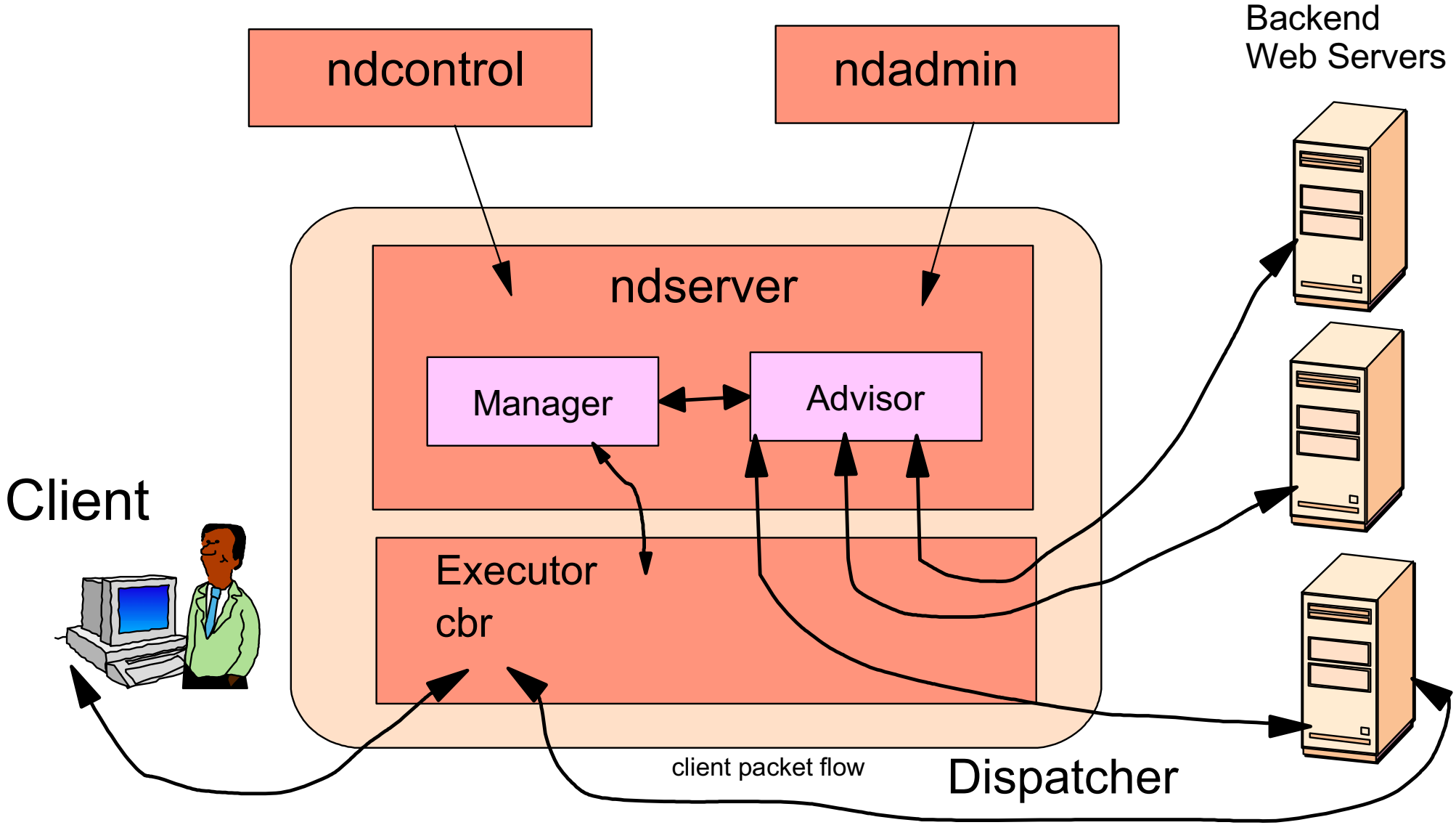
Response

Mac	IP	TCP	Data
from cc:cc to bb:bb	from 10.10.10.3 to 10.10.10.20	from 80	

WebSphere Edge Server...

IBM @server

Dispatcher content-based routing



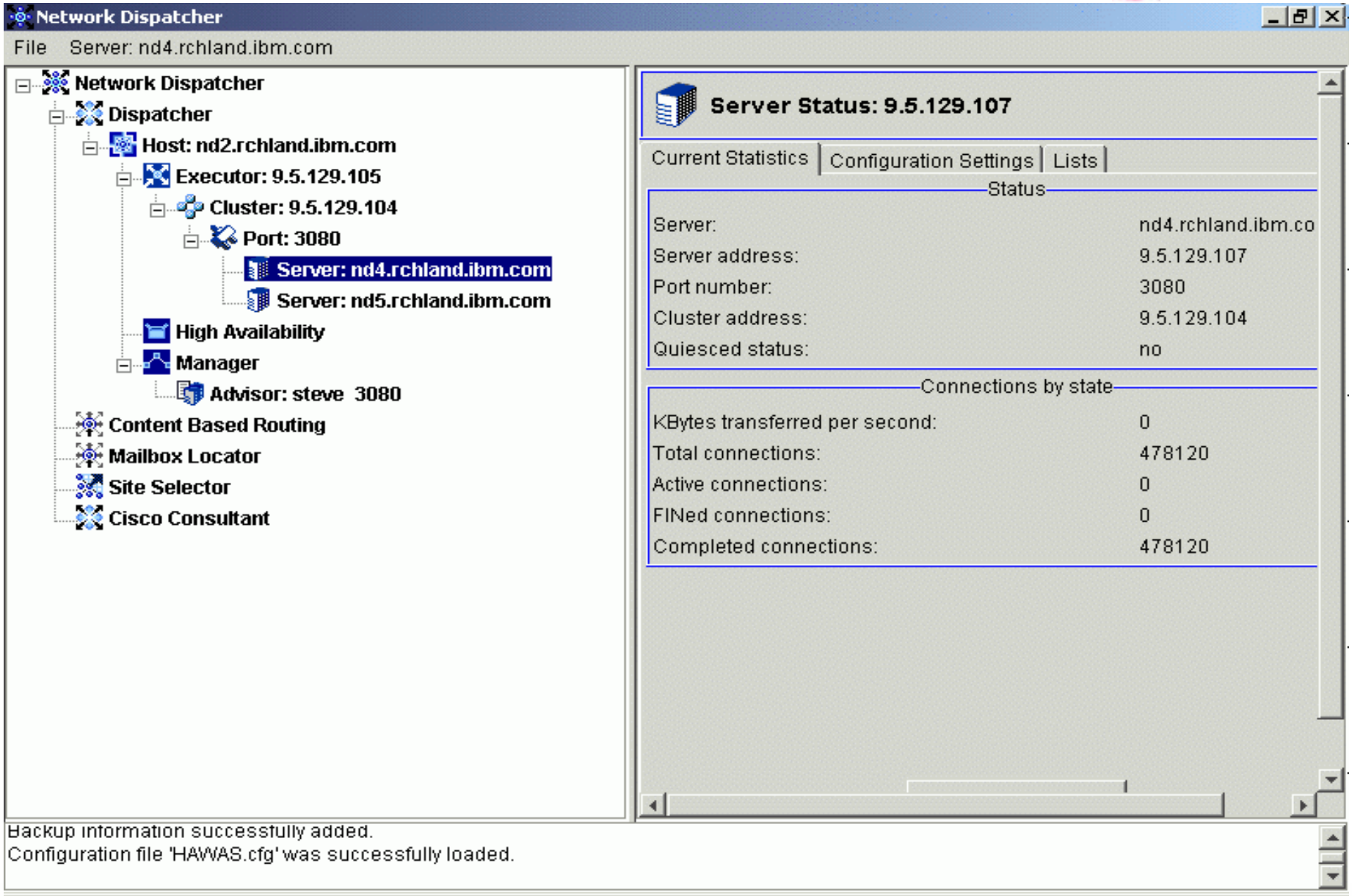
WebSphere Edge Server

IBM @server

- Custom Advisors
  - Provide wellness monitoring for various critical components in the application path
  - Will inform Network Dispatcher to not direct requests down a failed server path
  - Even when server/application is down, the advisor keeps monitoring wellness and will direct client requests as soon as the failed server/application is available after repair actions
- Redbook - WebSphere Edge Server New Features and Functions in Version 2 SG24-6511-00
- Redbook - IBM WebSphere V4.0 Advanced Edition Scalability (Chapter 4 - Adding Web server load balancing) SG24-6511-00
- Writing Custom Advisors for IBM Network Dispatcher - http://www-106.ibm.com/developerworks/library/ibm-cust/

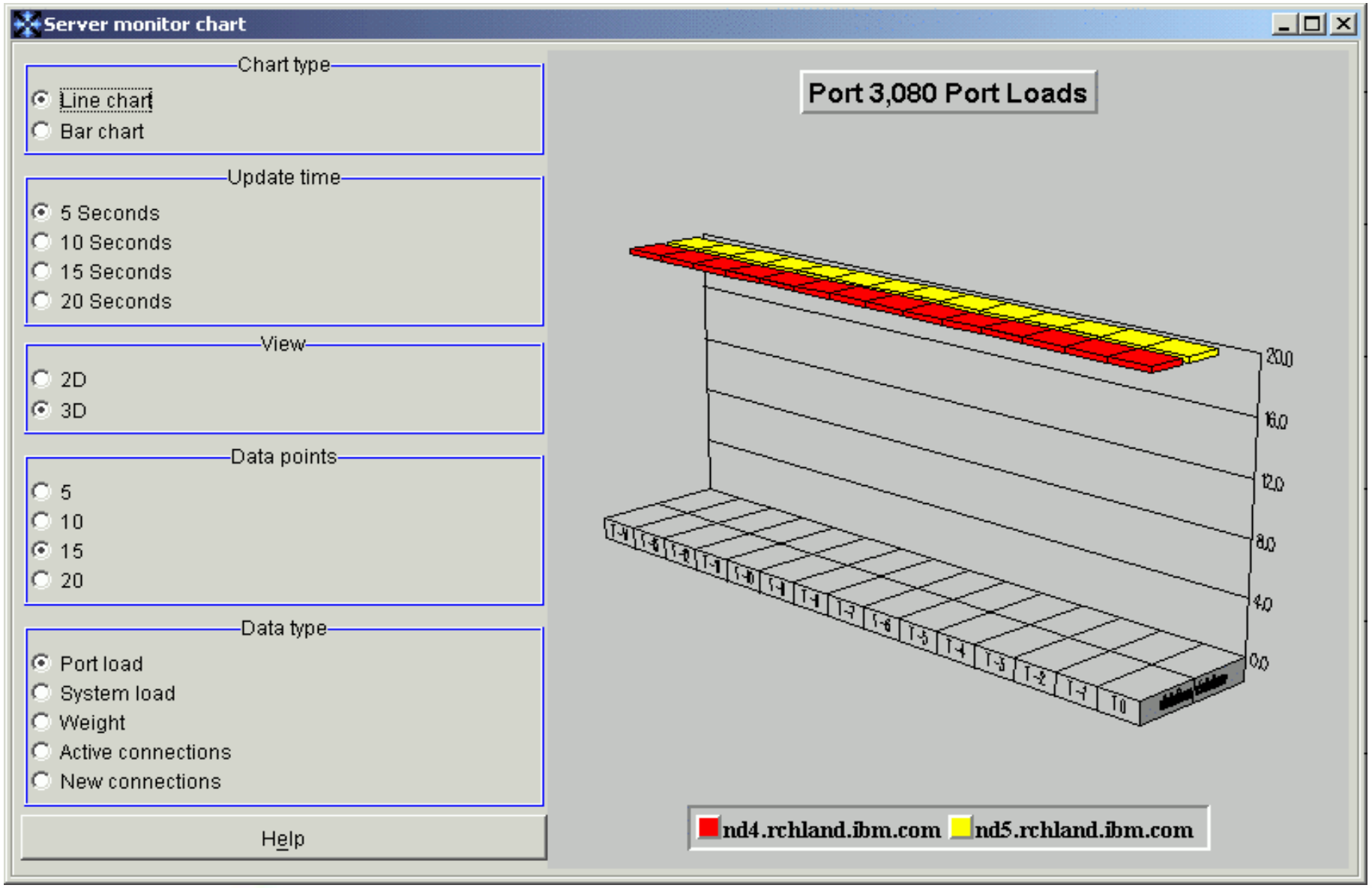
Network Dispatcher Manager Screen shot

IBM @server



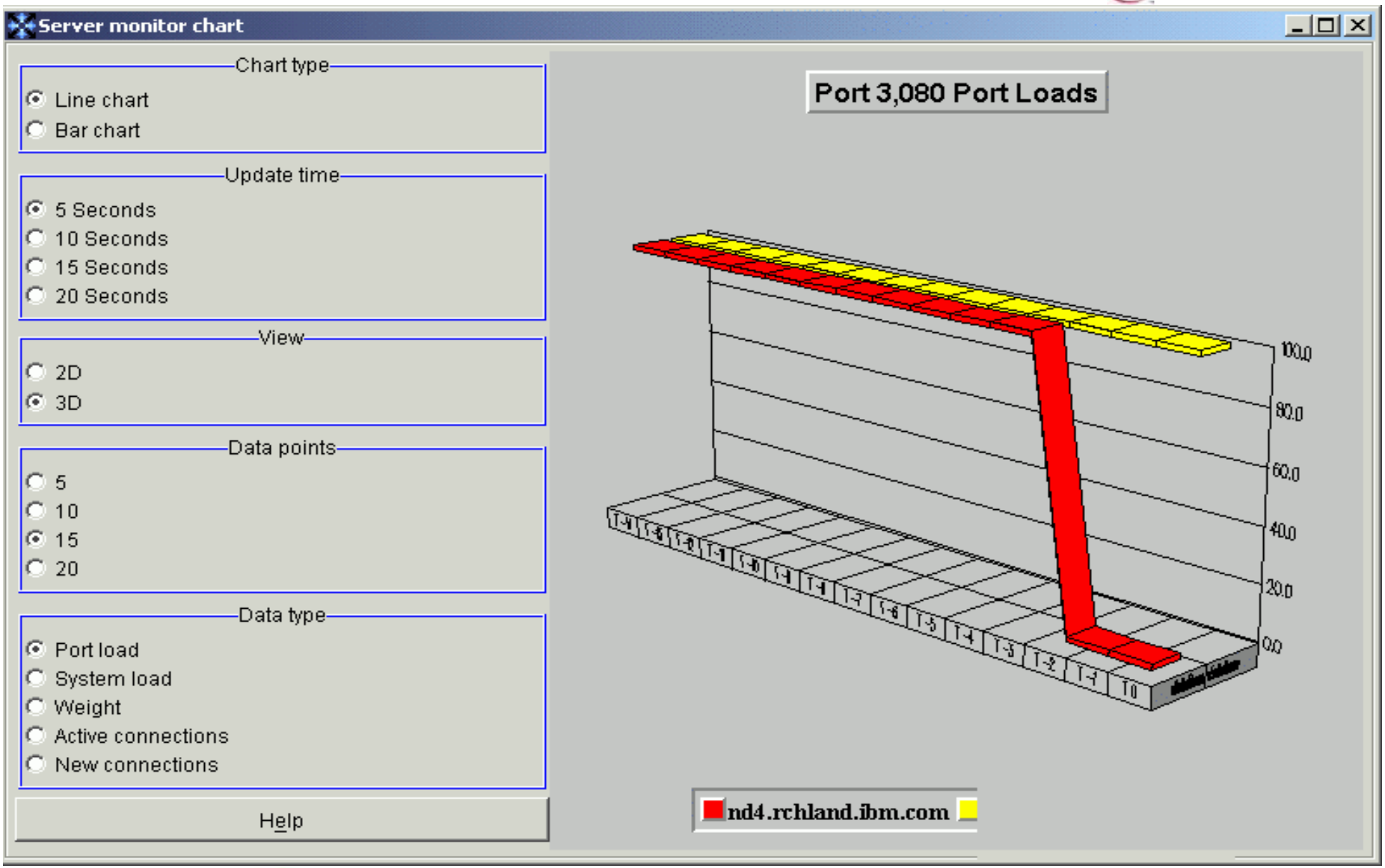
Network Dispatcher Monitor - All is Well

IBM @server



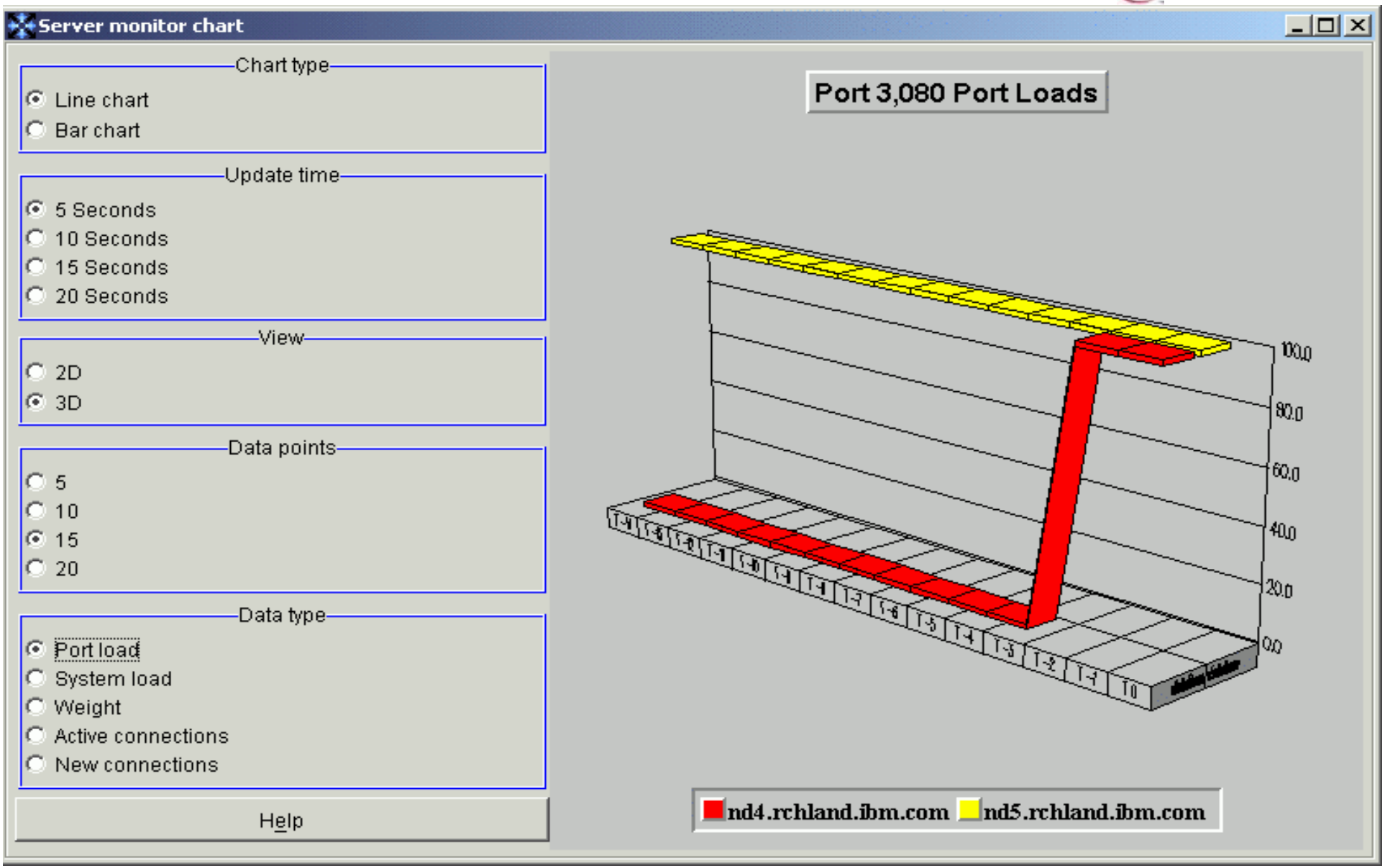
Network Dispatcher Monitor - Failure detected by Custom Advisor

IBM @server



Network Dispatcher Monitor - Custom Advisor senses repair action complete

IBM @server



High Availability HTTP (Powered by Apache) Server

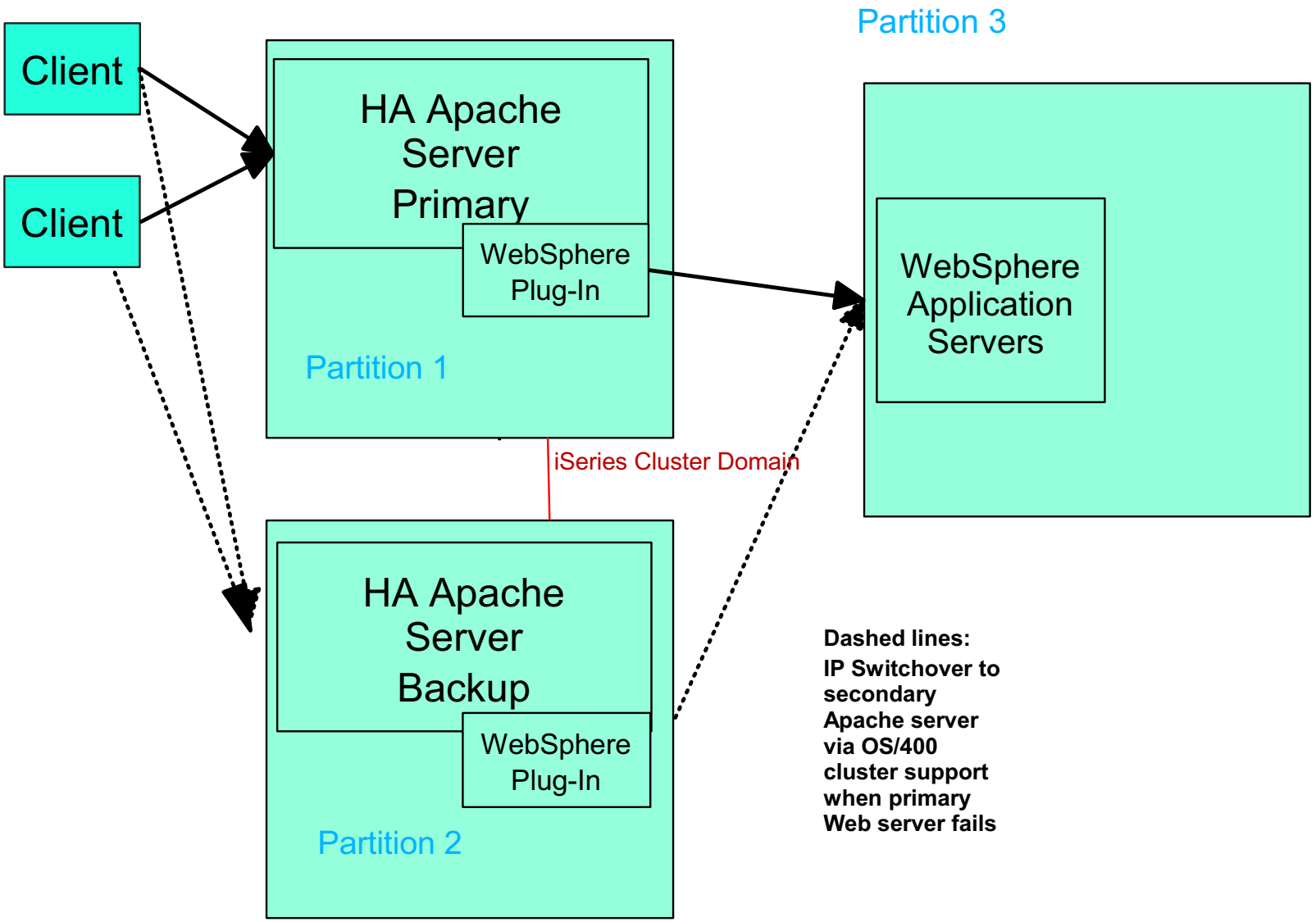
IBM @server

- Could be used as an alternative to Edge Server, but does not provide the same level of high availability for Java servlets
- Works with CGI programs
- Apache does not have a Custom Advisor type function
  - Not able to send URL request and specify unique text returned.
    - When WebSphere application servers are down, clients may encounter a blank page
- WebSphere does not use CHT (Clustered Hash Table), but instead servlet session data is persisted to a backend HA database.
- Best when used in conjunction with Edge Server
- More information on HA Apache can be found at:
  - http://publib.boulder.ibm.com/html/as400/v5r1/ic2924/index.htm?info/rzaie/rzaiehighavailability.htm

High Availability HTTP (Powered by Apache) Server

HTTP Failover option - Primary backup

IBM @server



WebSphere - Clustering WLM (Workload Management)...

IBM @server

- Workload management involves spreading multiple client requests for work over the cloned WebSphere resources. Workload mgmt provides better availability and scalability of your WebSphere resources. Types of resources and requests are
  - Application servers - Servlet requests
  - EJB containers - Enterprise JavaBeans requests
  - Admin servers - (Name resolution)
- Clients applications denote Web and/or (thick/thin) Java clients
- Clones can be distributed on one machine (vertical cloning) or multiple machines (horizontal cloning).
- Redbook - IBM WebSphere v4.0 Advanced Edition Scalability - Chapters 5, 6 and 7 - SG24-6511-00

WebSphere - App server (Workload Mgmt.)

IBM @server

- Stateful HTTP server state
  - Session manager (within servlet engine) preserves HTTP session state between dependent client requests.
    - Kept in memory within local JVM, and can be...
    - Written to backend database (persistent session data)
  - WLM load distribution facility honors server/session affinity
    - Use of persistent sessions will always work and provides failover. State can be shared amongst all cloned servers. There is performance degradation, but caching implemented by the Session Manager minimizes the impact for simple retrievals of session data.

WebSphere - EJBs (Workload Mgmt)...

IBM @server

- EJBs (Enterprise JavaBeans)
  - Can be WLMed. (Load balanced among cloned EJB instances).
    - Home interface object of either stateless or stateful session EJBs as well as entity EJBs.
    - Stateless session bean - there is no client-visible state associated with the bean. Each client request is independent of any previous request.

WebSphere - EJBs (Workload Mgmt)...

IBM @server

- Can be WLMed (Continued).
  - Entity beans - Notion of a session is replaced by the notion of a transaction. For the duration of one client transaction to which it participates, the entity bean is instantiated in one container (normally the container where the first operation within that transaction was initiated). All subsequent accesses to that same bean, within that same transaction, must be performed against that same instance in the same container. Between transactions, the handling of the entity bean is specified by the EJB specification in the form of caching options. Workload mgmt uses the concept of transaction affinity to direct requests for entity beans. Note: work load mgmt is only supported when option C caching is enabled in the container.

WebSphere - EJBs (Workload Mgmt)...

IBM @server

- Cannot be WLMed.
  - Stateful session bean instance - state persists across client requests. Same bean instance located in container needs to be accessed each time.
  - Entity beans when option A caching is used within owning container.

WebSphere - Admin Servers (Workload Management)...

IBM @server

- Administrative Servers
  - Can be cloned - are cloned by default in 4.0 to avoid single point of failure
  - com.ibm.ejs.sm.adminServer.wlm=true in admin.properties file.
- If running with one WebSphere domain, then locate the admin repository on your HA DB system and replicate.
- Administrative clients which need access to the admin server
  - Administrative Console - adminclient RCHAS841 900
    - If failure occurs, then another bootstrap port is used. Possible to write XML script to do this
  - Thin Java client - client must be coded to bootstrap to another admin server port in a failure scenario.

WebSphere - Admin Servers (Workload Management)...

IBM @server

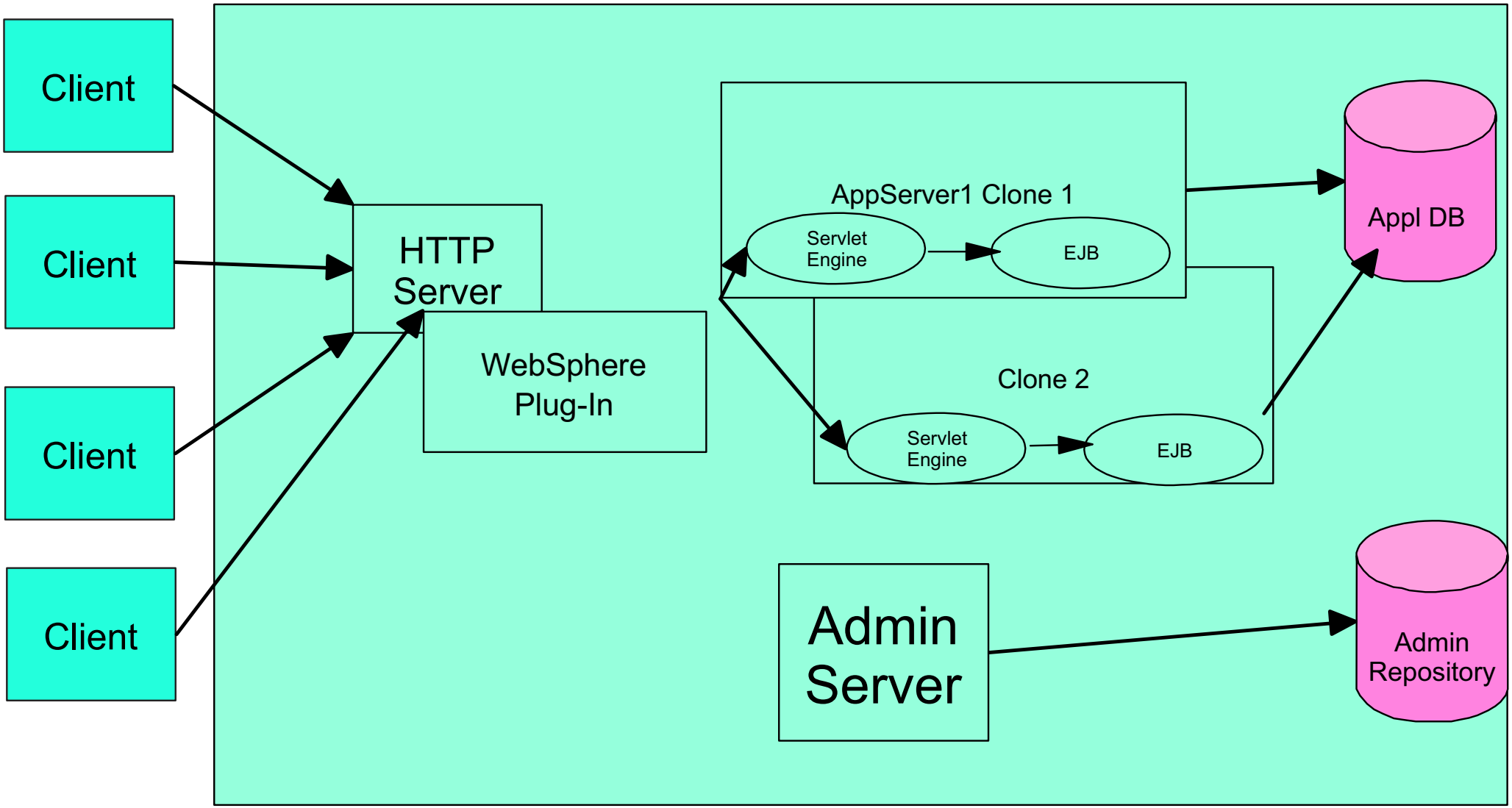
- Administrative server clients (Continued)
  - J2EE Java client - started via launchclient
    - if started, then the client has the necessary information to do automatic failover
  - Application server - failover automatic once server is started
  - Security requests - failover automatic once the containing application server is started.
- Redbook IBM WebSphere V4.0 Advanced Edition Scalability
  - SG24-6192-00 Chapter 7. Administrative server failover

WebSphere - Vertical Cloning - Single system...

Load Balancing and Failover

IBM @server

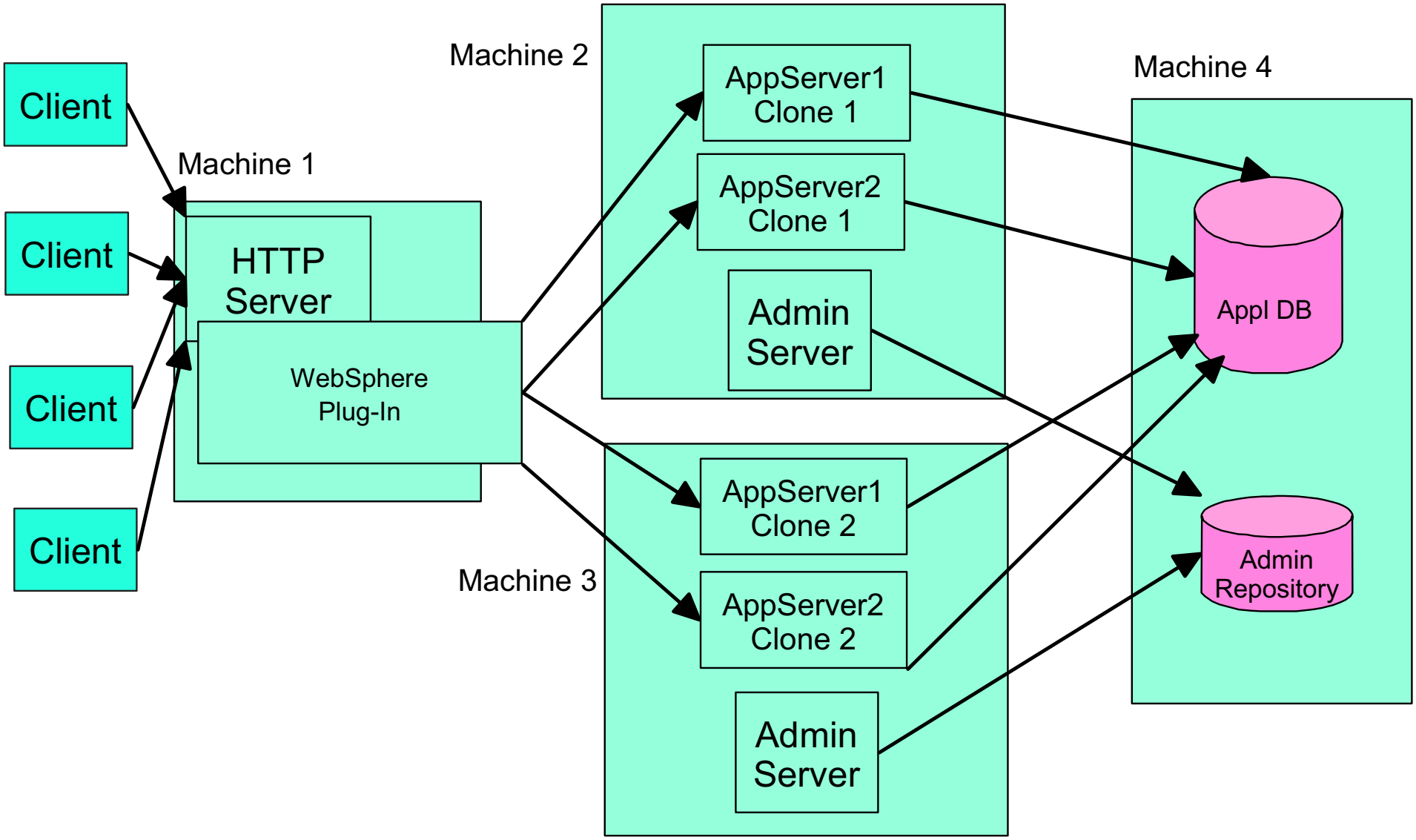
Machine or Partition A



WebSphere - Horizontal Cloning - Multiple systems...

Load Balancing and Failover+

IBM @server

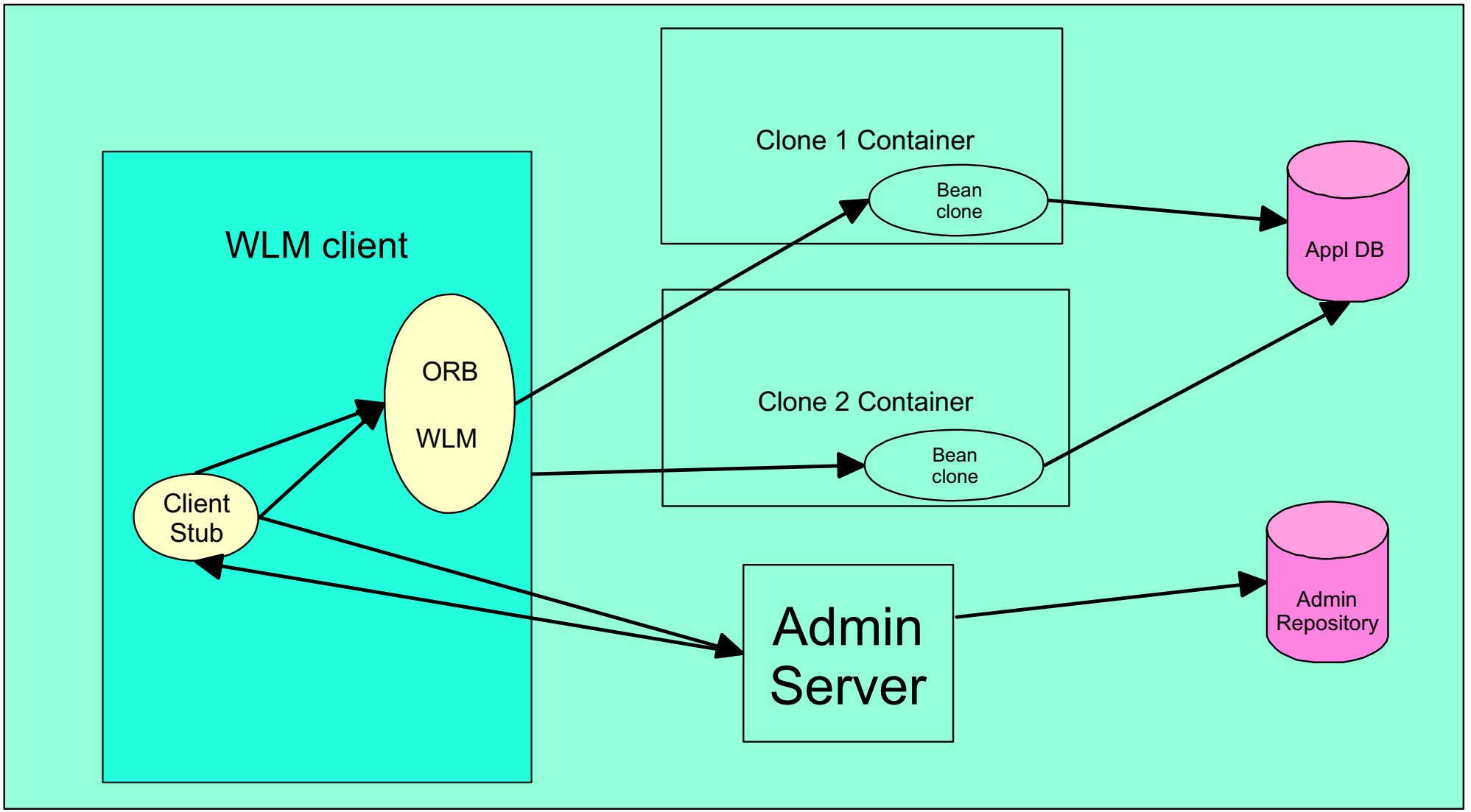


WebSphere - EJB Cloning...

Load Balancing and Failover

IBM @server

Machine A



WebSphere... HA Configuration used...

IBM @server

- Two nodes running identical WebSphere applications
  - Each node runs a separate WebSphere domain
  - Not horizontally cloned
  - Allows one node to be taken offline without noticeable interruption of service for planned maintenance activities
  - Could use just one domain and horizontal cloning, but the disadvantage here is downtime for some planned maintenance activities.
- WLM
  - iSeries scales well vertically
    - vertical versus horizontal needs vary among server platforms
  - Each node utilizes vertical cloning
    - application servers

WebSphere... HA Configuration used

IBM @server

- WLM (Continued)
  - Admin repository
    - each WebSphere node has its own repository

WebSphere - HA Configuration

IBM @server

- Persistent sessions
  - Session data stored on backend HA DB
- Data source location is defined by IP address
  - Works well with clustering and IP takeover
    - application data
    - persistent session data

WebSphere - Connection Manager...

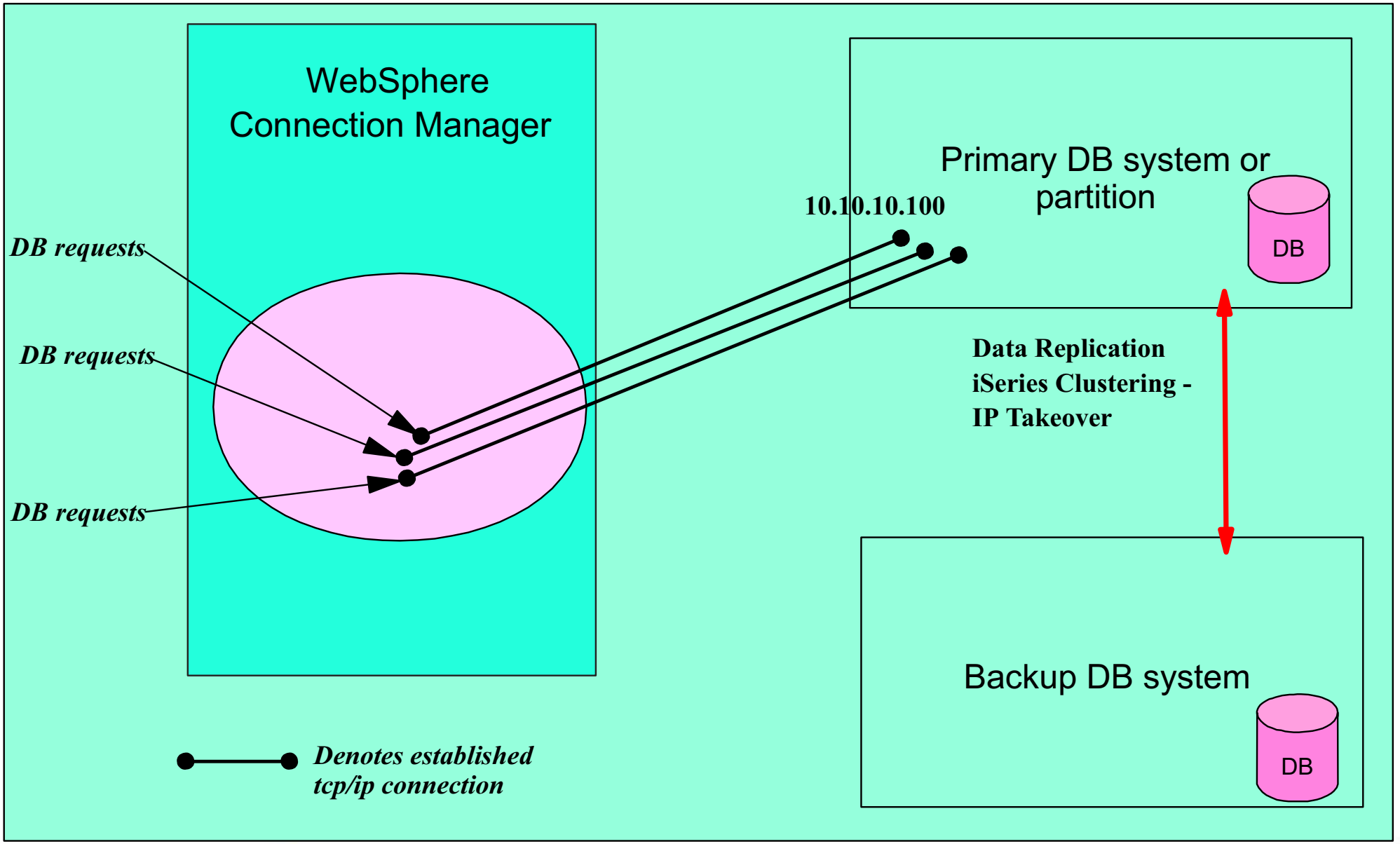
IBM @server

- So what happens with the primary DB during servlet or EJB database access? In order to answer this question one needs an understanding of the WebSphere Connection Manager.
- Each time a client (servlet or EJB) attempts to access a database, it must acquire a connection to the database (via JDBC provider). The connection made by the connection manager is a TCP/IP sockets connection. A new connection can be costly, so the WebSphere Connection Manager allows the administrator to create a data source that has a pool of database connections that can be shared by applications. Prepared SQL statement cache coupled with the Connection Manager provides performance benefits for applications. Another aspect of the Connection Manager is that it provides failover semantics when the DB fails.

WebSphere - Connection Manager...

*Interaction with backend HA DB - **All is well***

IBM @server



WebSphere - Connection Manager...

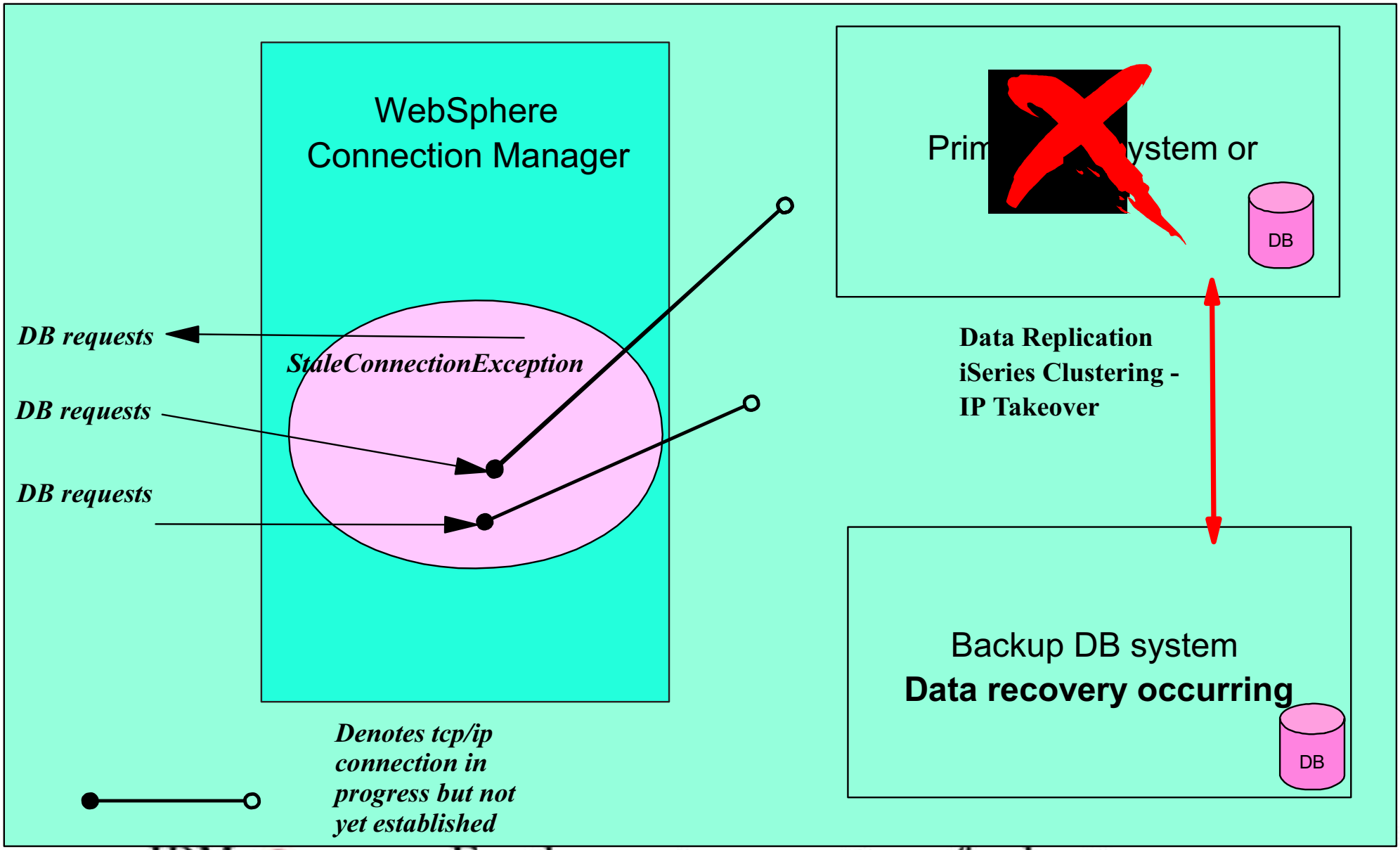
IBM @server

- When the backend database fails, one client per application server will encounter a stale connection exception. When this happens the entire pool of connections is destroyed and a new pool will be created to handle subsequent client requests.
  Note:StaleConnectionException is inherited from SQLExceptions. Most applications have catch blocks which already handle SQLExceptions
- The WebSphere connection manager, upon encountering a stale connection, will flush all DB connections in the pool and attempt to reestablish a connection with the backend database. This attempt will appear to wait because the IP address on the failed primary has been deactivated by iSeries cluster services

WebSphere - Connection Manager...

*Interaction with backend HA DB - **Database fails***

IBM @server



WebSphere - Connection Manager

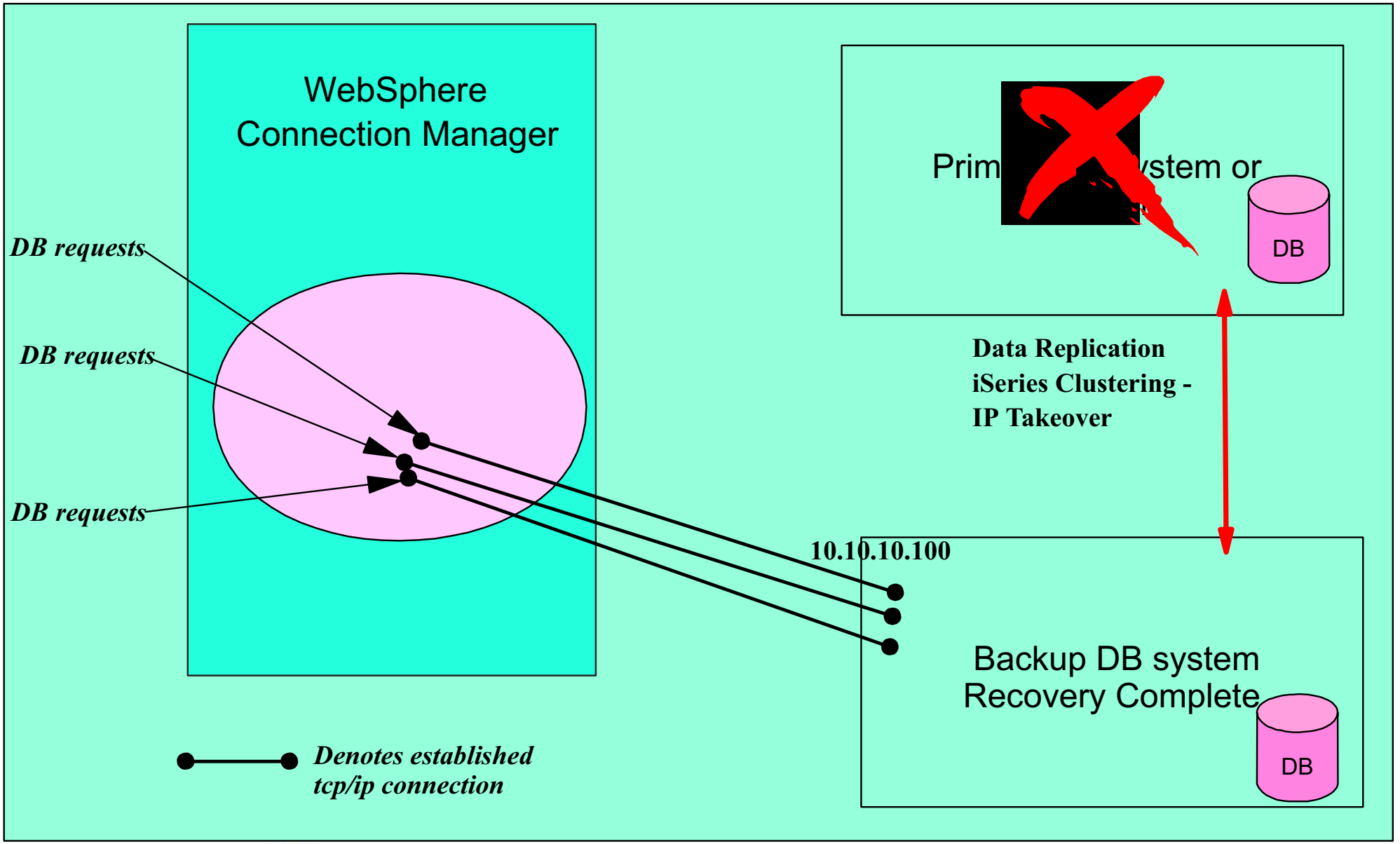
IBM @server

- At some point the backup DB node is ready and the IP address becomes active (IP takeover). At that time an ARP is sent out causing all ARP caches within the network to be updated with the backup DB MAC address
- The thread attempting the DB connection succeeds and most clients, if patient, may notice no failure. Keep in mind this outage may be several minutes; therefore, some clients may select 'stop' followed by 'reload' in their browsers.

WebSphere - Connection Manager

*Interaction with backend HA DB - **Backup is ready***

IBM @server



iSeries HA Database

IBM @server

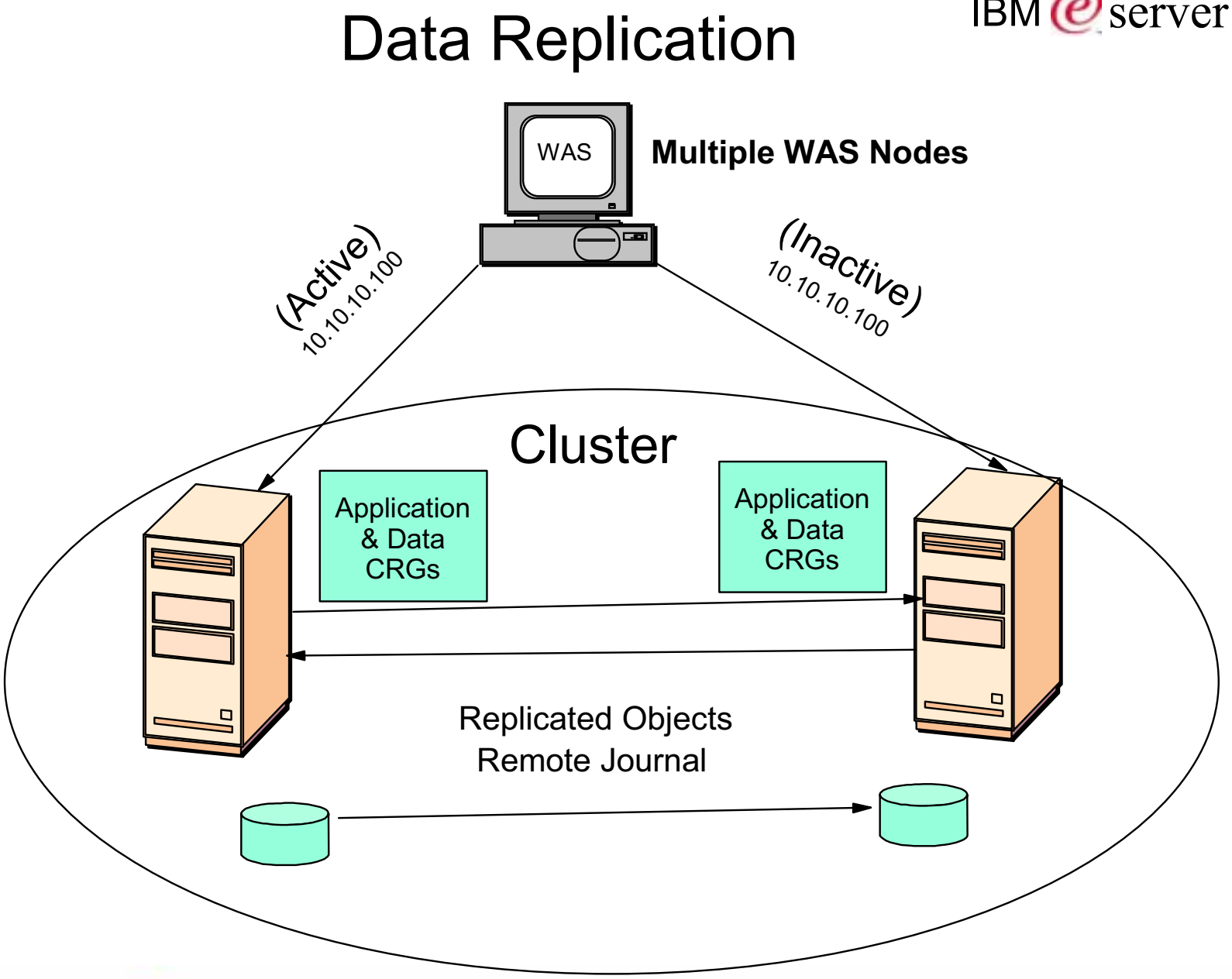
- Application resiliency
  - handled by WebSphere
- Data resiliency
  - Data replication
  - Switch disk

iSeries HA Database Replication

IBM @server

- iSeries Cluster framework
- Remote Journal technology
  - Assures hot backup
- Cluster Middleware Business Partner
  - Cluster Management
  - Configuration of application
  - Replication of data
    - apply of data from remote journal receiver
    - synchronous remote journaling used in our test environment
  - Switchover and failover management
    - Switchover - planned switch
    - Failover - unplanned - iSeries cluster framework defines a list of events which would cause a failover to occur.

Pictorial view (Data Replication)



DataBase Replication - iSeries Clustering

IBM @server

- Uses a peer architecture where each node has all of the characteristics of the cluster. All related nodes comprise a recovery domain.
- Cluster Resource Group Manager synchronizes services across all affected nodes within the cluster
- A request for cluster action can be initiated from any active node in the cluster
- Heartbeat monitoring determines if a node is active
- Cluster resource services manages IP takeover switching
- Cluster Resource Group - CRG - Provides a single resource view across all nodes within the recovery Domain
  - Data CRG - controls data replication activity for each application via an exit program written by the cluster middleware partner

DataBase Replication - iSeries Clustering

IBM @server

- Cluster Resource Group - (continued)
  - Application CRG - controls application via an exit program
  - Device CRG - controls switchable towers
- Within our topology both WAS domains are always active at all times. The application CRG at switchover or failover ensures that the corresponding data CRG is complete before activating the IP address on the backup. An example application CRG program that may be used for loosely-coupled WebSphere applications is available.
- At switchover the sample application CRG exit program will, on the former primary, release locks held on all files listed in the cluster object specifier file. This avoids a deadlock when a switchback to the former primary occurs.

Database Replication - Remote Journal

IBM @server

- Remote Journal
  - Preferred mechanism used to replicate data from primary to backup system
  - Synchronous and asynchronous modes supported
  - Provides disaster recovery - systems can be on different power domains or even in different countries
  - Remote journal performs very well (iSeries remote journal is state-of-the-art implementation)
    - transaction state enforced versus mirrored device approach
    - microcode-to-microcode transfer; application layer not involved
    - data harvesting of remote journal receiver keeps the backup DB current to assure fast failover
    - cluster middleware will resync the data on the failed node once back online (see following foils)
    - Journal Batch Caching (PRPQ 5799-BJC in V5R1, priced optional feature 42 in V5R2)

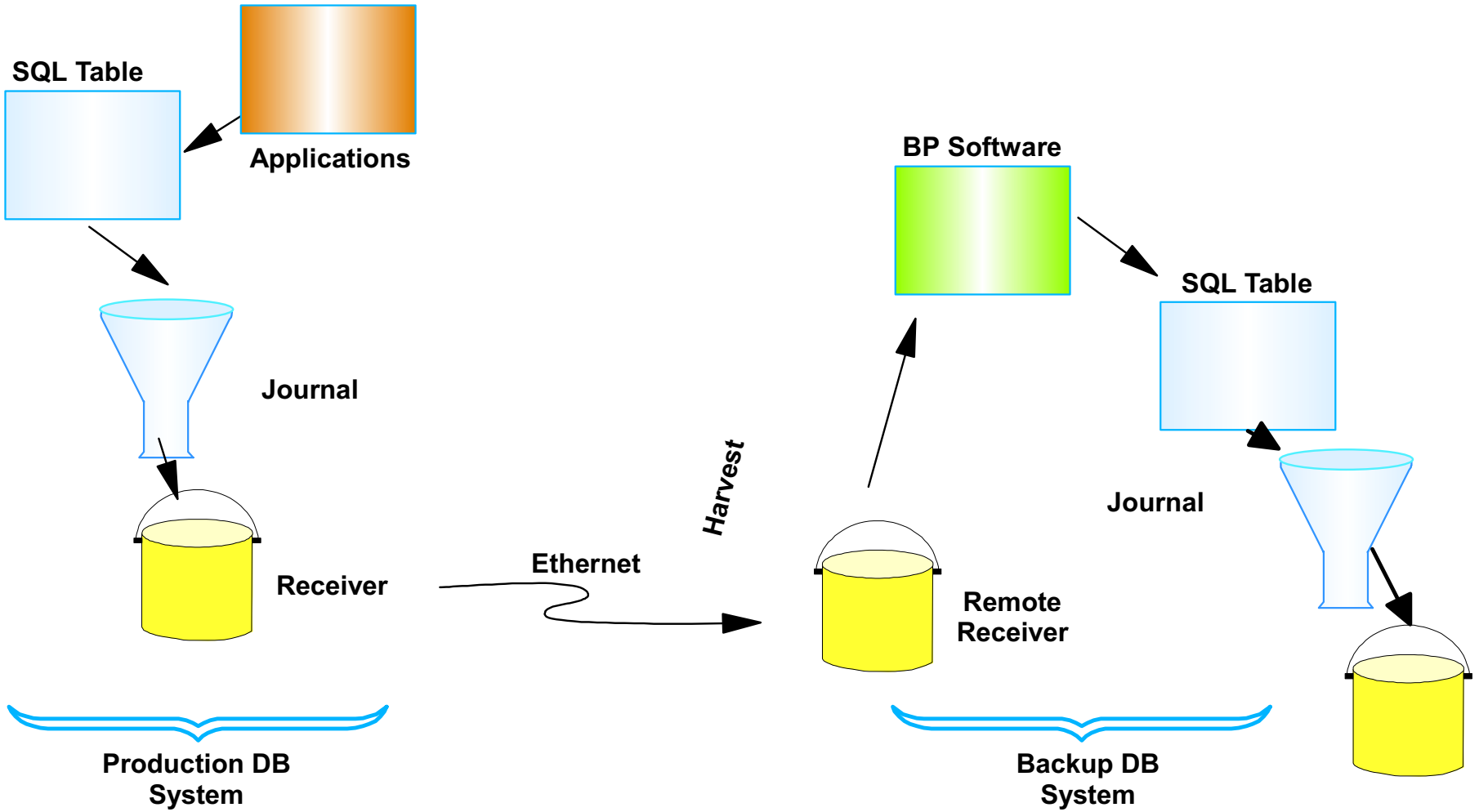
New Journal Performance Redbook

IBM @server

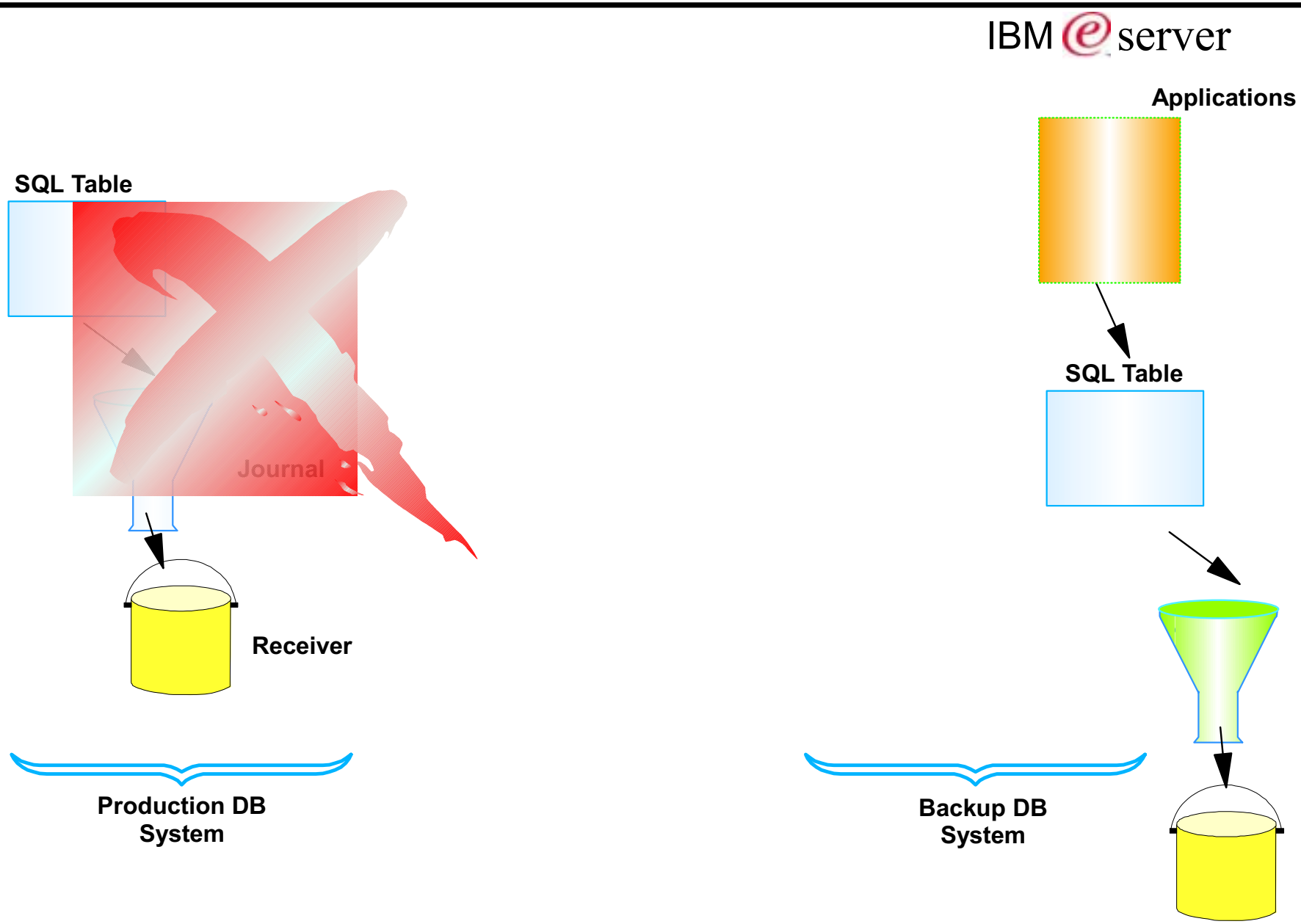
- iSeries conducted journal performance studies recently
- The results are documented in a resulting **Redbook:**
  - **Striving for Optimal Journal Performance**
  - Order number
    - **SG24-6286-00.**
  - Web Site:
    - **http://www.redbooks.ibm.com**
      - **Then search for the word: "Striving"**

Database Replication - Remote Journal

IBM @server

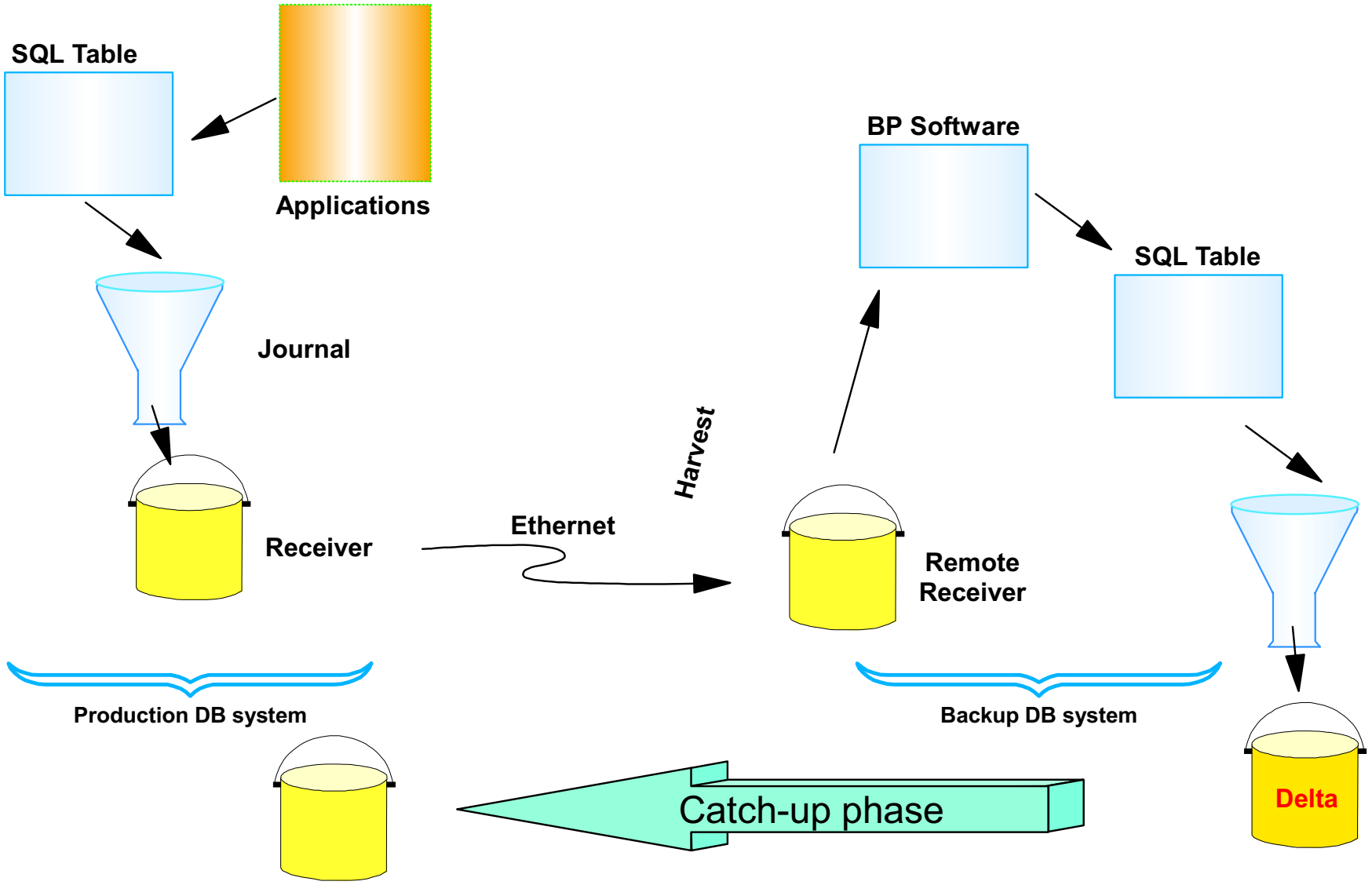


Database Replication - Remote Journal - Failover

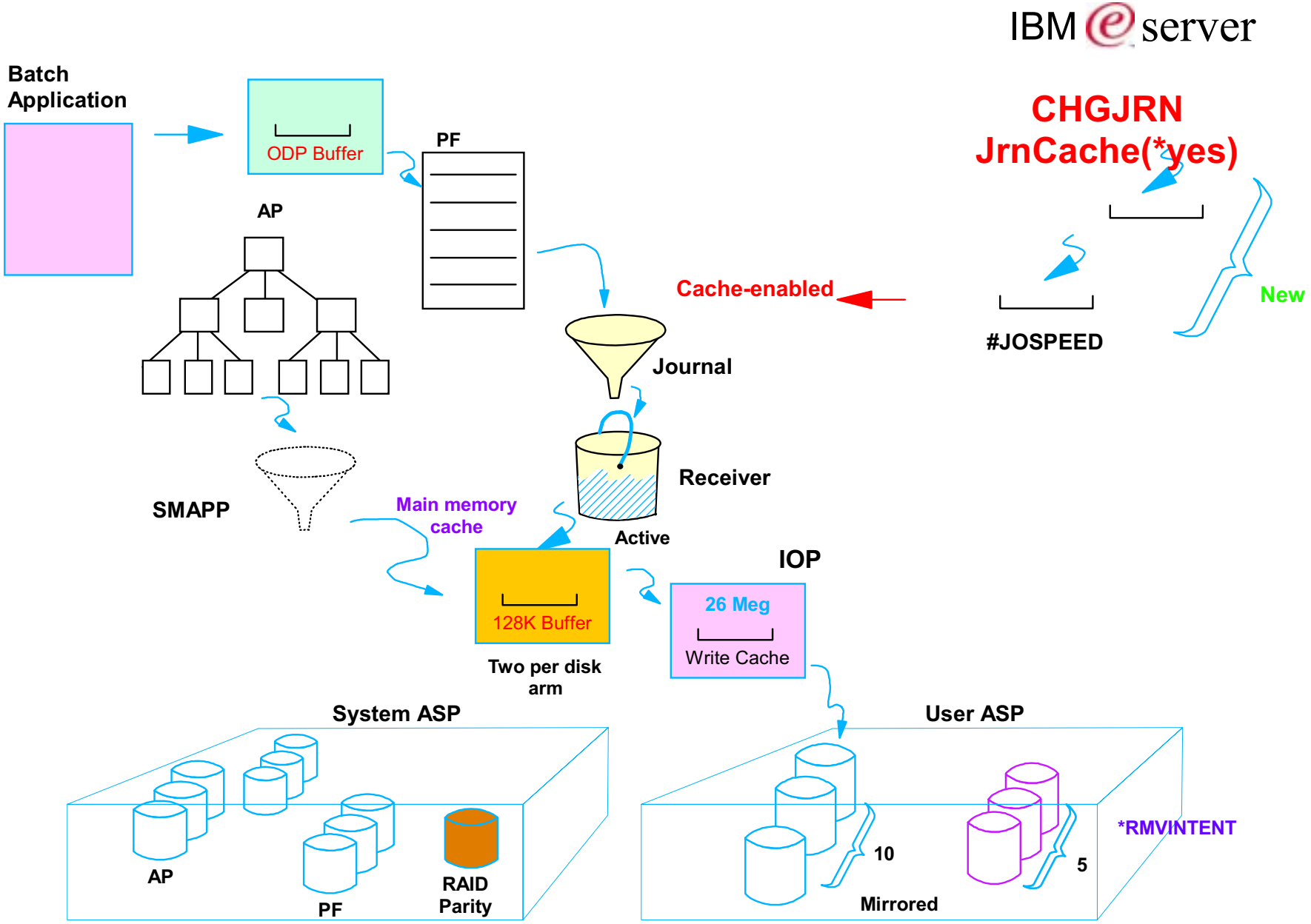


Database Replication - Remote Journal - Switch Back

IBM @server



Data Replication - Journal Caching:



Data Replication - Potential Performance Benefits of PRPQ

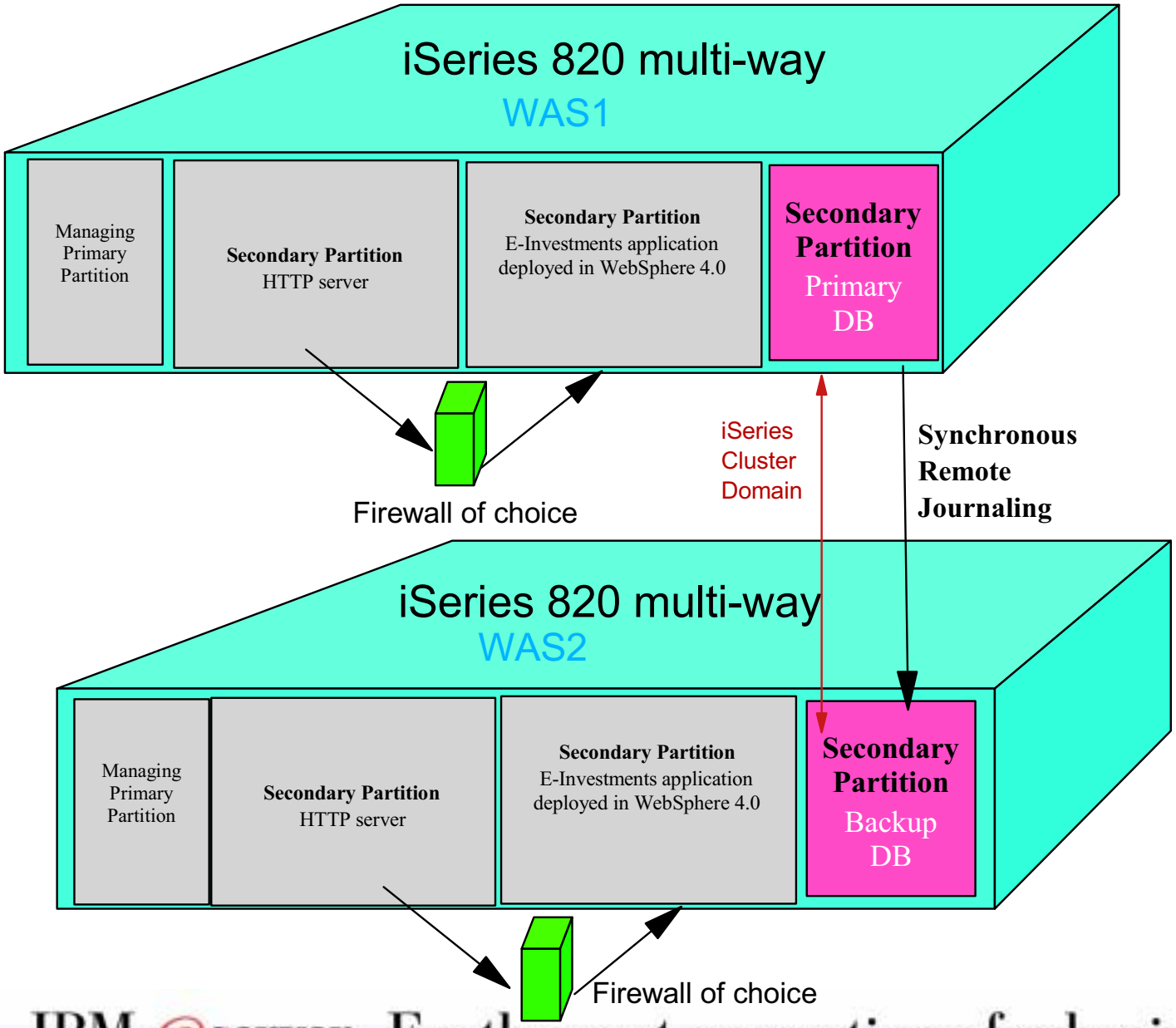
IBM @server

(For Target Machine's Keep-up Mode - critical for hot-backup)

	Apply rate on Target machine
Before PRPQ	600,000 transactions/Hr
With PRPQ on target	2,400,000 transactions/Hr

(For more detail see the Redbook mentioned on a previous page)

Recommended Implementation Order for Topology I ...



Recommended Implementation Order for Topology I ...

IBM @server

- Configure LPAR partitions on WAS1 and WAS2
  - Partitions
    - Primary (Managing/Thin)
    - Secondary (HTTP server)
    - Secondary (WebSphere)
    - Secondary (Database) -- Ensure enough disk arms are configured to meet the particular performance requirements
  - Redbook - LPAR Configuration and Management Working with IBM iSeries Logical Partitions - SG24-6251-00
- On WAS1
  - Move enterprise database to database secondary partition
  - Install WebSphere in Secondary WebSphere partition

    http://publib.boulder.ibm.com/was400/40/AE/english/docs/avindex2.html

    http://publib.boulder.ibm.com/was400/40/AE/english/docs/iic.pdf

Recommended Implementation Order for Topology I...

IBM @server

- On WAS1 (Continued)
  - Create a WAS1 WebSphere instance and deploy pertinent WebSphere application within the WebSphere secondary partition. Create vertical clones for application servers and enable persistent sessions. Install the IBM Toolbox for Java JDBC driver and create separate datasources to access your enterprise database and WebSphere persistent session database.
    - Deploy in each mission-critical application server an NDAdvisor servlet to be used for wellness monitoring by the Edge Server
    - Redbook - IBM WebSphere V4.0 Advanced Edition - SG24-6176-00 Chapter 17 - Server groups and workload mgmt.
  - Install WebSphere (Base Option) on HTTP server secondary partition - http://publib.boulder.ibm.com/was400/40/AE/english/docs/avindex2.html
  - Install Apache Web server on HTTP server secondary partition
    - Redbook - HTTP Server (powered by Apache): An Integrated Solution for IBM eServer iSeries Servers - SG24-6716-00

Recommended Implementation Order for Topology I...

- On WAS1 (Continued)
  - Create WebSphere WAS1 instance in the Web server secondary partition. Configure HTTP server instance, set **--ServerInit entry--** to your WAS1 instance
  - Copy the WebSphere secondary partition's \QIBM\UserData\WebASAdv4\WAS1\config\plugin-cfg.xml file to Web server secondary partition's \QIBM\UserData\WebASAdv4\WAS1\config\plugin-cfg.xml file
  - Start HTTP server
  - Test application
- Perform the same steps on WAS2 and in addition ...
  - Copy enterprise database from DB partition on WAS1 to DB partition on WAS2
  - Point datasource to database located on WAS1

Recommended Implementation Order for Topology I...

- Perform the same steps on WAS2 (Continued)
  - Utilize the services of an HABP (High-Availability Business Partner) company to replicate the database. Use modified application exit program supplied by HABP
  - Test and make sure no files go on hold when performing switchover and normal runtime activities while workload simulation testing is in progress
- Configure the Network Dispatcher
  - Add Web server cluster to the configuration. This cluster address will be the host/IP address used in the Web URLs
  - Add the HTTP port to the Dispatcher configuration

Recommended Implementation Order for Topology I...

- Configure the Network Dispatcher (Continued)
  - Add the HTTP Web server IP addresses to the Dispatcher configuration
  - On each Web server machine the Web server cluster address must be specified as a *VIRTUALIP address under CFGTCP option 1. This is pertinent for MAC forward routing.
  - Write a custom advisor to check wellness of mission critical application servers. Configure the frequency this advisor should be called. See custom advisor example supplied.
  - Ensure each mission-critical application server contains an NDAdvisor servlet. This servlet will be called by the custom advisor, and it will return a string denoting current status as defined by the custom advisor code. See example.

Recommended Implementation Order for Topology I

IBM @server

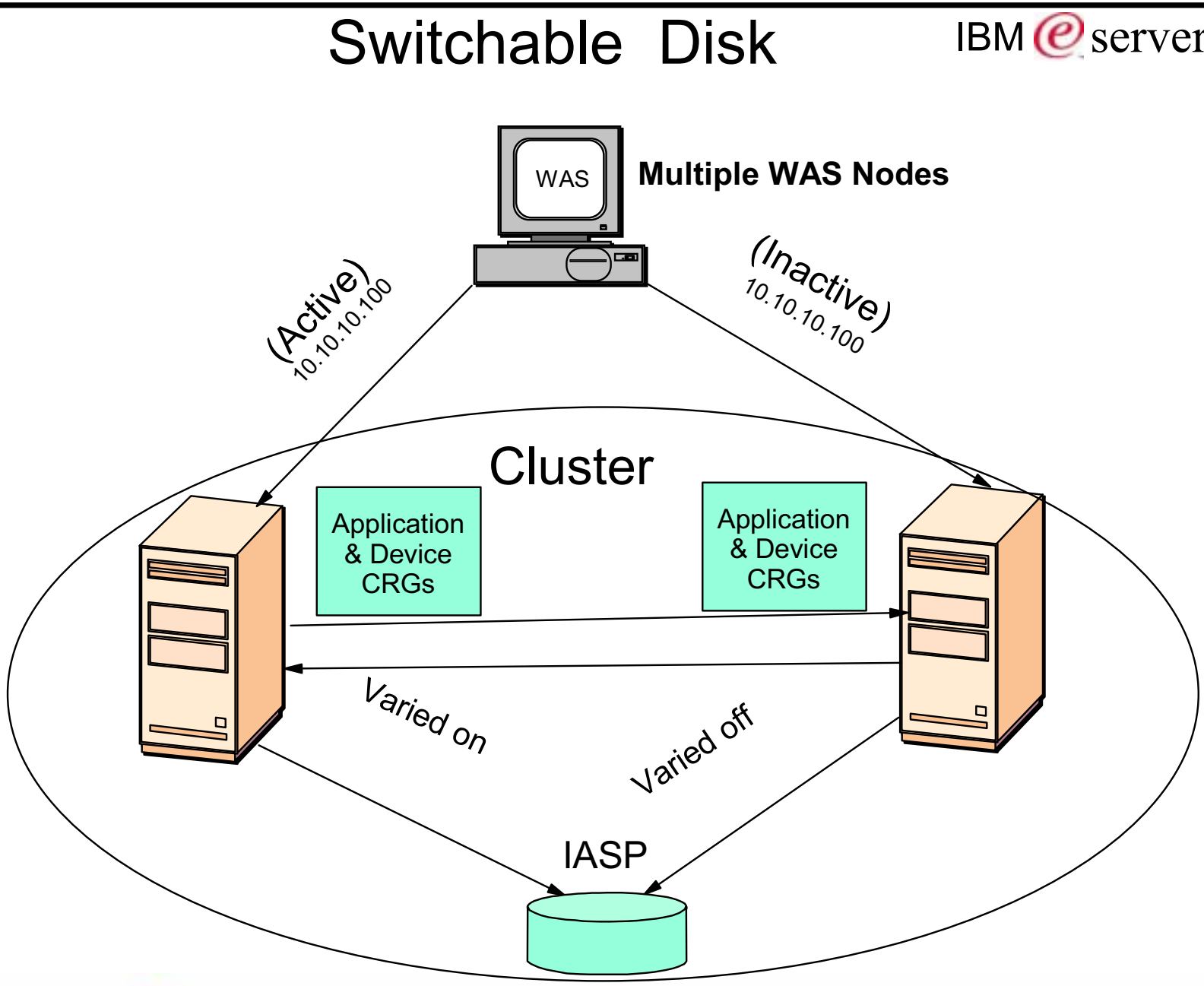
- Configure the Network Dispatcher (Continued)
  - Redbook - IBM WebSphere V4.0 Advanced Edition - SG24-6176-00 Chapter 4 - Adding Web server load balancing
- Test composite configuration via workload simulation tools

iSeries HA Switchable Disks

IBM @server

- Components
  - OS/400 V5R2 required for the nodes in the iSeries cluster
    - WebSphere systems may remain at V5R1 if the IBM Toolbox for Java JDBC driver is utilized and a V5R2 version of the jt400.jar file is in the application server classpath
  - iSeries Clustering
  - Local Journal
  - Independent ASP

Pictorial view (Switchable Disk)



iSeries Configuration - Switchable Disks...

IBM @server

- Switchable disks
  - No replica of data maintained which could get out of sync
  - All DB objects stored in Independent ASPs (IASP).
  - In case of failure, IASP can be switched between two or more nodes
  - Downside is that DB is single point of failure (recommend device parity or mirrored disk drives)
    - iSeries single system availability ~ 99.9 %
  - Primary and backup nodes need to be in close proximity
    - Maximum of 15 meter copper HSL cable
    - Maximum of 250 meter fiber optic HSL cable
    - Does not meet disaster recovery requirement
  - SMAPP (system-managed access path protection) value may need to be lowered from the default setting to assure an acceptable outage time for the database application
    - Use the 'EDTRCYAP' CL command to adjust the value

iSeries Configuration - Switchable Disks...

IBM @server

- iSeries Clustering used for:
  - Device CRG used to switch IASP
  - Wellness monitoring of the backend DB partitions

iSeries Switchable Disk Advantages

IBM @server

- Data still available despite single system outage
- Disk failures isolated within independent disk pools
- OS/400 remote journaling and replication software not required
  - Use local journaling with commitment control instead
- Reclaim storage can be done just for an individual IASP
- Self-contained applications can be moved to IASPs for server consolidation
- Workload switching across OS/400 servers
- Infrequently-used data can remain offline in an IASP to reduce overall system IPL time and the duration of system save operations
- Mirroring or device parity protection for an IASP may be activated through iSeries Navigator instead of using traditional DST functions

iSeries Switchable Disk Considerations

IBM @server

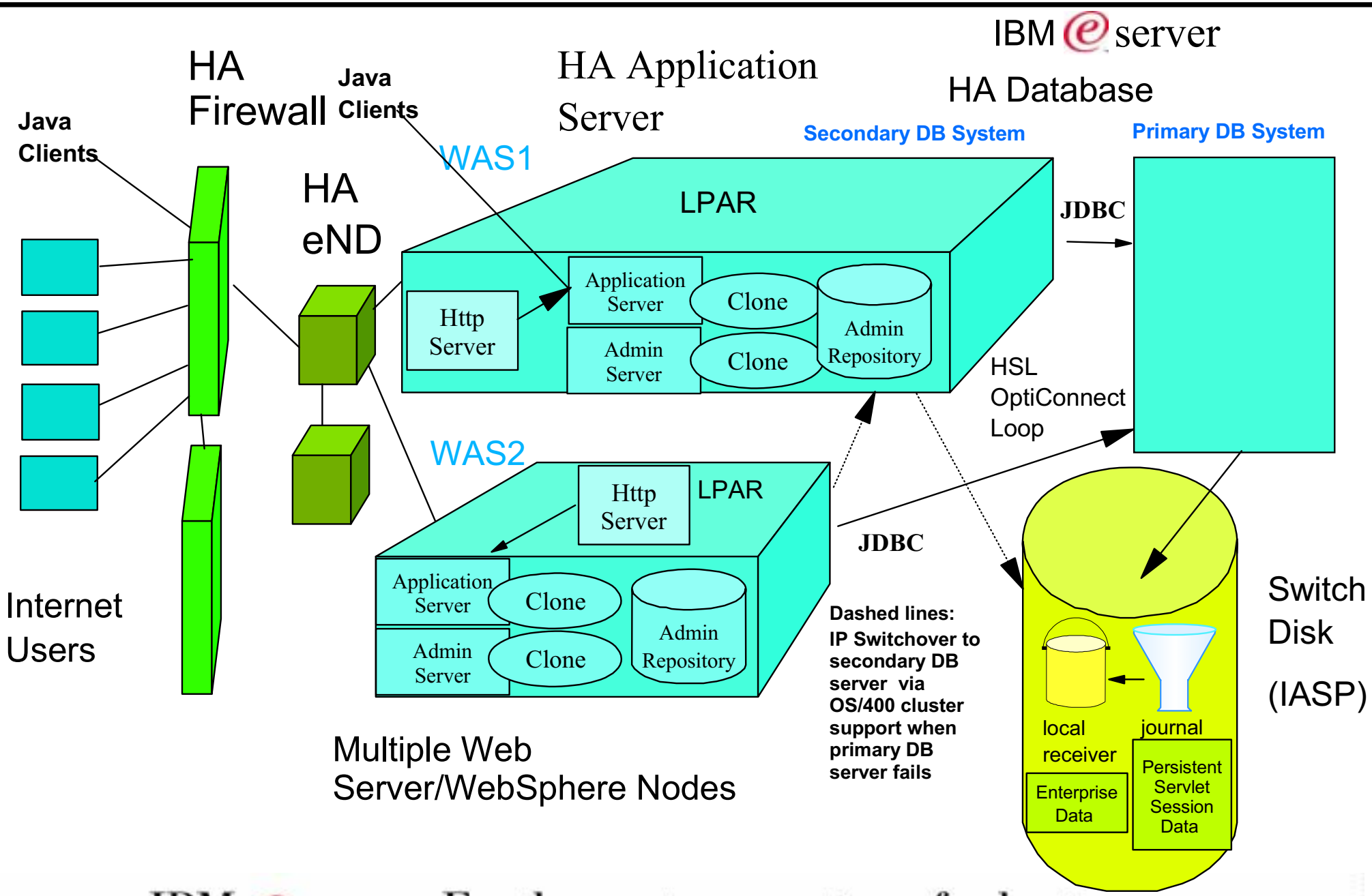
- The IASP itself is a single point of failure when disk failures occur
  - Make sure to configure mirroring or device parity for critical IASPs
- Does not meet disaster recovery requirements
  - Systems attached to IASP must be within 15-250 meters of each other due to loop limitations associated with HSL
  - Sustained power outages, floods, fires, and other disasters will disrupt access to the IASP
- No workload balancing across multiple systems since only one system can access the IASP at any one time
- Systems with IASPs should have a limited number of database files stored in *SYSBAS so that vary-on processing is relatively quick and transparent
- Need awareness of user profile, security, supported object types, and object identification considerations related to IASPs

Topology II (Switchable Disks) - Commentary

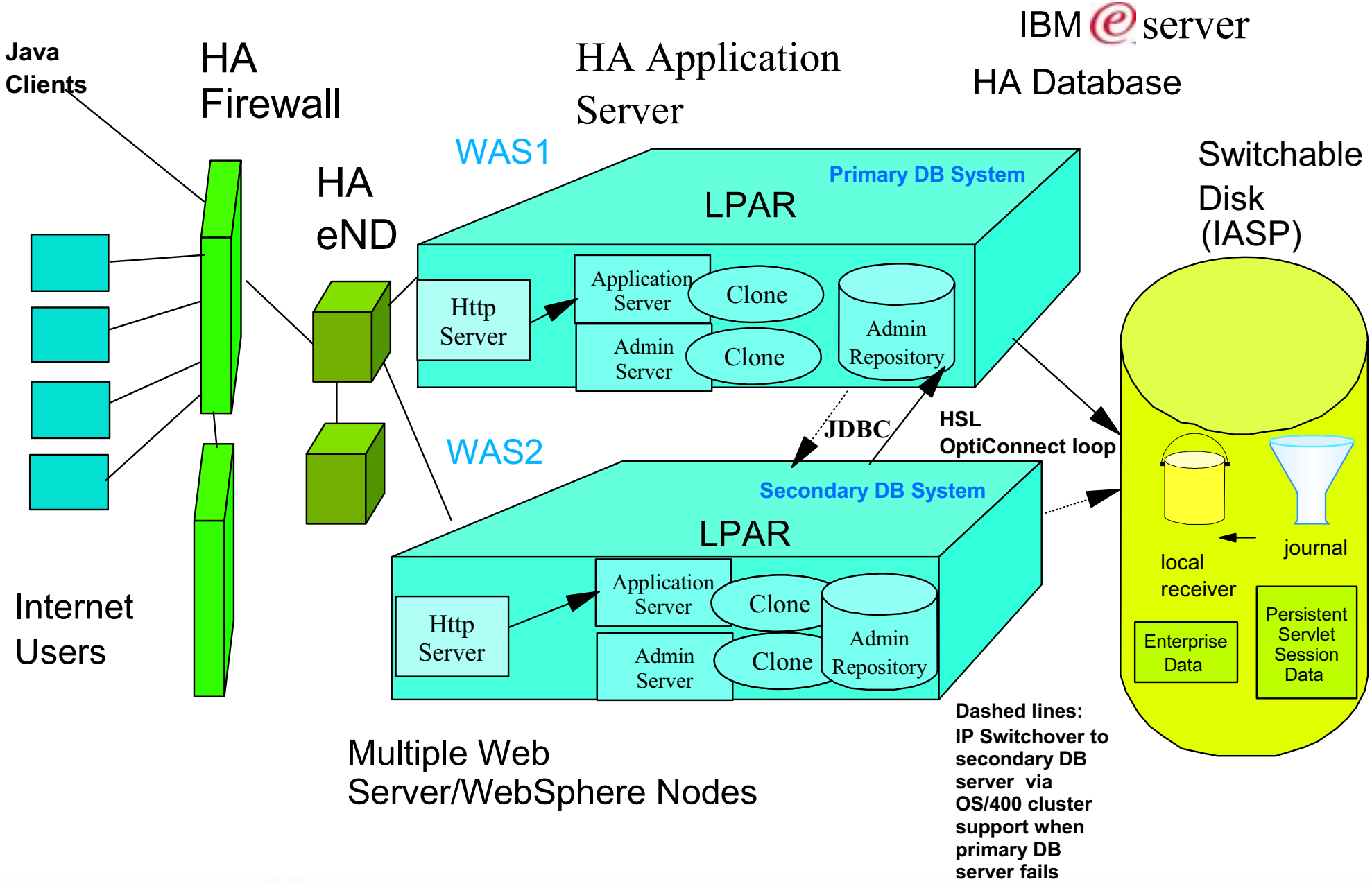
IBM @server

- HA Switchable Resources (OS/400 option 41) must be installed on all the nodes in the iSeries cluster
- WAS1 and WAS 2 simultaneously run identical applications in two administrative domains.
  - Provides not only failover but also planned switchover for maintenance.
- WAS1 and WAS2 use vertical cloning for both application servers and admin server.
- Session state persistent to backend DB. Data source directed via IP address for clustering IP takeover
  - Failure of WAS1 or WAS2 will be almost transparent except for active clients at time of failure
- Local journal used to preserve database transaction boundaries
- Both the cluster and switchable hardware need to be started

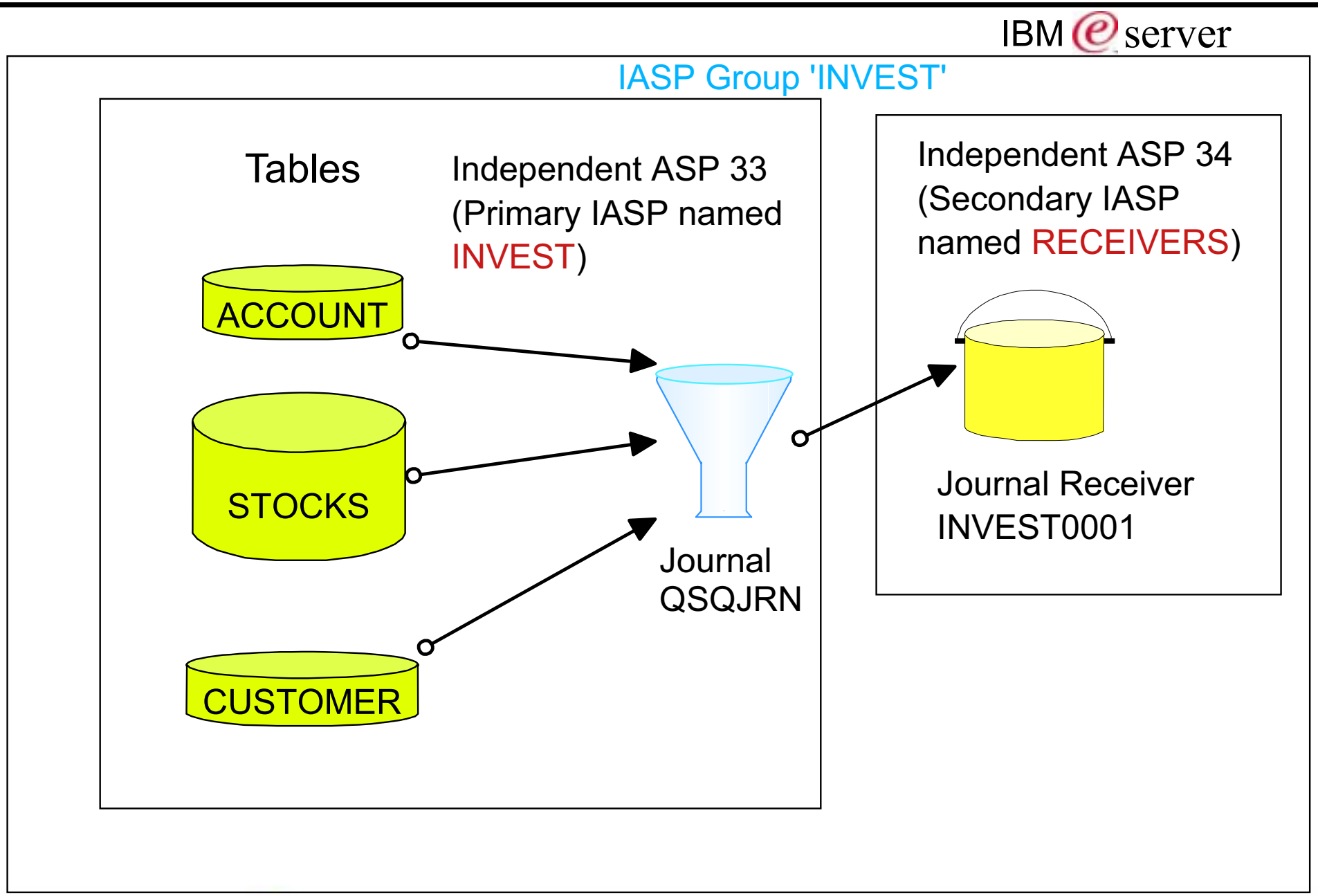
Topology II - Switchable Disks - Remote DB system



Topology III - Switchable Disks - Partition DB



Example Disk Pool Group 'Invest'

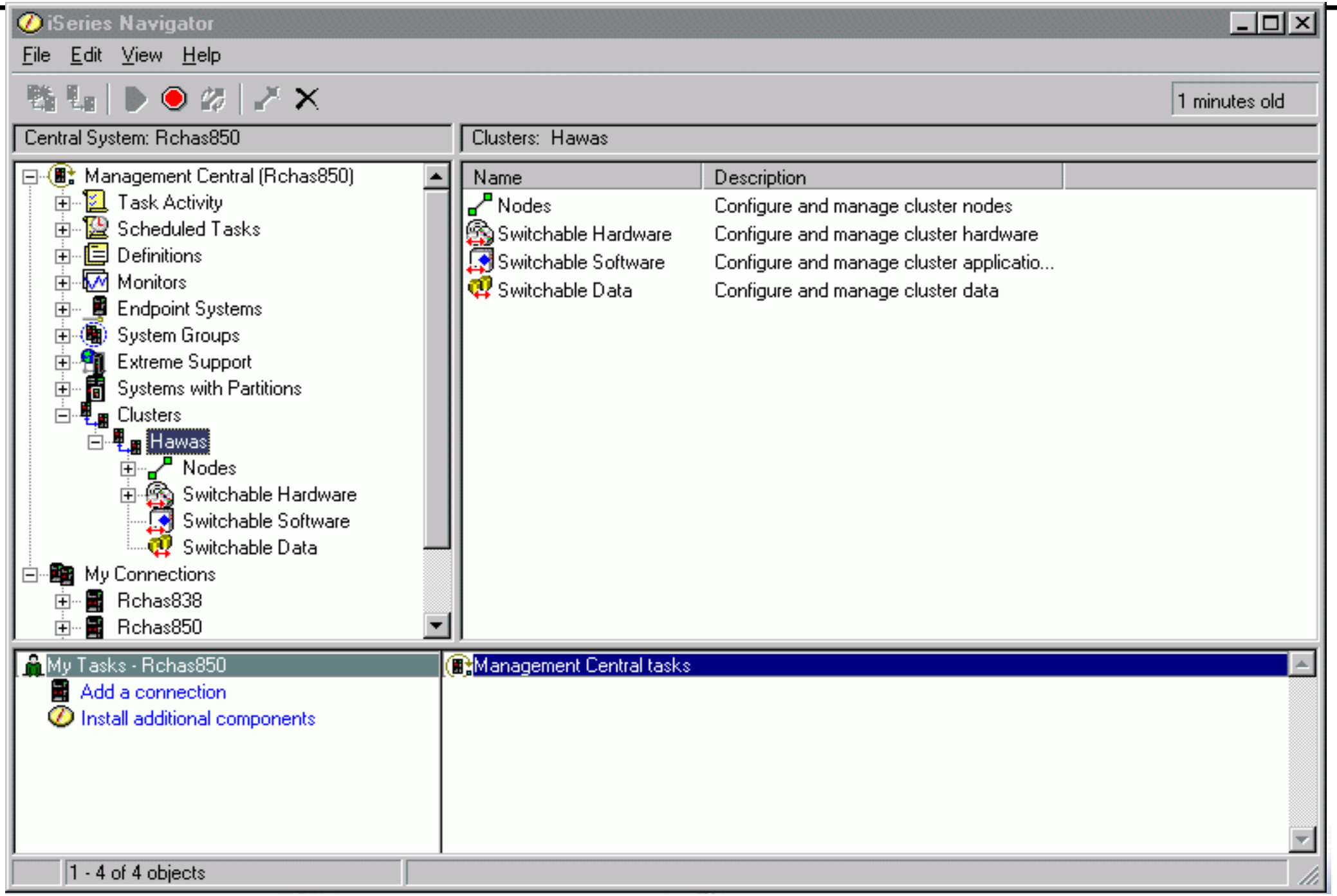


How did we move our database collections to a disk pool group which contains primary and secondary IASPs?

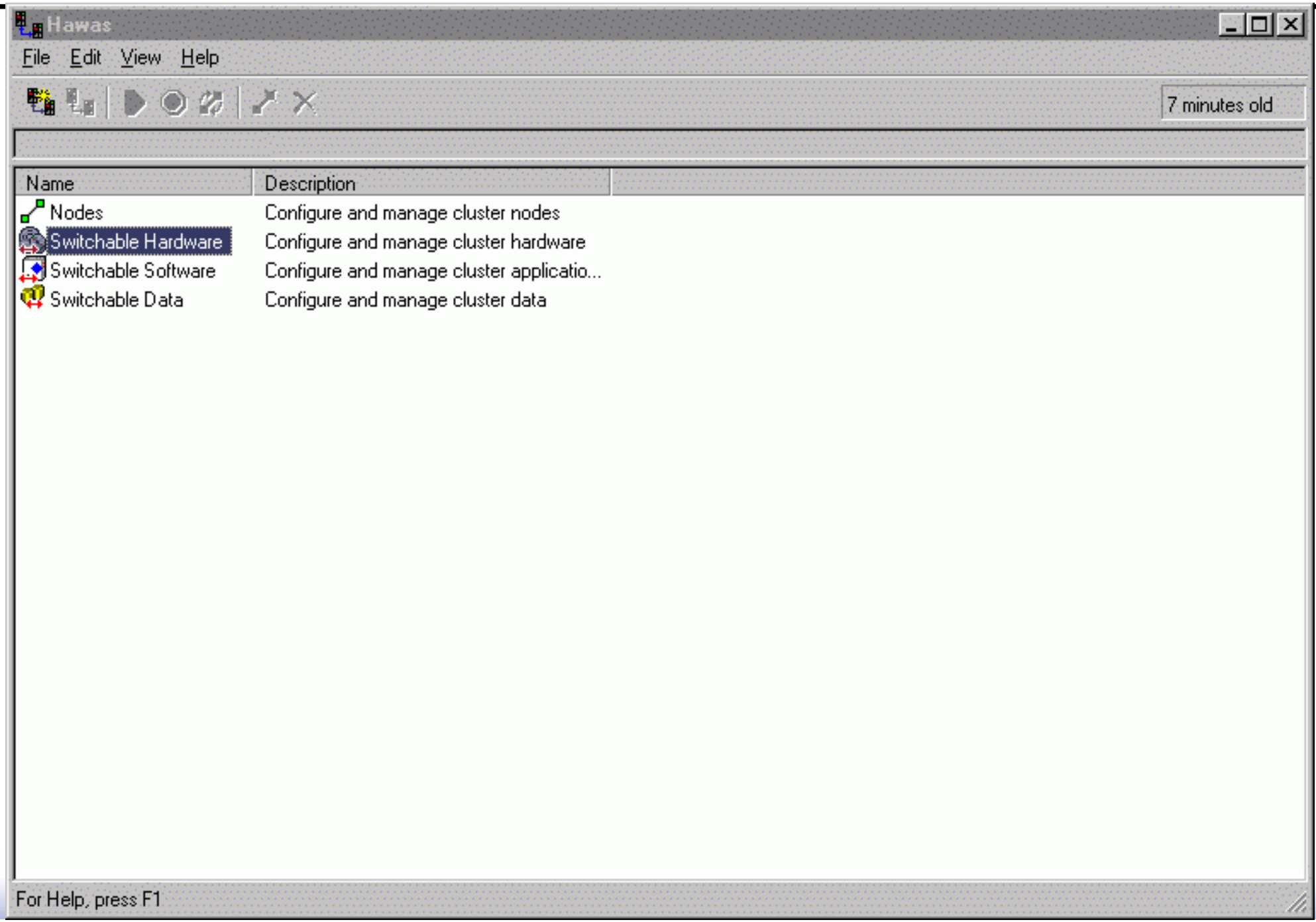
IBM @server

- Performed SAVLIB command to a save file for our existing INVEST SQL collection located on a production iSeries server
- Transferred the save file to the primary node in the iSeries cluster and issued the following CL commands:
- RSTLIB SAVLIB(INVEST) DEV(*SAVF) SAVF(DARIN/INVEST) RSTASPDEV(INVEST)
- CRTLIB LIB(HAWAS) ASP(*ASPDEV) ASPDEV(INVEST)
- CRTLIB LIB(INVESTJRN) ASP(*ASPDEV) ASPDEV(RECEIVERS)
- SETASPGRP INVEST
- CRTJRNRCV JRNRCV(INVESTJRN/INVEST0001) THRESHOLD(100000)
- CHGJRN JRN(INVEST/QSQJRN) JRNRCV(INVESTJRN/INVEST0001) MNGRCV(*SAME) RCVSIZOPT(*RMVINTENT) MINENTDTA(*FILE)

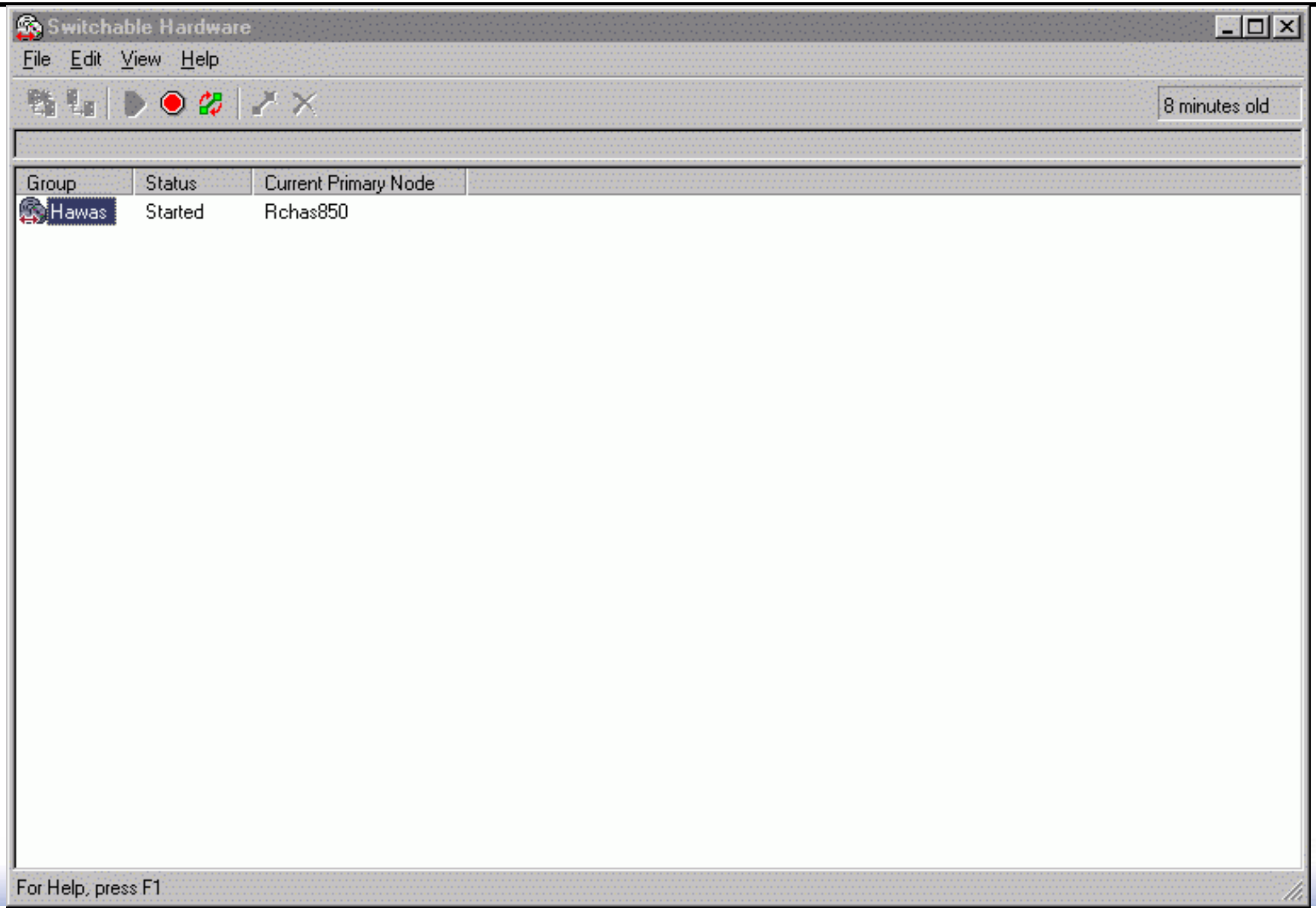
Switchable IASP: iSeries Navigator Screenshots



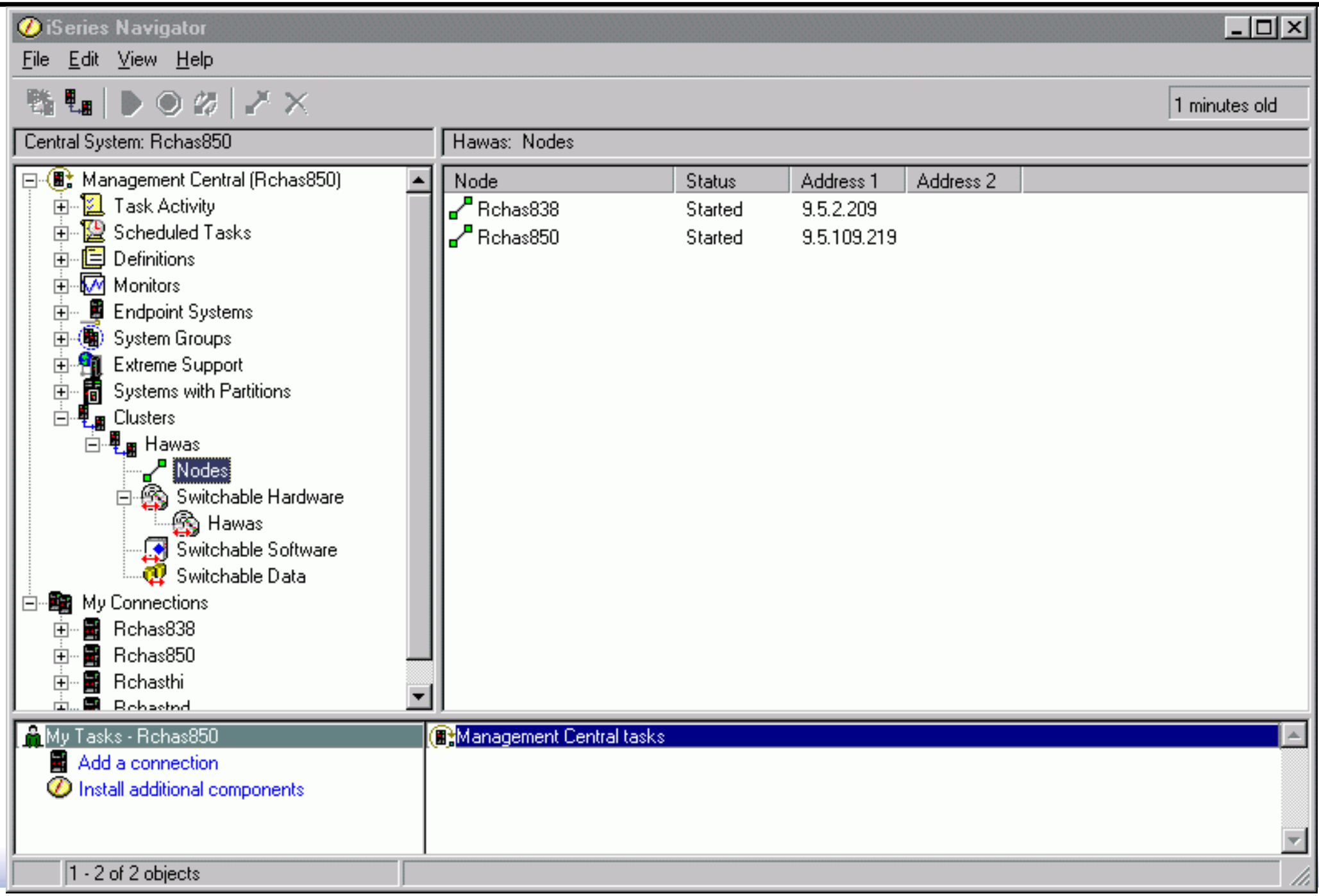
Switchable IASP: iSeries Navigator Screenshots



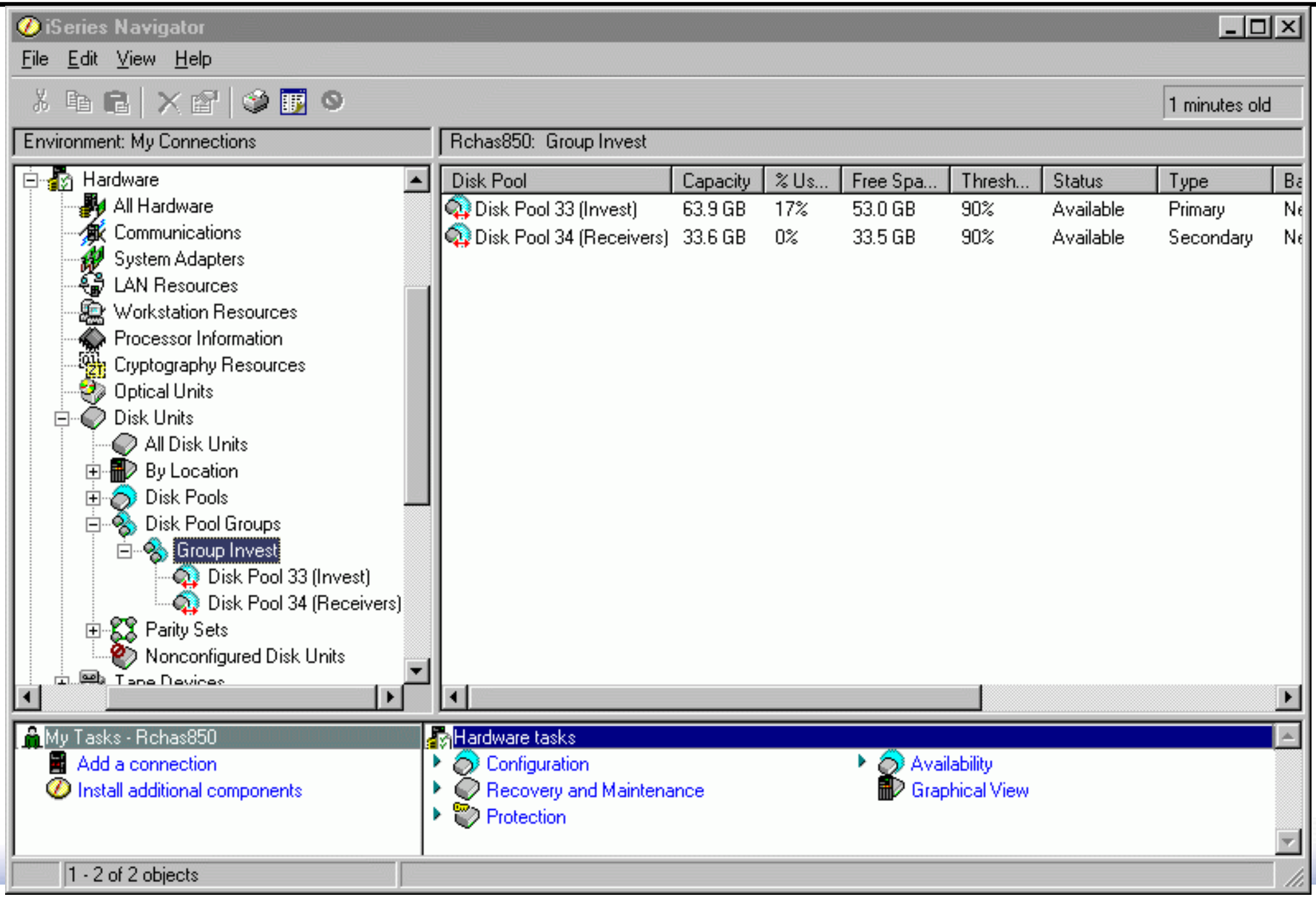
Switchable IASP: iSeries Navigator Screenshots



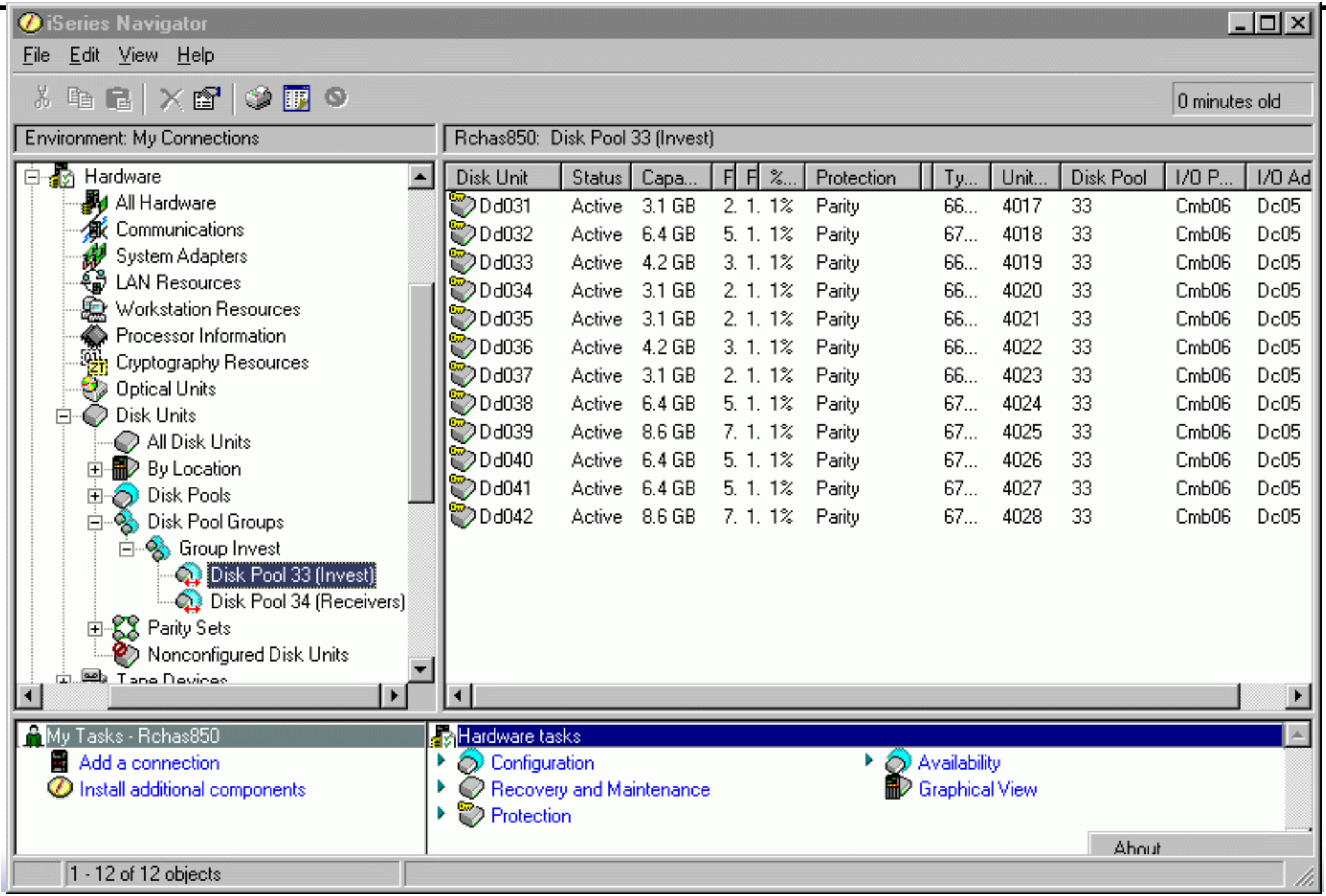
Switchable IASP: iSeries Navigator Screenshots



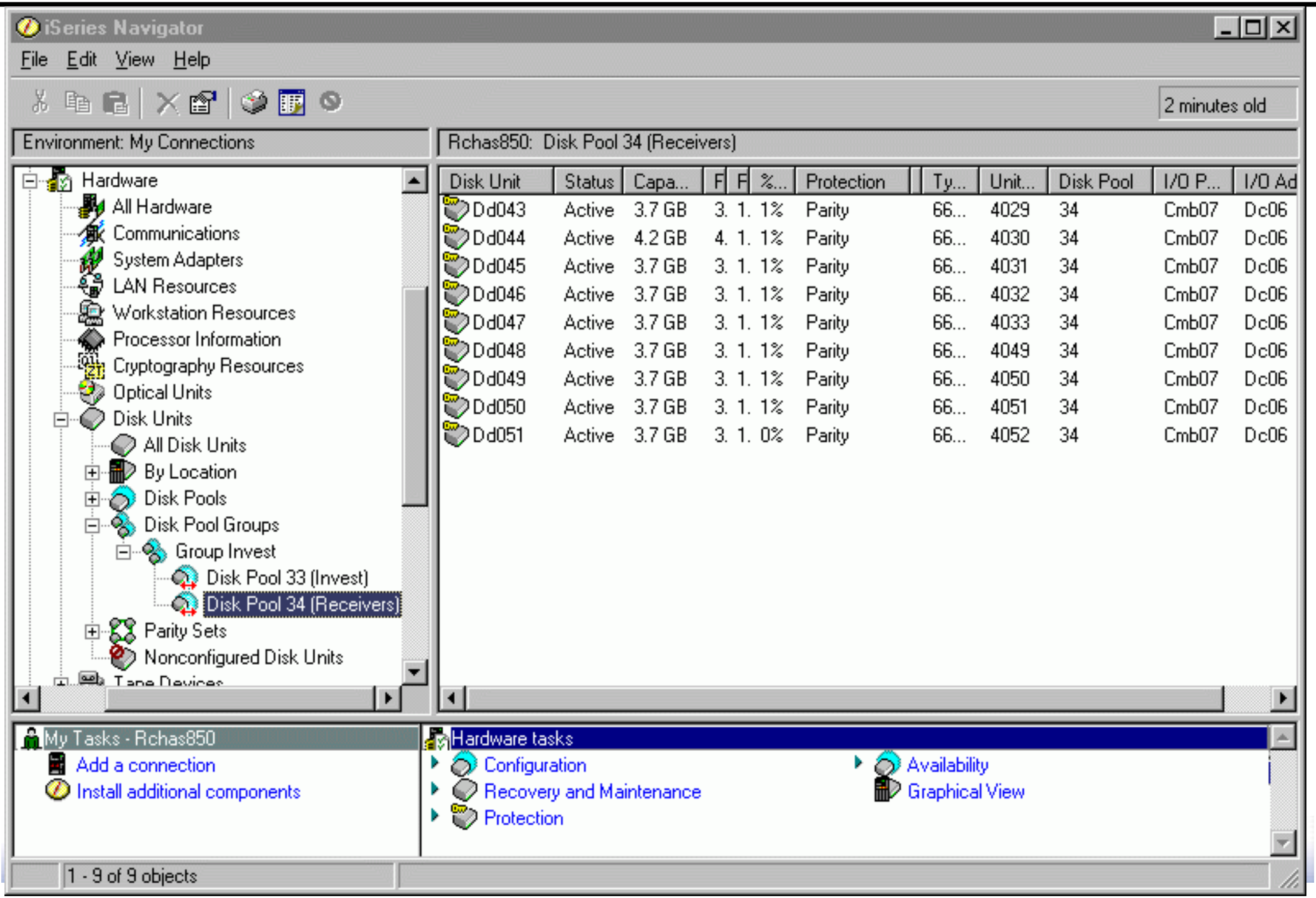
Switchable IASP: iSeries Navigator Screenshots



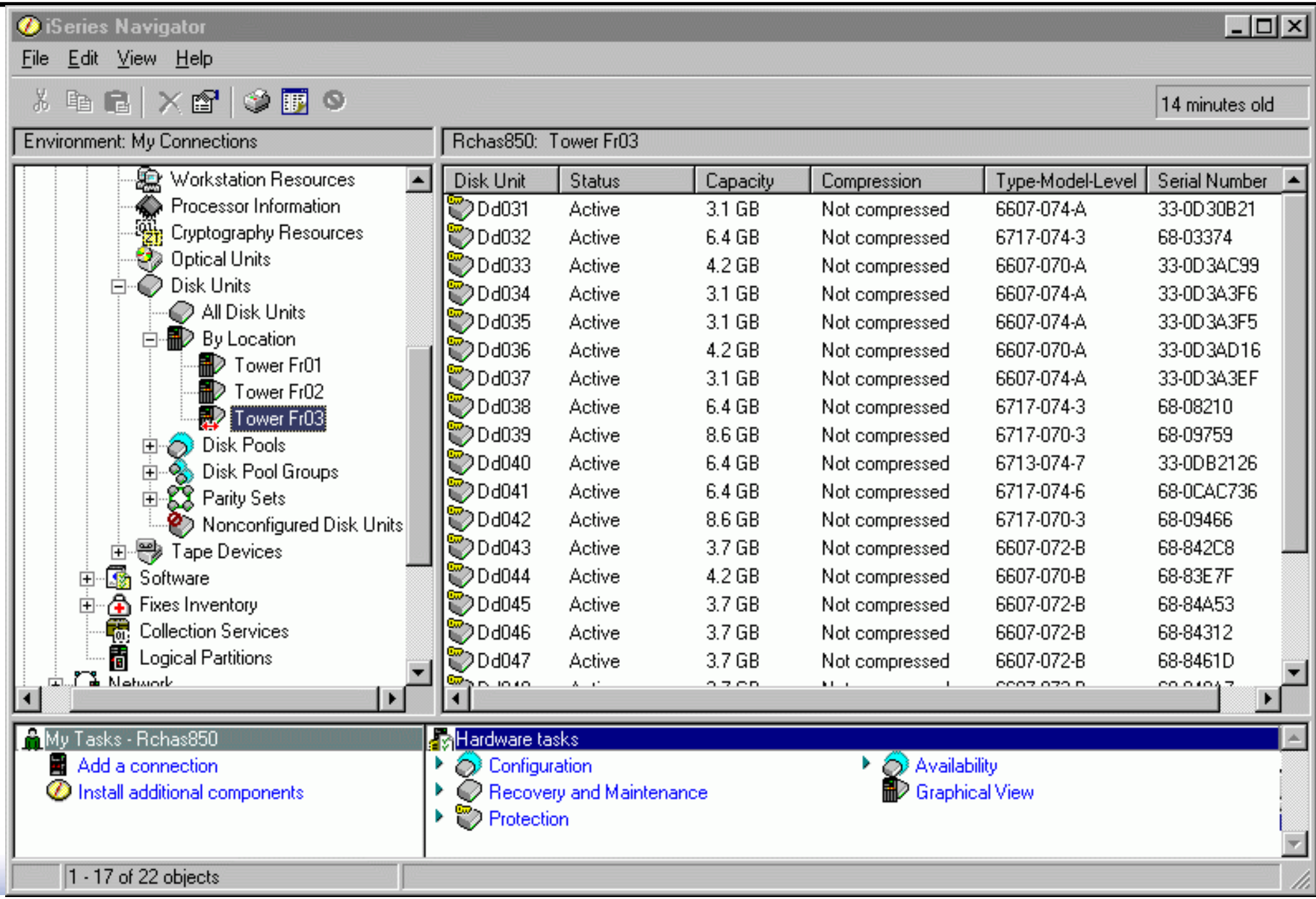
Switchable IASP: iSeries Navigator Screenshots



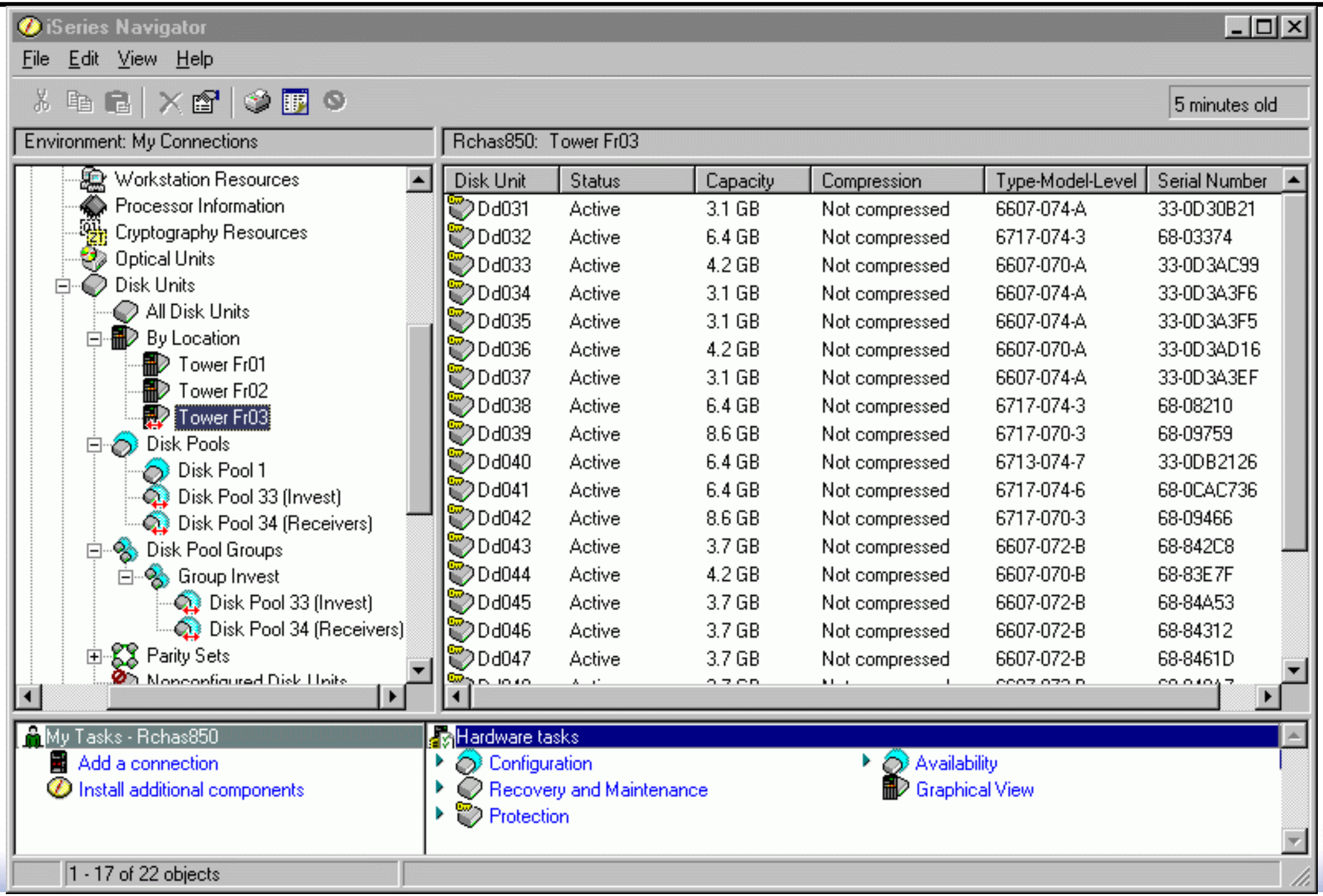
Switchable IASP: iSeries Navigator Screenshots



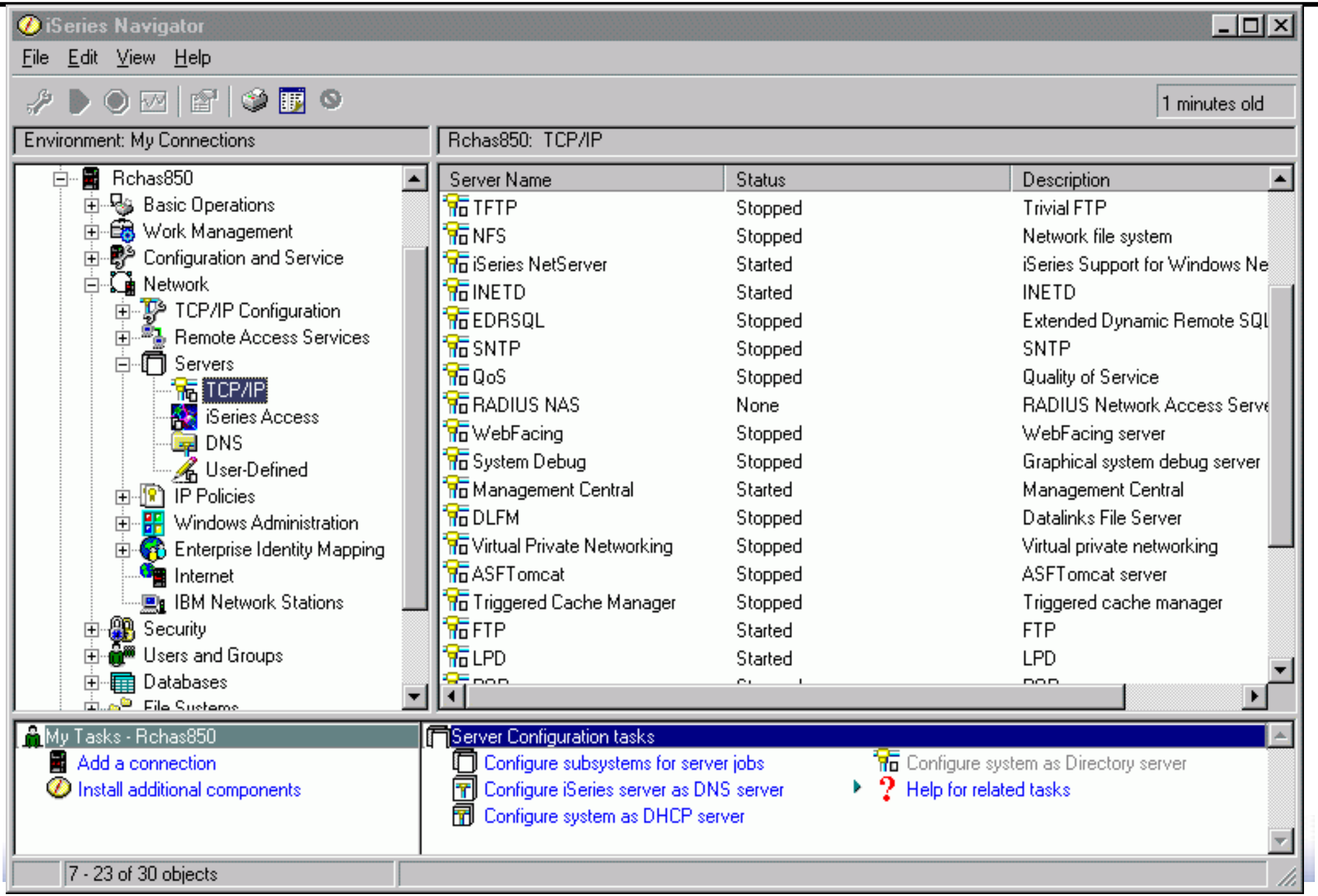
Switchable IASP: iSeries Navigator Screenshots



Switchable IASP: iSeries Navigator Screenshots

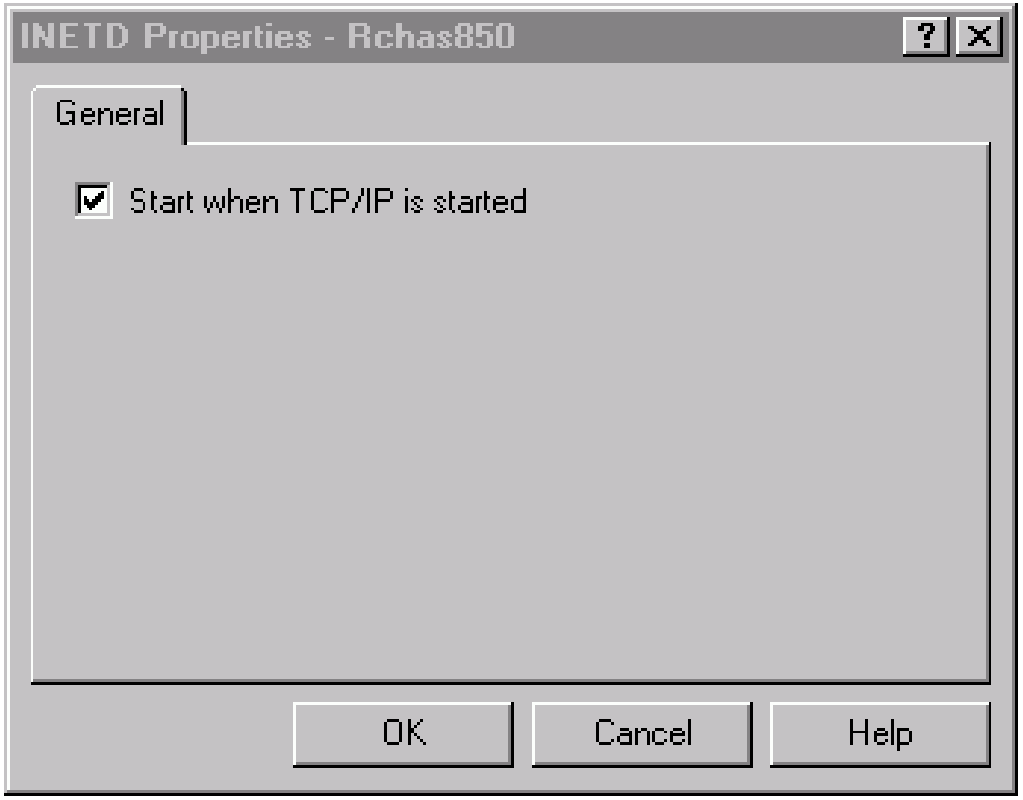


Switchable IASP: iSeries Navigator Screenshots

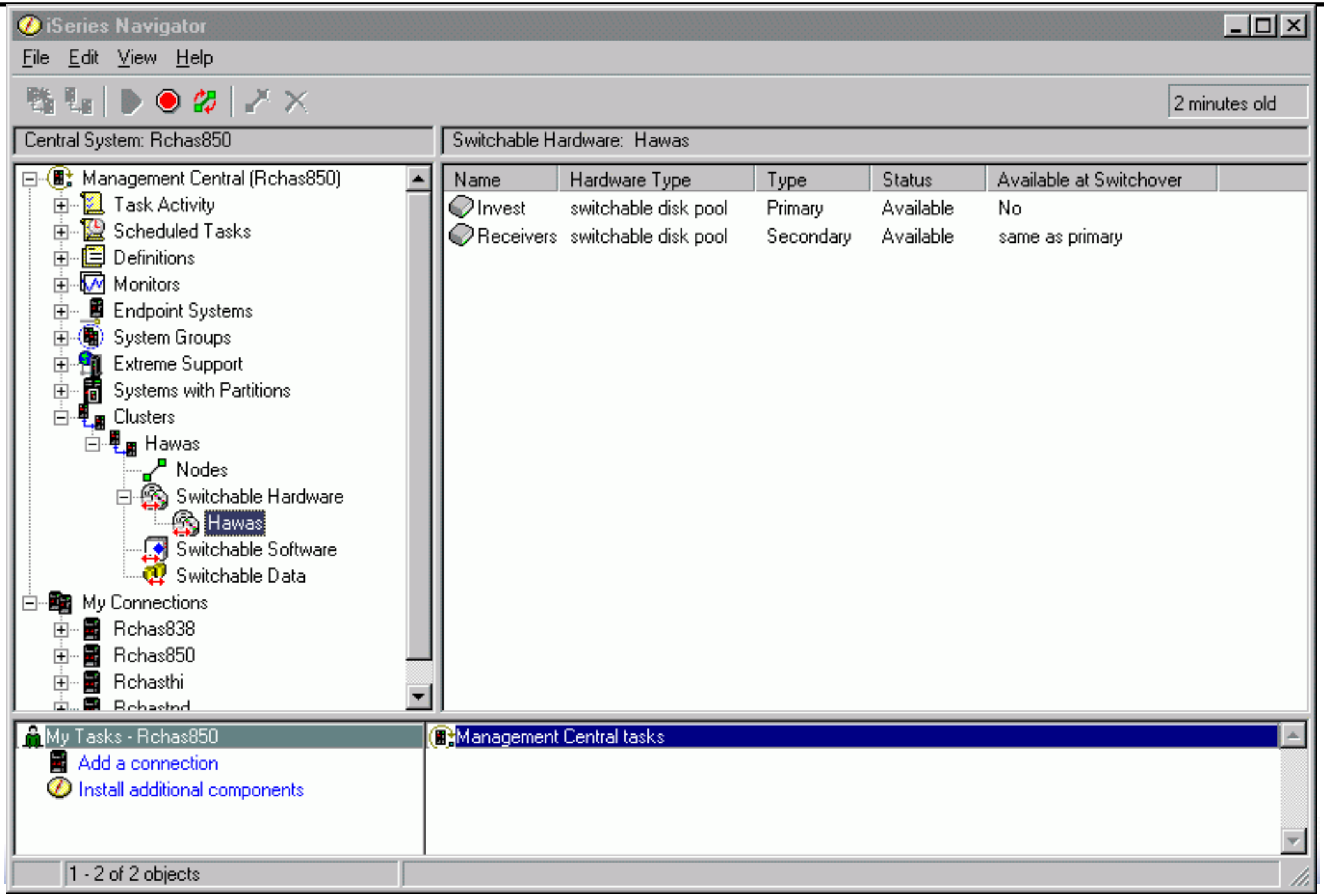


Switchable IASP: iSeries Navigator Screenshots

IBM @server

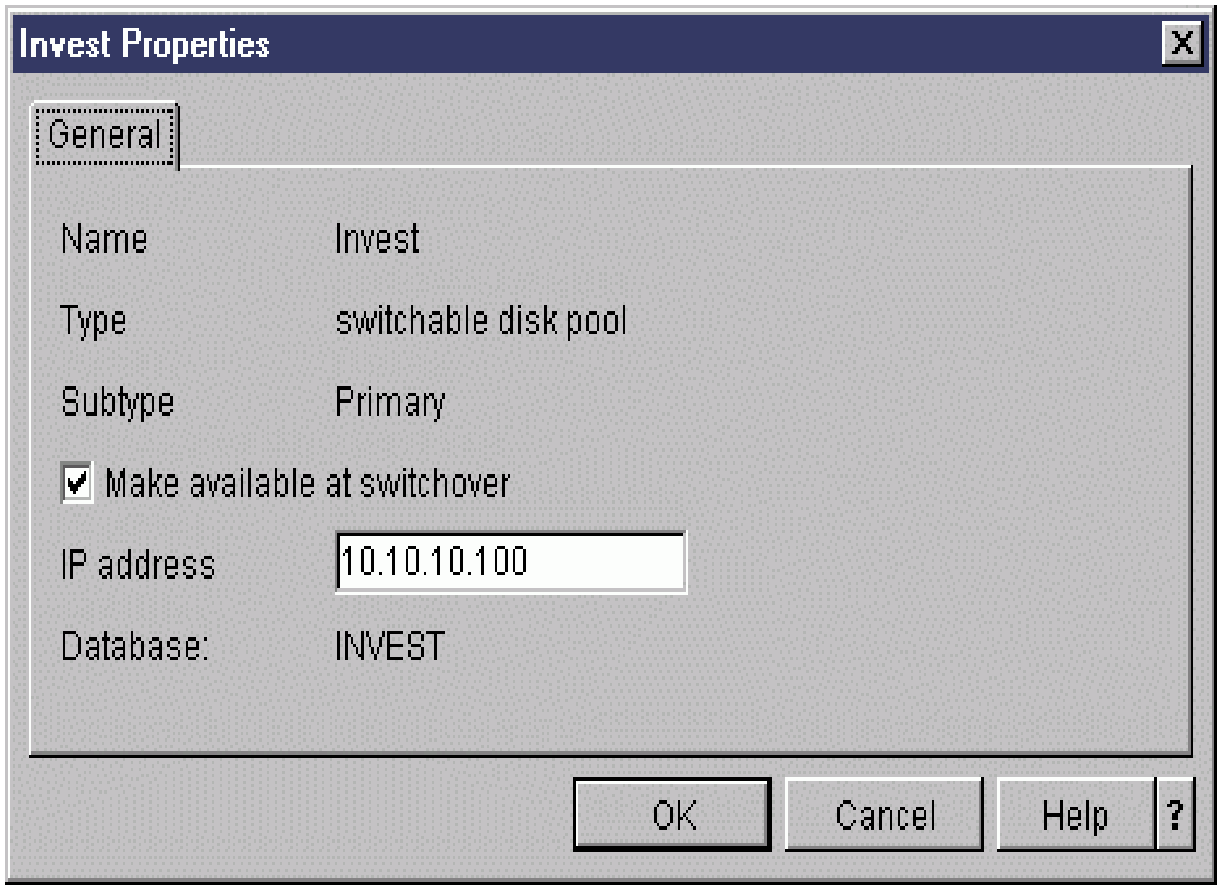


Switchable IASP: iSeries Navigator Screenshots

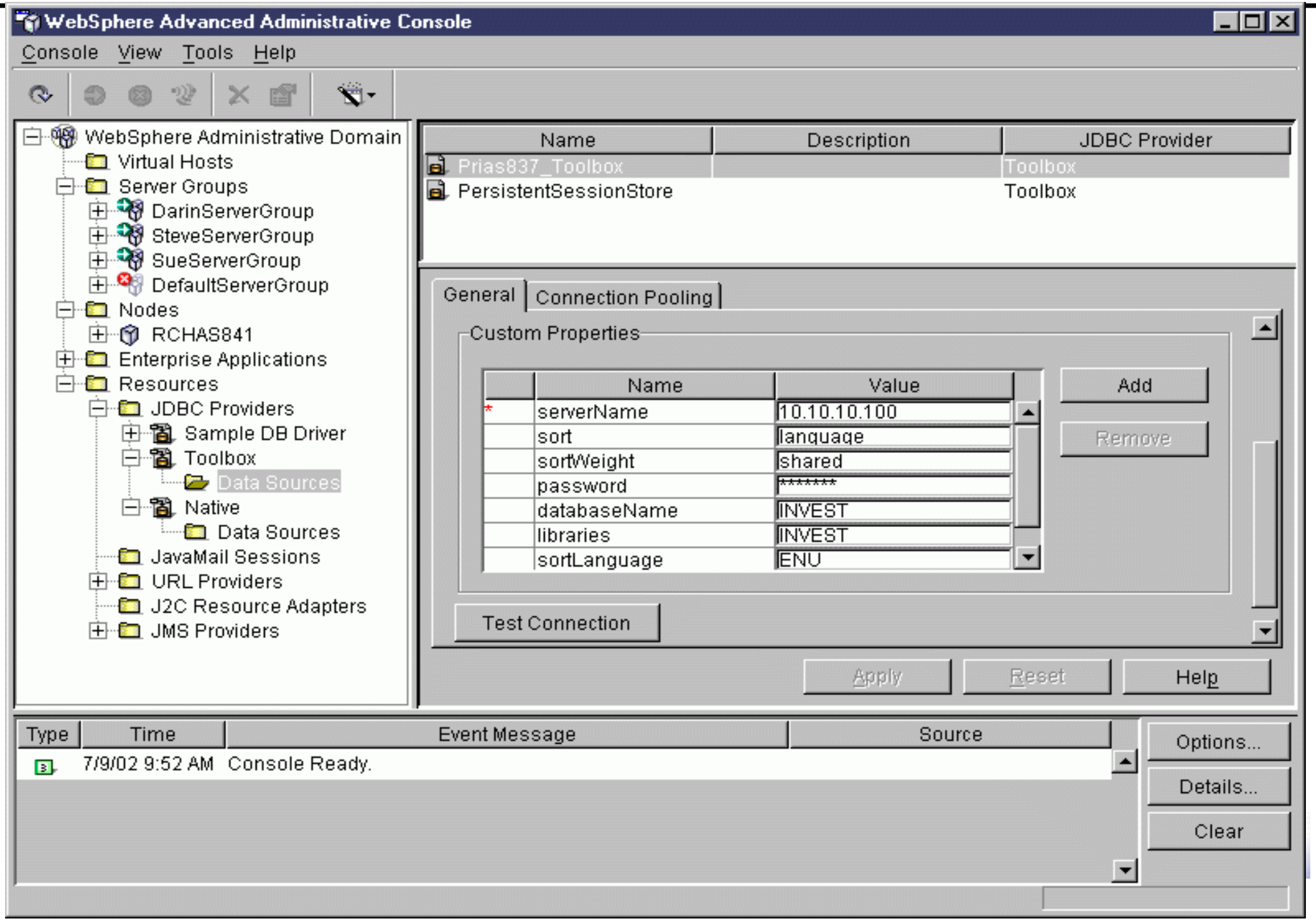


Switchable IASP: iSeries Navigator Screenshots

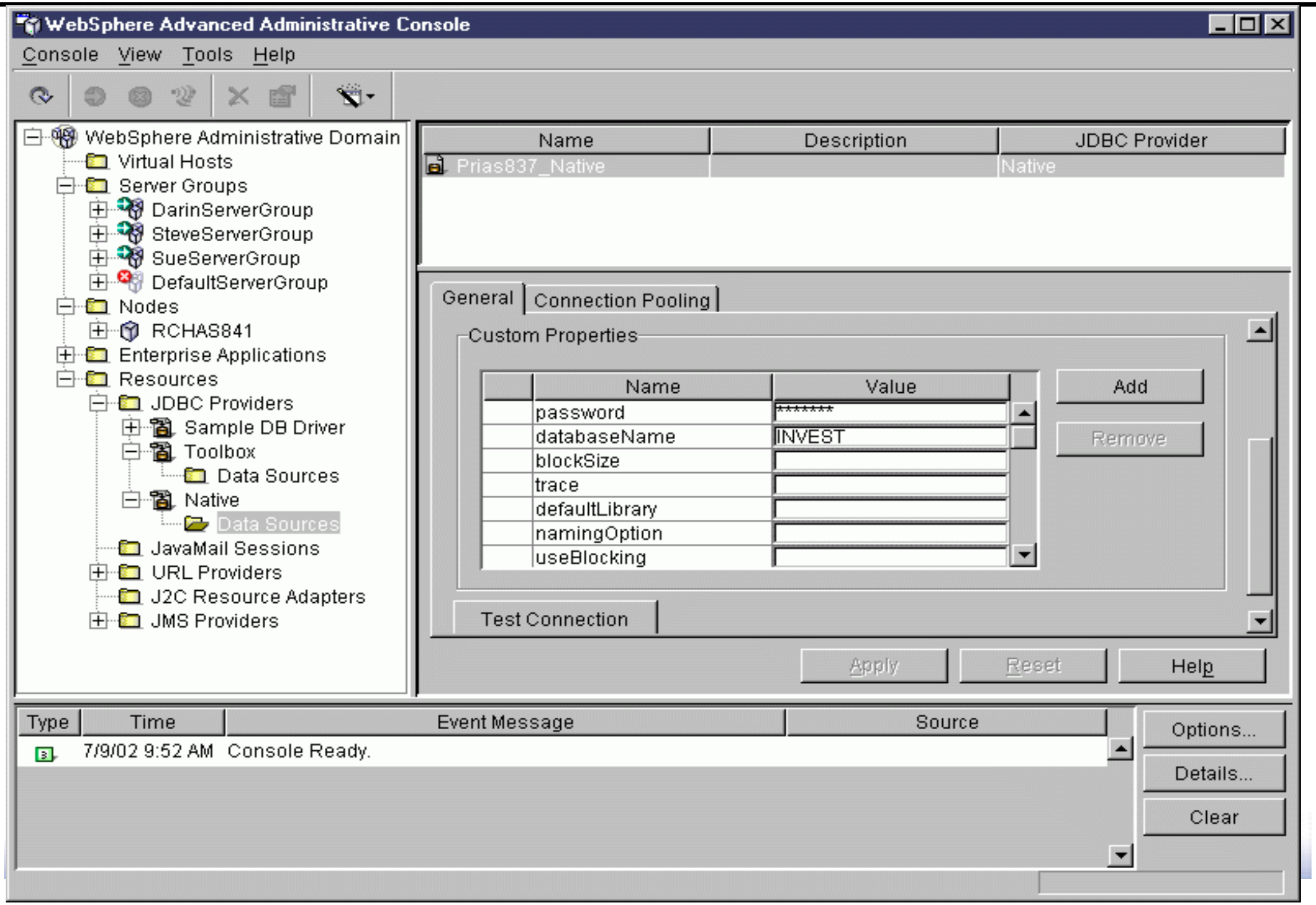
IBM @server



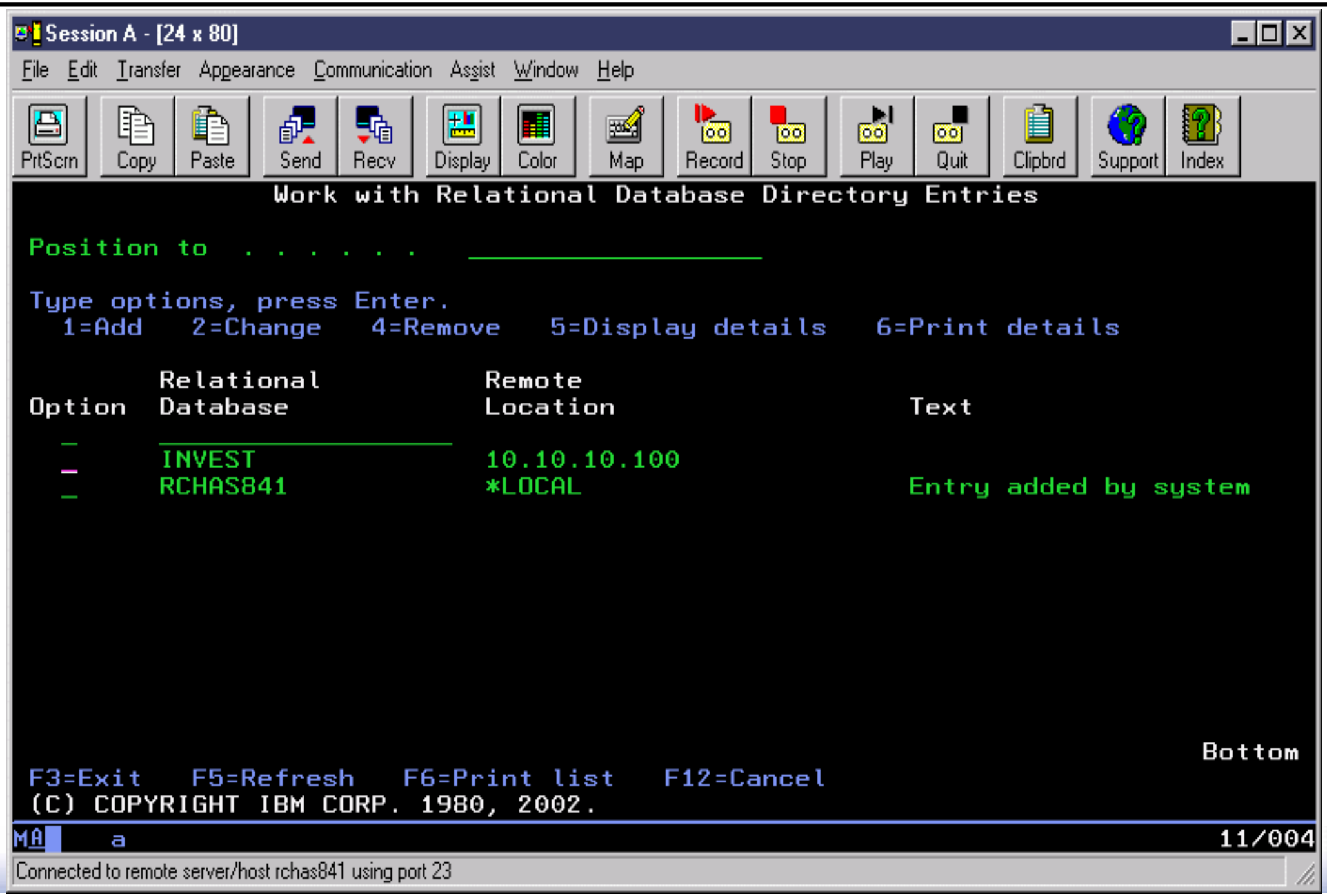
Switchable IASP: WebSphere 4.0 Screenshots



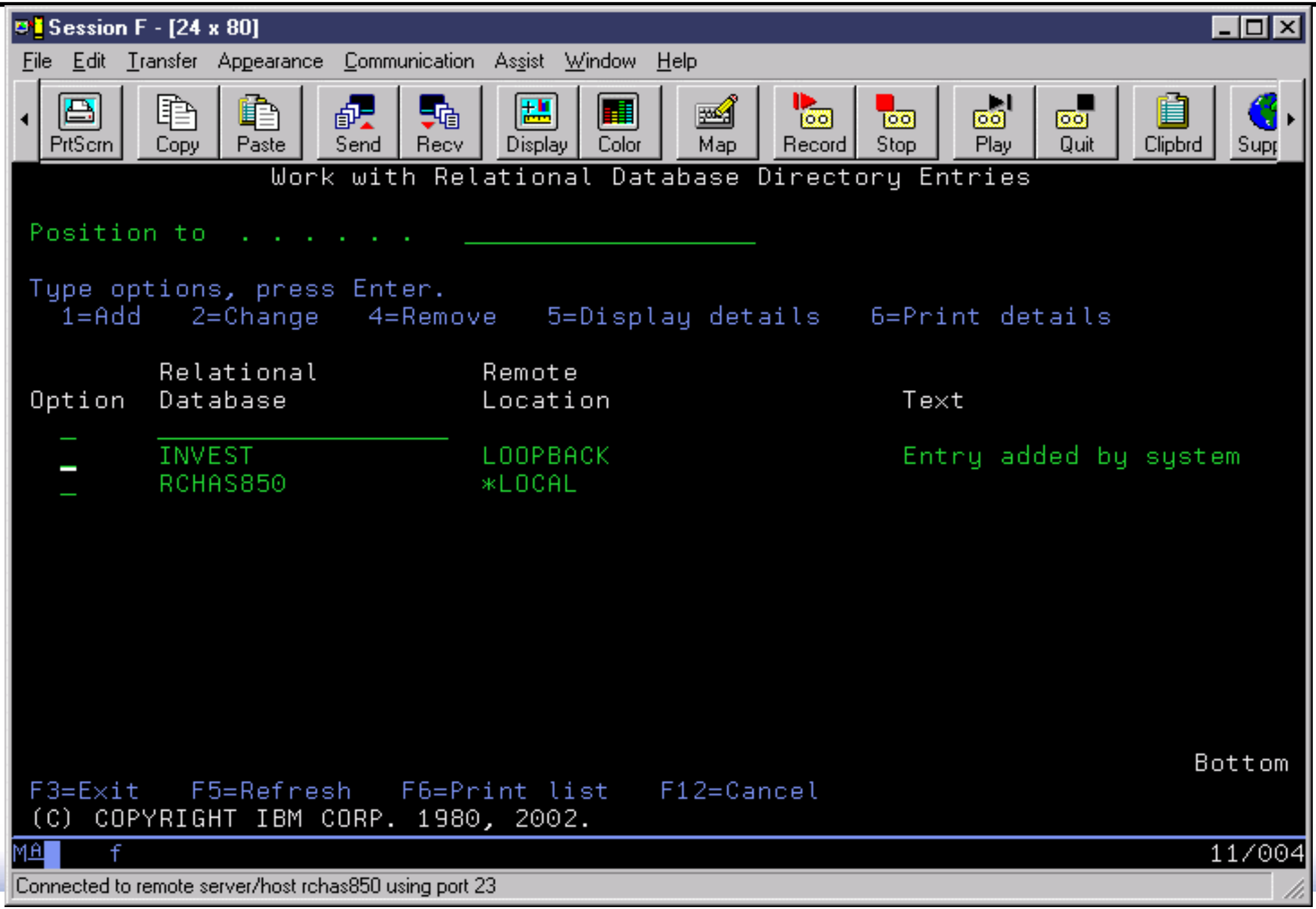
Switchable IASP: WebSphere 4.0 Screenshots



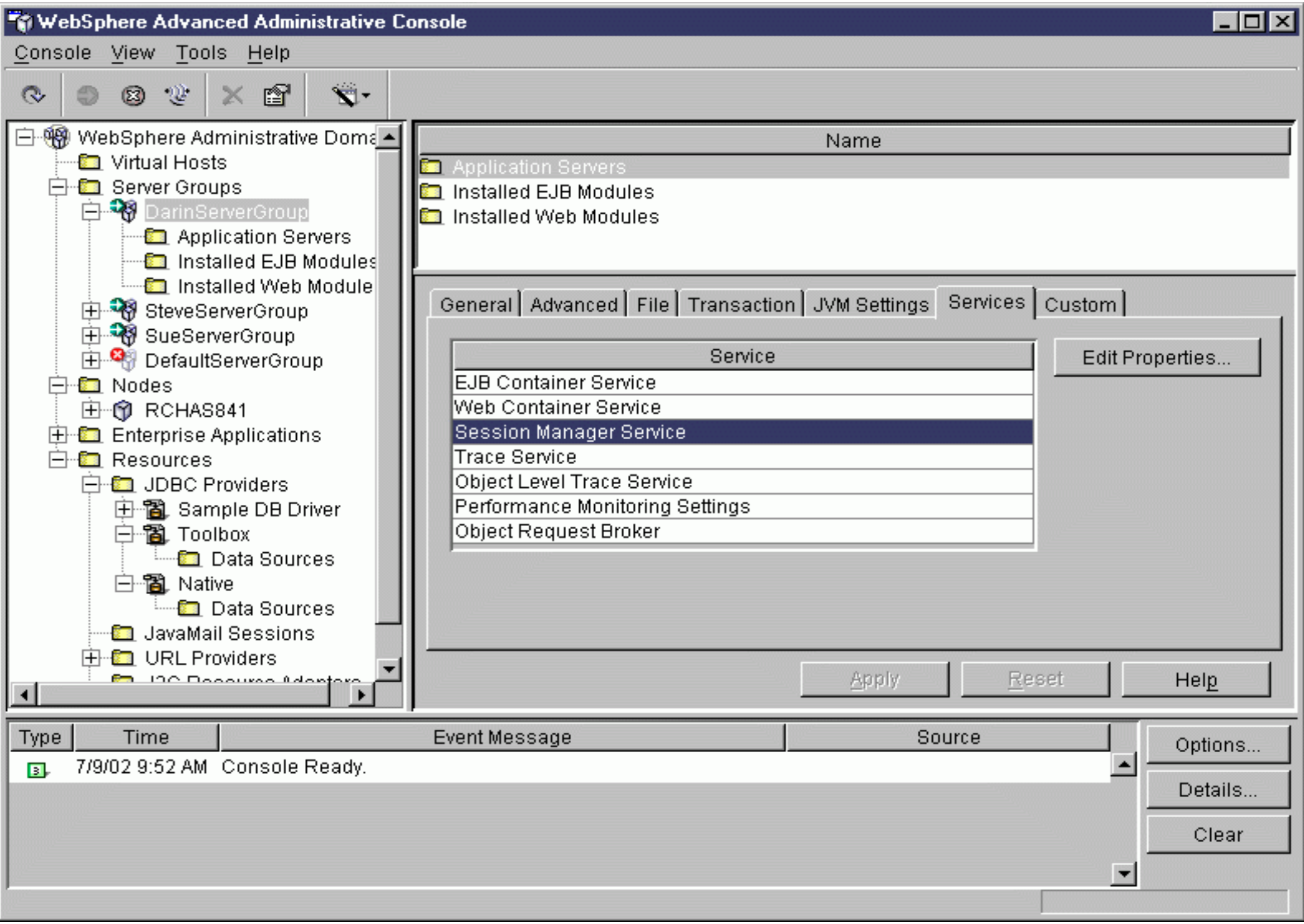
Switchable IASP: 'wrkrdbdire' command screenshots



Switchable IASP: 'wrkrdbdire' command screenshots

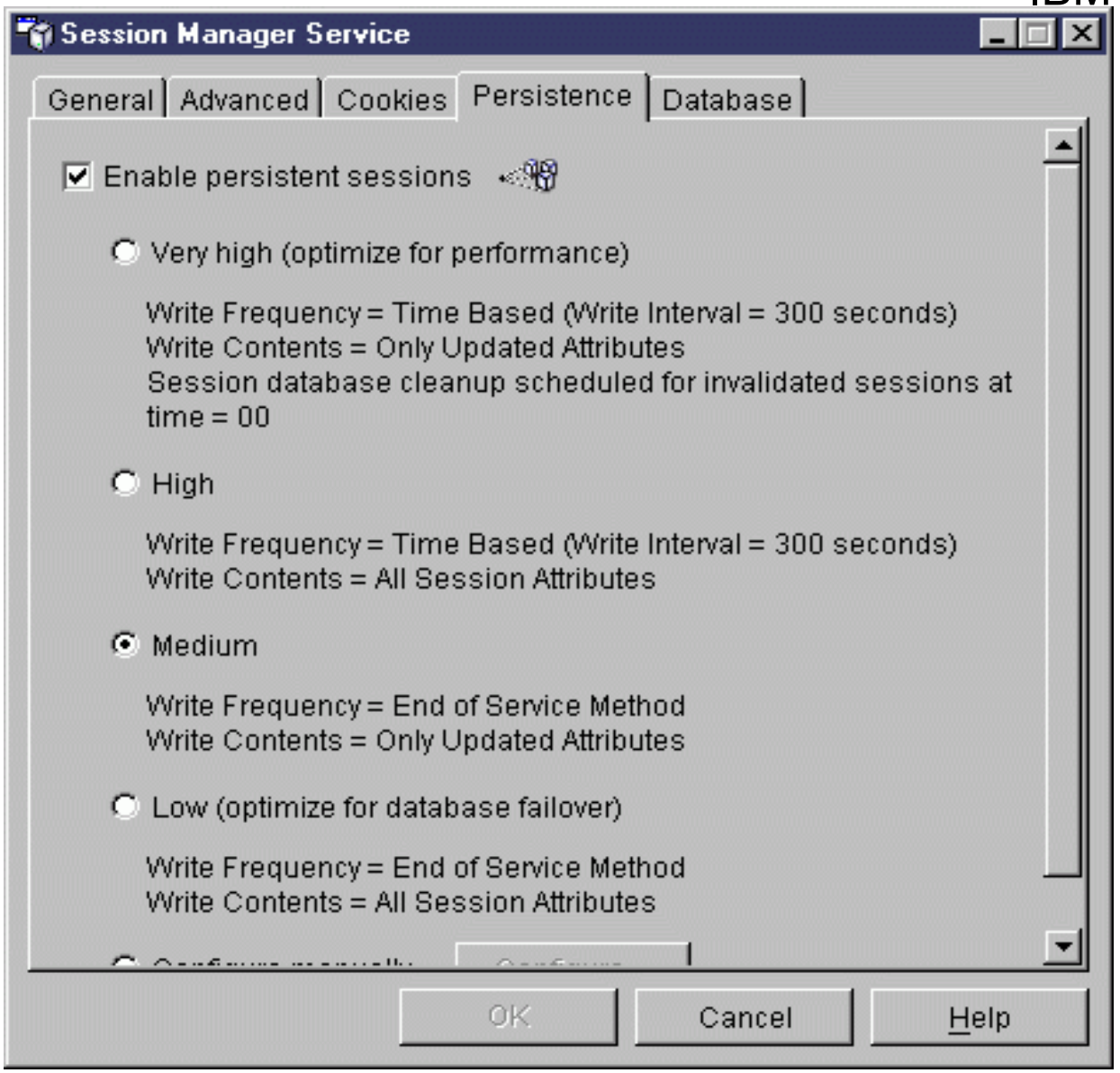


Switchable IASP: WebSphere 4.0 Screenshots



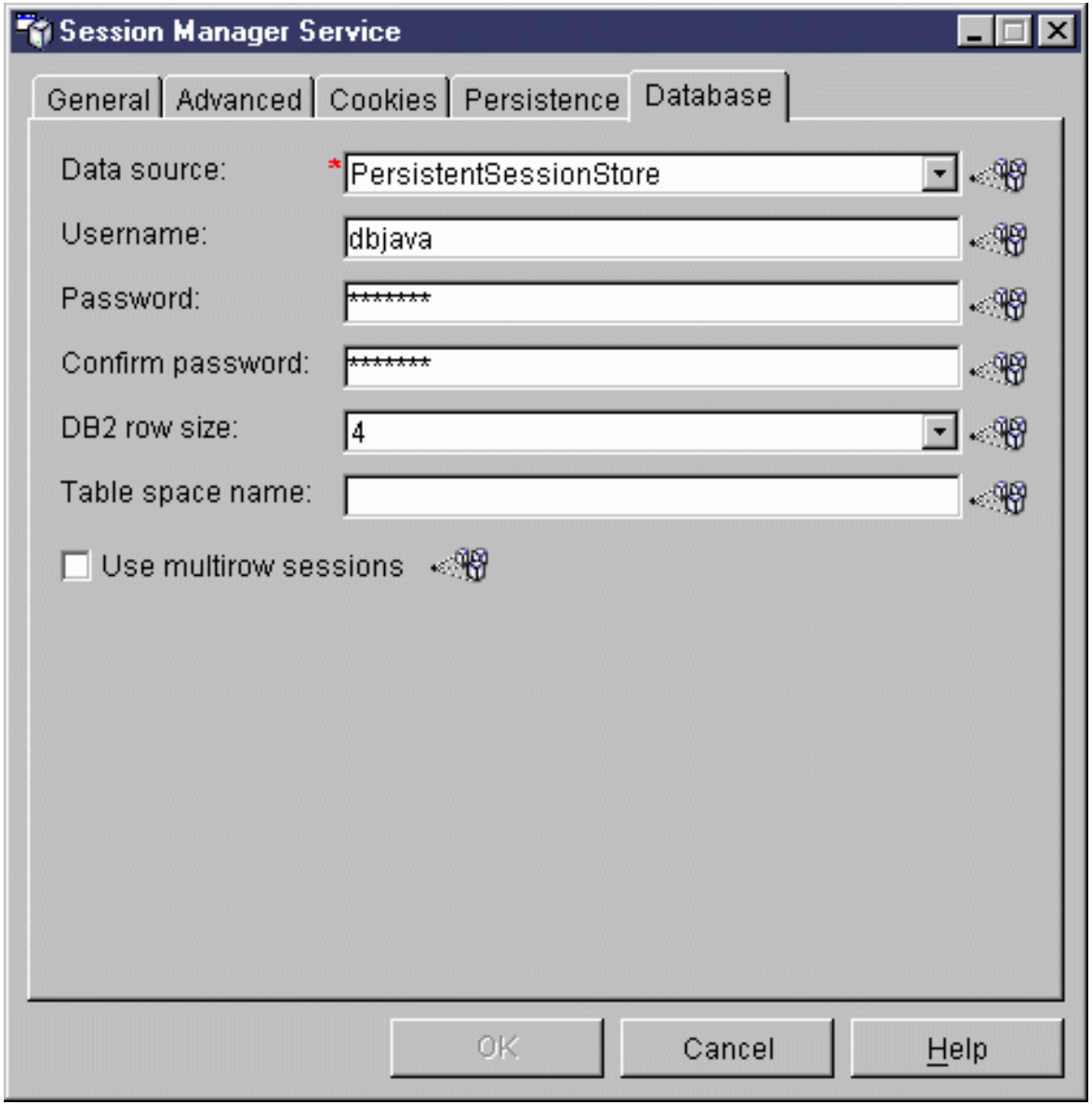
Switchable IASP: WebSphere 4.0 Screenshots

IBM @server

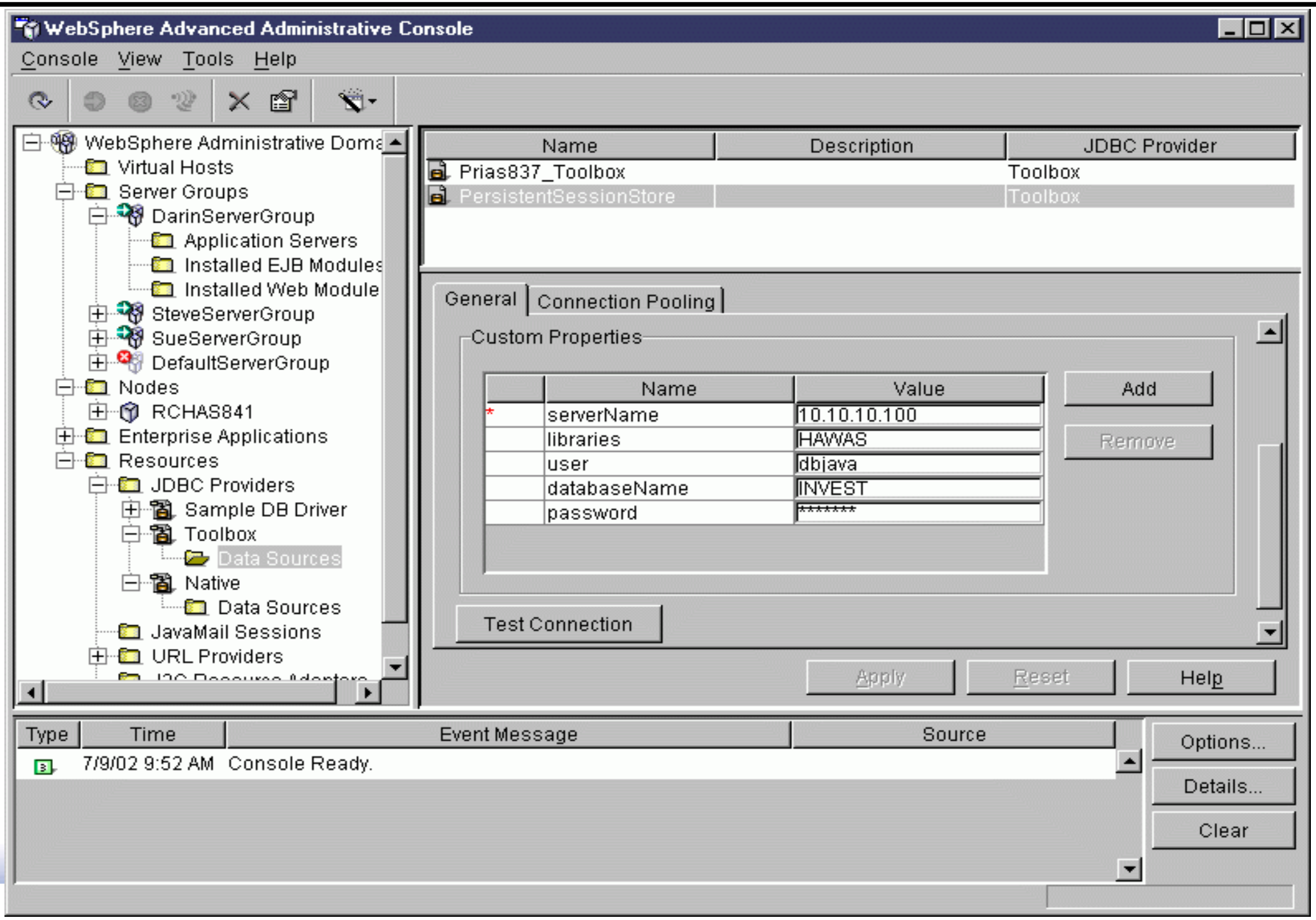


Switchable IASP: WebSphere 4.0 Screenshots

IBM @server



Switchable IASP: WebSphere 4.0 Screenshots



Summary...

IBM @server

- WebSphere and iSeries strengths complement one another to provide an HA eCommerce solution
- The ongoing effort documented in this presentation will facilitate customer support groups with implementing HA topologies
- Expertise in networking, Java, WebSphere, and DB2 is required for successful deployments

Summary

IBM @server

- Recommendations
  - Topologies
    - Dual domain WebSphere or
    - Single domain with horizontal cloning
  - OS/400 V5R1 or later with latest PTF package
    - OS/400 V5R2 required for IASP database support
  - WebSphere Application Server, Advanced Edition, version 3.5.5 (or later point release) or version 4.03 (or later point release)
  - Control programs
    - Edge Server Network Dispatcher and custom advisors
    - iSeries cluster application CRG exit program
  - The latest version of IBM Toolbox for Java JDBC driver
  - iSeries managed solution using redundant LPAR systems
  - Latest HABP software if database replication is desired
  - Commitment control usage in database applications

References...

IBM @server

- Redbook - IBM WebSphere V4.0 Advanced Edition Scalability
  - SG24-6192-00
- White paper - Failover and Recovery in WebSphere
  - http://www7.software.ibm.com/vadd-bin/ftpdl?1/vadc/wsdd/pdf/modjeski.pdf
- Redbook - LPAR Configuration and Management - Working with iSeries Logical Partitions
  - SG24-6251-00
- Redbook - WebSphere Edge Server New Features and Functions in Version 2
  - SG24-6511-00
- Redpaper - Roadmap to Availability on the iSeries 400
  - REDP0501

References

IBM @server

- Redbook - Clustering and IASPs for Higher Availability on the IBM iSeries Server
  - SG24-5194-01
- Redbook - iSeries IASPs: A Guide to Moving Applications to Independent ASPs
  - SG24-6802-00
- iSeries Information Center link to V5R2 IASP information
  - **http://publib.boulder.ibm.com/iseries/v5r2/ic2924/index.htm?info/rzaly/rzalycreateswitchableiasp.htm**
- WebSphere Application Server 4.0 (online documentation)
  - http://publib.boulder.ibm.com/was400/40/AE/english/docs/pvindex10.html
- WebSphere Application Server 3.5 (online documentation)
  - **http://www-1.ibm.com/servers/eserver/iseries/software/websphere/wsappserver/docs/as400v35/docs/**