IBM Systems and Technology Group University 2005

iSeries High Availability

Best of the HA Deep Dive Workshop - *Unplugged*

Eric Hess - iSeries, Americas ATS
texas@is.ibm.com

ON DEMAND BUSINESS™

Today's objectives

- Focus on key items that will help your customer meet their business availability expectations
- Agenda
  - **In 2005, approach HA differently**
    - **Assess, validate, plan, staff, project manage, test, execute, test, validate, test**
  - **Use availability technologies as building blocks for HA solutions**
  - **SCON and HA - design them to work together**
  - **If you leave with anything today, remember this . . .**

Why we are here - the iSeries HA evolution

1988
$$$$$/CPW
AS/400-B60, 5.5 CPW, 480 Users Max, 56GB Disk

2004
$/CPW
i5 595, 165000 CPW, Unlimited, 300TB+ Disk

Flexibility - capacity - power
(and complexity)

Good old days!

- Data entry and reporting (traditional)
- Token Ring, SNA, SNADS, 5250
- Disk measure in MBs - simple, single pool
- Out of the box solution
- 8 x 5 IT operations
- IT supports the business

- Traditional, real-time, ondemand, e-biz, collaboration
- LPAR, multi-apps, OnDemand, Multi-OS
- TCP/IP, internet, ODBC, virtual resources
- Disk measured in TBs - slow, fast, SAN, WORM
- Mission critical - highly secure
- 24 x 365 business, international
- IT is the business !!

HABP owns HA

- 1 HA Business Partner

Multi-server HA technologies
(from one to tons)

- 6 HA Business Partners
- IBM IASPs and Switched Disk Clusters
- IBM iSeries Copy Services for ESS
- IBM Cross Site Mirroring
- IBM Domino Clusters
- IBM AIX, Linux and Windows Clusters

IBM & HABP owns HA

Resources, skills, planning, consulting, systems mgt. disciplines

Good old days!

Managed availability & complexity

Hess 2004

Reliability vs. availability

- High availability must not be confused with great reliability
  - A loss of control can result when there is a total dependence on vendor quality and support in order to meet business HA requirements
    - *"IBM, if you didn't have quality problems, I wouldn't have availability problems."*
    - *No product is error-free - if it's built by a human, it will fail*
  - But, a 'Chain Can Never Be Stronger Than Its Weakest Link', *for example*

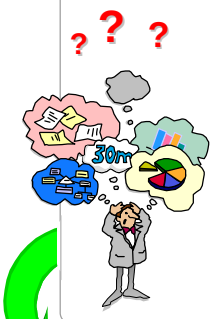
Availability of Technology alone:		**99.999%**
Availability of People alone:		**98%**
Availability of Process alone:		**98%**
The overall availability is now:	**99.999% * 98% * 98% =**	**96.039%**

  - Your system has gone from a theoretical down time of less than 5 ½ minutes for the technology alone to a combined down time for the service of over 14 days!!

Misconceptions, mistakes, bad habits, etc.

- **AS/400 740 - international finance - *misconceptions***
  - CPU failure in one partition caused an outage of all 6 partitions
  - CIO and IT assumed that LPAR isolated outages to the affected partition
  - *Incorrectly believed their RS/6000s did not have this exposure*
- **iSeries 890 and SAN - retail - *bad systems management***
  - Accidental LUN removal on ESS by inexperienced operator destroyed production ASP
  - Cause long unplanned outage - full partition system reload (HA impl. never completed)
  - *Wanted flexibility of SAN, but made no investment in operator education or change management, documentation and procedures*
- **iSeries 830 - manufacturing - *technology under utilization***
  - Multi drive failure within the same RAID-5 parity set caused a complete system loss
  - First disk failure message in QSYSOPR missed for 30 days - System Agent inactive
  - *Business expected 24x7, but assumed that RAID-5 was enough protection*
- **AS/400 640 - banking - *bad design***
  - Application testing on production server caused 4+ hour outage
  - Sr. developer keystroke error caused test batch job to run at priority 04 (rather than 40). Consumed entire server for 4.5 hours
  - *A sandbox should never be placed inside the house - it belongs outside (e.g. LPAR)!*

Common inhibitors to high availability



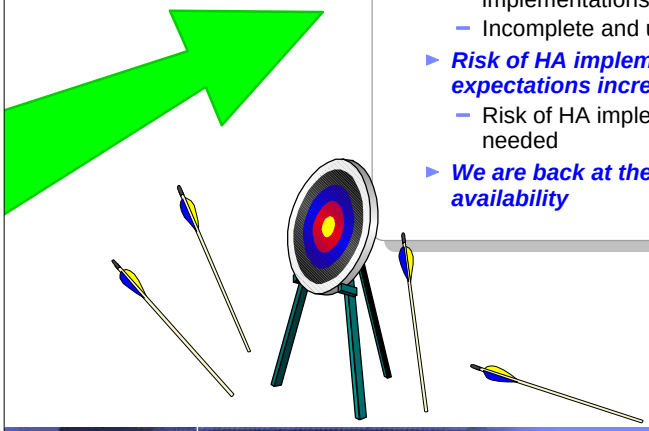
- **It can start with our customer:**
  - ***Availability is owned by operations / analysts - but not the business***
    - No executive owner of Business Continuity
    - Unknown availability requirements or SLA (including RTO/RPO)
    - Investment in availability is not fully justified
    - Lack of application design for availability
  - ***Poor commitment to systems management disciplines***
    - Lack of change management and of testing contributes to instability
    - Undocumented, informal procedures
    - Problems unresolved due to lack of identification, process, follow-up
  - ***High availability is confused with reliability***

...and...

- **We (IBM and Partners) can compound this problem:**
  - **High availability is oversold - selling a technology, not a solution**
    - A crisis is used to improperly position or sell HA
    - Lack of guidance for planning, resources, and timelines
    - Reluctance to sell consulting services - "we need to make this server sale"
    - Lack of focus on single server design and Single Point of Failures (SPOFs)

Inhibitors to high availability . . . *cont*

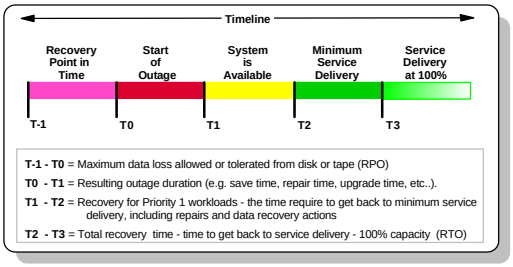
- **The Results can be:**
  - ***A business over simplifies the problem and solution***
    - Customer desire or perception of plug and play HA technology answer
    - Expectations are for easy, quick turn around HA implementations with minimal skills
    - Incomplete and untested HA implementation
  - ***Risk of HA implementations failing to meet expectations increases***
    - Risk of HA implementation not masking outage when needed
  - ***We are back at the customer rebuilding or reselling availability***



If it cannot be defined, it probably cannot be built

Availability requirements must be analyzed and defined

- **RTO:** Recovery Time Objective - *Time to get back to key production delivery*
  - More than 4 days is acceptable
  - 1-4> days
  - <24 hour
  - <4 hours
  - <1 hour
  - Approaching zero (near immediate)
- **RPO:** Recovery Point Objective - *Maximum data loss allowed or tolerated*
  - Last save (weekly, daily, ...)
  - Start of last shift (8 hrs)
  - Last major break (4 hrs)
  - Last batch of work (hours)
  - Last transaction (seconds to minutes)
  - In-flight changes may be lost (power loss consistency)
  - Now (near immediate)



Example 1. If it cannot be defined, it probably cannot be built

Identify recovery requirement for each application, LPAR, and Server

Customer example - pop quiz

- Customer: **Single iSeries 820, V5R1, JD Edwards - PeopleSoft**
  - Stable environment, though customer bumping up against backup windows
  - No HA in place - customer considering HA in future
- Activity: **Upgrades to 570 and V5R3 (same PeopleSoft release and design)**
  - 820 is removed as part of the upgrade
- Outage: **New V5R3 QSQL/QTEMP defect causes application failure**
  - 23 hours downtime in 28 days
  - 2-3 weeks for IBM to create i5/OS fix
- ***How could this situation have been avoided?***

- How could this situation have been avoided?
  - n-1 release strategy integrated with HA
    - Keep 820, implement HA - run at V5R1 or V5R2
  - Stress testing/benchmark
- Complications
  - 820 may have been lease machine, had to be returned - what would be used for HA?
  - A 2nd HA 570 (V5R3) may experience the exact same OS defect
  - n-1 application support not always possible with HA-ISV combination
  - Stress testing and benchmarks are difficult, complex, expensive

IBM Systems and Technology Group University 2005

General rules (summary)

- Reliability is not the same thing as availability
- Look for single points of failures that a customer may not be able to tollerate
- Crisis time is rarely the right time to plan and propose HA
  - In a crisis, people may agree to anything
  - Instead, *commit* to a follow up HA Planning Session later
- Make sure the customer has defined their HA requirements
  - Invest the proper amount of time analyzing customer requirements
  - Make sure what you design and sell meets those requirements
- Educate your customers on the reality of their HA requirements, and the commitments required of them to be successful
- Increase availability design early in the sales cycle
- Plan early, and utilize consulting services early in the sales cycle to address complex planning

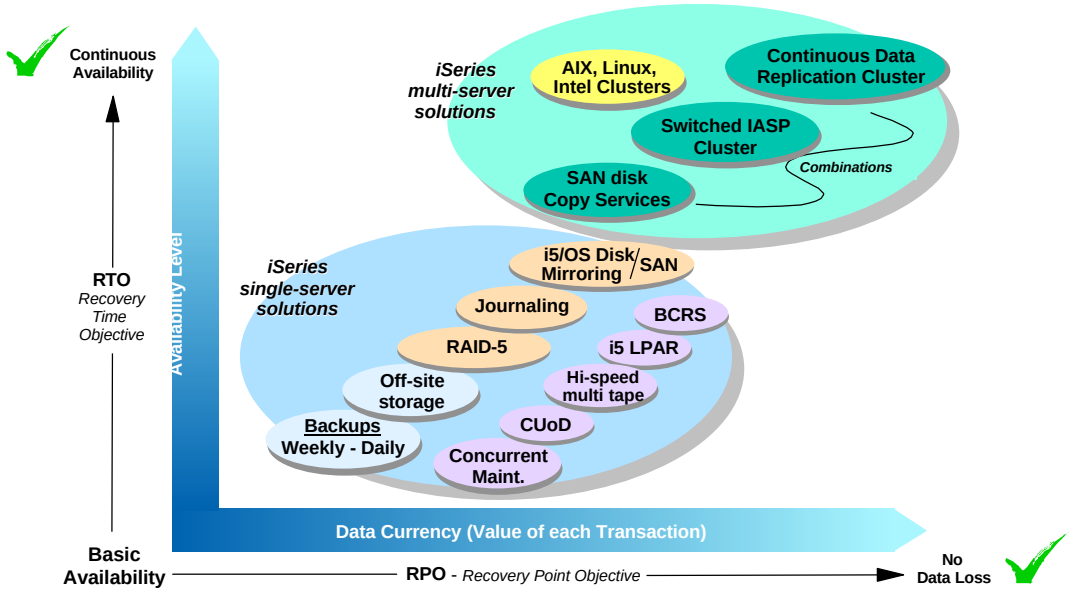
© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Availability Technologies

Building blocks for HA

IBM Systems and Technology Group University 2005

iSeries availability building blocks and strategies



© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Journaling planning considerations

- All journaling implementations must review the need for the OS/400 **'High Availability Journal Performance'** feature
  - Priced feature of OS/400 (V5R2) - OS/400 option 42
    - For more details see Doc 21070117 at: www.ibm.com/support
  - Previously the *'Batch Journal Cache PRPQ'*, PRPQ can be upgraded to OS/400 option
- Recommendations
  - Use the **latest disk and I/O** technology available
    - IO Adapters with largest write cache
    - Fastest disk arms
  - Utilize IOP level disk mirroring - the extra write cache helps
  - Add sufficient disk arms to the journal receiver ASP
  - Don't just focus on disk capacity, think about bandwidth and parallelism
  - Specify ***MAXOPT1** to get the broadest arm usage
  - Use **Journal Minimal Data** as needed
  - Utilize consulting resource expertise when required
- More details:
  - Information Center (- >Systems management' -> Journal management)
  - Striving for Optimum Journal Performance - Redbook - SG24-6286

What HA technologies require journaling?

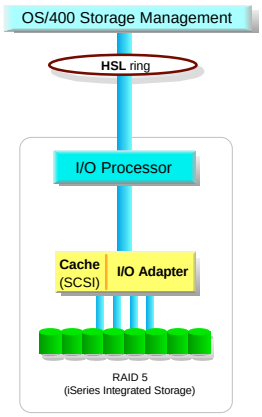
Answer: ALL OF THEM!! No DB server runs w/o journaling

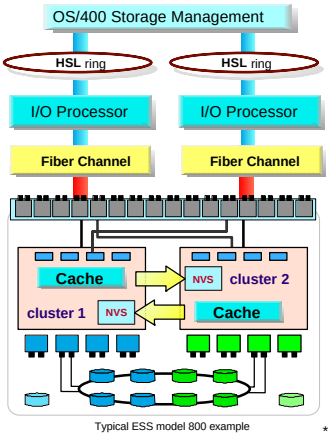
Use the appropriate level of storage protection

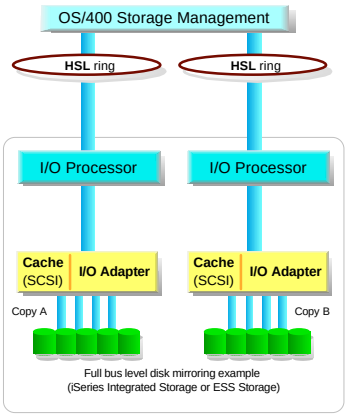
Integrated RAID 5
- ***Good*** storage availability
- Considerations:
  - Single path to drives
  - Single copy of cache
  - Risk of multi-drive failure

ESS 800 RAID 5
- ***Better*** storage availability *
- Considerations:
  - Redundant cache in ESS ✓
  - Reduced risk from multi-drive loss - hot spare drive ✓
  - Multiple fiber paths to drives *(V5R3)* ✓
  - Chance of outage from SAN maint.

Integrated i5/OS Disk Mirroring
- ***Best*** storage availability
- Considerations:
  - Multiple paths to drives ✓
  - Redundant cache on IOA ✓
  - Minimal risk from multi-drive loss ✓
  - Concurrent maintenance ✓



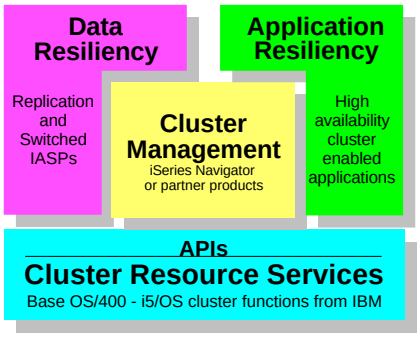




* Can be a 'best' solution if designed for maximum availability

Clustering

- A property of the Operating System
- Provides the logical connections between resilient data groups
- Can enable the automation of physical and logical switching
- Can enable a resilient application to be “switched”, activated and repositioned to a defined state
- Enables the automatic sequencing of events that bring the user, application and data to a coherent production state automatically
- Application design can be the primary limiting factor



- ***Heart beating***
- ***IP Address Takeover***
- ***Reliable internal cluster communications***
- ***Switchover administration***

OS/400 - i5/OS Option 41 – HA Switchable Resources

IBM Systems and Technology Group University 2005

Notes: iSeries and i5 clusters

A Cluster can be defined as a configuration or a group of independent servers that appear on a network as a single machine. Or stated another way, a cluster is a collection of complete systems that work together to provide a single, unified computing resource. The cluster is managed as a single system or operating entity. It is designed specifically to tolerate component failures and to support the addition or subtraction of components in a way that is transparent to users.

The major benefits that clustering can offer a business are continuous or high availability of systems, data, and applications, simplified administration of servers by allowing a customer to manage a group of systems as a single system or single database, and increased scalability and flexibility by allowing a customer to seamlessly add new components as business growth requires.
Here are some of the attributes normally associated with the concept of clustering.

High availability/continuous availability
Simplified single system management
Scalability/flexibility
High-speed interconnect communications
Shared resources
Workload balancing
Single system image

It is important to note that there are several interpretations or implementations of what a cluster is. Various computer manufacturers have different cluster solutions. Most of these clusters were brought about to solve the limited horizontal growth in distributed systems, the basic idea of which was that a number of systems closely coupled together can provide the capacity required for a growing business. Sometimes this is referred to as load balancing: when a client job addresses a server in a cluster to get some work done, it is automatically directed to the server with less workload running at the moment.

Some application software packages running on the iSeries, as for example SAP, can accomplish the same thing.

However, the most important aspect of clustering is high availability, the ability to provide businesses with resilient processes.

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Notes: iSeries and i5 clusters

Clusters on the iSeries and i5: An iSeries clustering solution offers continuous availability to meet your operational business demands 24 hours a day, 365 days a year (24 x 365). This solution, called OS/400 Cluster Resource Services, is part of the OS/400 operating system and provides failover and switchover capabilities for your systems that are used as database servers or application servers.

If a system outage or a site loss occurs, the functions that are provided on a clustered server system can be switched over to one or more designated backup systems that contain a current copy (replica) of your critical resource. The failover can be automatic if a system failure should happen, or you can control how and when the transfer will take place by manually initiating a switchover.

The iSeries cluster uses the separate server or shared-nothing model. That is, cluster resources are not physically shared between multiple systems. Critical resources are replicated between nodes. You might view the resource as shared since it is accessible from other nodes. However, at any given moment, each resource is owned, or hosted, by a single system.

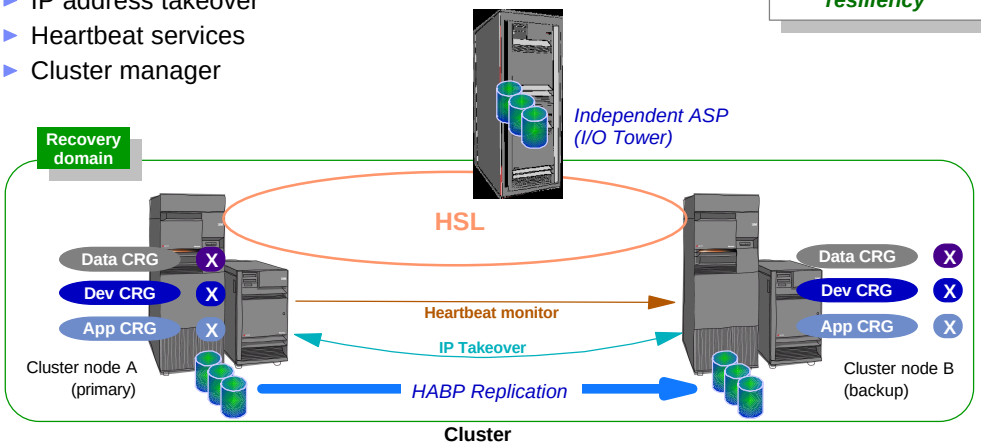
The most important aspect of clustering is high availability, the ability to provide business with resilient processes.

IBM Systems and Technology Group University 2005

Cluster Resource Services example

- Integrated Cluster Resource Services
  - Objects to define clusters and cluster resource groups
  - APIs for cluster control
  - IP address takeover
  - Heartbeat services
  - Cluster manager

Used for both IASP based resiliency and HABP based resiliency



© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Notes: Cluster example

Cluster Resource Services, a component of OS/400, provide:

Tools to create and manage clusters, the ability to detect a failure within a cluster, and switch over and fail over mechanisms to move work between cluster nodes for planned or unplanned outages.

A common method for setting up object replication for nodes within a cluster. This includes the data objects and program objects necessary to run applications that are cluster enabled.

Mechanisms to automatically switch applications and users from a primary to a backup node within a cluster for planned or unplanned outages.

Heartbeat monitoring utilizes a low-level message function to constantly ascertain that every node can communicate with other nodes in the cluster. If a node fails or a break occurs in the network, heartbeat monitoring tries to reestablish communications. If communications cannot be reestablished within a designated time, the heartbeat monitoring reports the failure to the rest of the cluster.

This clustering framework is built around a set of system APIs, system services, and exit programs. This clustering architecture requires teamwork between IBM and business partners to provide the total solution. IBM's clustering initiatives include alliances with high-availability business partners (HABPs) and independent software vendors (ISVs) and the development of standards for cluster management utilities. More about this later in this presentation.

Clustering is the latest iSeries technology for high availability. The clustering functions from V4R4 and subsequent releases build on existing iSeries availability support such as journaling, commitment control, mirroring, and OptiConnect. OS/400 clustering delivers a standard for transferring applications, and their associated data, programs, and users, from one iSeries to another.

Note that combined with the benefits of continuous availability, the second server can be utilized as a production server by having it perform tasks such as save operations, database queries, batch reporting, and to act as a web server for inquiries. In this way some of the work is offloaded from the production server.

The message function of cluster resource services keeps track of each node in a cluster and ensures that all nodes have consistent information about the state of cluster resources. Reliable messaging uses retry and time-out values that are unique to clustering. These values are preset and cannot be changed. These values are used to determine how many times a message will be sent to a node before a failure or partition situation is signaled. For a local area network (LAN), the amount of time it will take to go through the number of retries before a failure or partition condition is signaled is approximately 45 seconds. For a remote network, more time is allowed to determine whether a failure or partition condition exits. You can figure approximately four minutes and 15 seconds for a remote network.

Heartbeat monitoring ensures that each node is active. When the heartbeat for a node fails, the condition is reported so the cluster can automatically fail over resilient resources to a backup node. A heartbeat message is sent every 3 seconds from every node in the cluster to its upstream neighbor. In a network, the nodes expect acknowledgment to their heartbeat from the upstream node as well as incoming heartbeats from the downstream node, thus creating a heartbeat ring. By using routers and relay nodes, the nodes on different networks can monitor each other and signal any node failures.

OS/400 cluster service jobs are a set of multithreaded jobs. When clustering is active on an iSeries, the jobs run in the QSYSWRK subsystem. The jobs run using the QDFTJOBD job description. Should any cluster resource services job fail, no job log is produced. In order to provide a job log, change the LOG parameter of the job description to a logging level that produces job logs.

IP takeover is the ability of a backup system to take over the IP address of a primary system in case of a failure on the primary system.

IBM Systems and Technology Group University 2005

Notes: Resiliency

A very important concept in the discussion of clustering is **resiliency.**

The purpose of clustering is improved availability, which depends on two interrelated concepts: data resiliency and application resiliency. Technologies such as **Device resiliency** and **Data resiliency** ensures that the backup system has all the information necessary to run critical production jobs when control is transferred from the primary system. Device resiliency insures the storage containing data is available. IBM provides Device Resiliency with technologies such as Switch Disk Clusters. Data resiliency requires synchronizing objects across the nodes in the cluster resource group.

HABPs have many tools to deliver iSeries data resiliency. **IBM has chosen to support those tools rather than to create competing data resiliency solutions**. Existing high-availability solutions synchronize data files, programs, and related objects such as data areas, job queues, and user profiles using a combination of custom applications and OS/400 functions (e.g. remote journaling). All these functions are needed to support clustering.

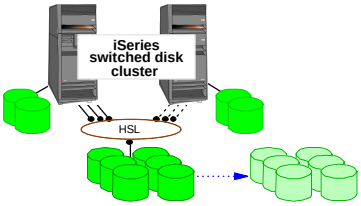
Application resiliency can be defined as the ability to run an application on more than one node in a cluster. Ideally, when an application switches from one node to another, the user experiences no disruption at all and is not even aware that the job has been switched to a different server. Realistically, the disruption the user experiences can range from a slight delay to an extensive application restart. The user may have to sign on to the new server, restore or resynchronize data, restart the application, and reenter any partially completed transactions. The more resilient an application is, the more this disruption is minimized.

The problem with solutions which only focus on data is that they can not be 24x365 or provide zero downtime. Switching between systems requires application resiliency and transaction signaling. That is why clustering technology was introduced in V4R4 and why the focus now is to include the application and the data together in a comprehensive solution called the cluster. The external disk vendors (including IBM's Shark box) provide at best a data copy function and therefore cannot integrate their disk replicating technologies into a clustering solution as they have no knowledge of the data currency, the transaction status and of the application architecture.

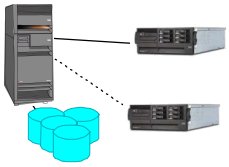
Having given the importance to application resiliency, it is equally important to recognize that in order to obtain continuous availability, the applications have to be designed in a way which allows them to return to their previous known failure state. In other words, the job state and the application state have to be maintained. This cannot be controlled through only a data copy or through any disk storage subsystems.

iSeries cluster solutions

- Data and application resiliency through replication
  - Local and remote coverage; unlimited distance
    - HA/CA/DR/backups
  - a.k.a. HABP, Logical replication, data resiliency, host replication



- Switched Disks: device and application resiliency
  - Local; provides HA
- Switched Disks with Cross Site Mirroring - XSM
  - Local and remote coverage; unlimited distance
    - HA/DR
- Switched Disks with ESS Copy Services
  - Local and remote coverage; unlimited distance
    - HA/DR, backups



- Intel
  - Intel; Hot spare, switched disk, Microsoft Cluster Services, replication cluster, virtual disk
- Linux, AIX
  - Switched Disks, replication, native OS HA

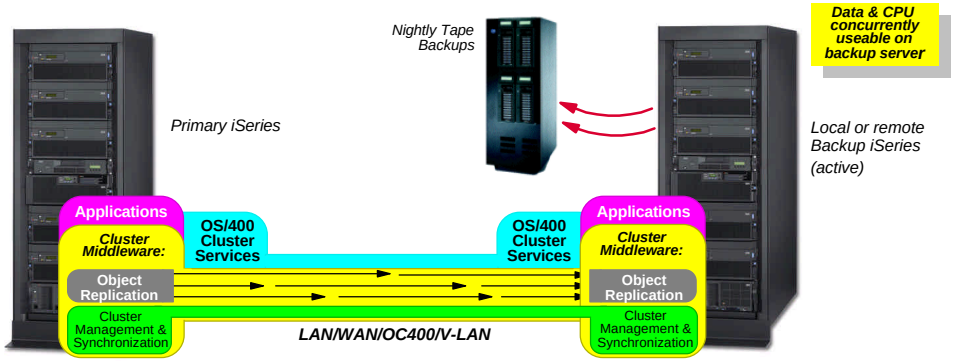
IBM Systems and Technology Group University 2005

HABP solution review and planning

iSeries data replication HA clusters

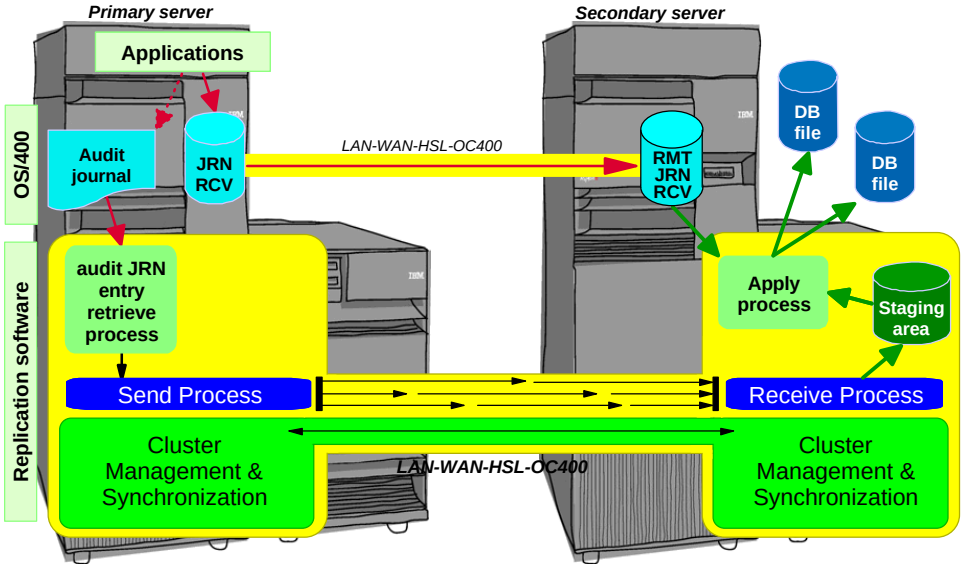
© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

iSeries data and application resilient clusters



- Logical, host based replication
  - Utilizes journaling to enable fast replication of changed data, high currency levels
- Better asset utilization
  - Backup server remains active - IPL/recovery is not required for switchover or failover
  - Offload tape backups to backup server
  - Hardware agnostic
  - Allows the use of different server, storage and OS levels

HABP Replication (remote journal example)



HA Business Partners *(HABPs)*

- 6 HA Business Partners (HABPs) today *
  - Data Mirror (High Availability Suite)
  - iTera (Echo²)
  - Lakeview (Mimix)
  - NoMax (Maximum Availability)
  - Trader's S.A. (Quick EDD)
  - Vision Solutions (Orion)
- 2 others now advertising HA capability - TBD
  - Halcyon Software (Halcyon HA Suite)
  - OS Solutions (OSD-SM HA)
- Customer needs to own responsibility for choosing
  - Avoid trying to own this decision!
  - Common attributes for HABP decision:
    - Functionality - performance - concurrency with latest OS/400 functions - service - support - education - consulting - local skills
  - Customer must use reference

* See iSeries HA website HABP product details and decision making material

From LUG website (customer)
Date: Mon 10/25/2004 07:54 AM
Subject: RE: 24/7 Availability

Discussion: ABC Co. had also done an analysis between the big three (Lakeview, DataMirror and Visions) and we chose to implement HABP-X for our replication.

We are replicating a model 870 system (with 18 data groups configured) to a model 825 system approximately 70 miles away over a DS3 line (approx. 45 MB/s). We are replicating approximately 25 million to 30+ million transactions per hour, and have performed successful switch tests. We plan to perform a switch test every quarter in the future as well. We are very happy with the solution ... it is working quite well for us.

IBM Systems and Technology Group University 2005

HABP planning considerations

- **Latest version of HABP solutions have seen major performance, reliability and automation enhancements**
  - Customers using old versions are dealing with unnecessary work or problems
- **Larger iSeries and i5 server performance and capacities means more focus on HABP sizing required**
  - More workload on a single server can mean an increased number of data groups may add additional overhead
  - The chosen BP solution must plan ahead for this
- **A large volume of non-journaled objects can slow down replication currency**
  - e.g. save/restore checkpoint method required
  - Always journal non-DB objects when possible
    - More efficient, better data currency
- **Customer must plan object replication**
  - Customer should avoid planning short cuts using the *'replicate everything because it's a simpler'* planning strategy
    - May end up replicating temporary work spaces and other unnecessary data
  - Invest in detailed replication requirements
  - **This requires assistance from what customer resources ???**

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Notes: HABP planning - application objects

Temporary objects or work files replication can significantly increase the bandwidth requirement you will need to successfully replicate your data. You want to make sure that you only replicate items absolutely critical to recovery. Another consideration for temporary objects is the potential they could be deleted from the primary system before getting replicated to the secondary system due to job end. Temporary objects of this type could make it impossible for you to keep your systems in sync.

Data queue and data area objects need special consideration depending on they way they are used. The business partner packages give you two ways of replicating these objects, similarly to DB files, meaning as each change arrives it gets replicated to the second system or an option to update the secondary system at timed intervals. The dependency of your data integrity on data areas and data queues and how your application uses these objects are going to be the factors that determine which option you select.

User spaces and user indexes may not be replicated reliably due to the existence of some programming interfaces that do not record changes to these objects in the audit journal. You will have to evaluate whether your application is using these interfaces.

The contents of **message queues and job queues** cannot normally be replicated. If your applications are highly reliant on these object types, you may have to invest in application redesign to deploy a high availability solution. Please note that there is a package in the marketplace by Total Solutions Group that will monitor JOBQ entries, the primary purpose of this package is to do load balancing, but it may be adaptable to high availability.

OS/400 job scheduler or IBM Advanced Job Scheduler or other job scheduling package entries also need to be replicated to the second system. These items are best handled by your change control procedures. Either all entries can be held on the secondary system until it is acting as the primary. For IBM Advanced Job Scheduler, you may use dependency scheduling to control the release of jobs at appropriate times when the system is acting as the primary.

Spool files have special challenges when replicating. Primarily it is the bandwidth required to replicate them. It would be better to change applications to not only produce a spool file, but also a database file for critical reports.

MQSeries requires special planning in any HA environment.

IFS objects, better known as byte stream files, are replicated in their entirety with each change. That means that if your application changes 2 bytes in a 40MB file, the whole file must be transmitted to the secondary system. Special planning for the communications bandwidth, potentially a second line to handle these types of bulk transfers may be required. This is a temporary situation. In an upcoming release of OS/400 BSFs will be included in journaling just as DB2 files are today.

DLOs are handled just like BSFs. However, there are no future plans for journaling these objects, so you may want to consider migrating your DLOs to BSFs.

IBM Systems and Technology Group University 2005

Notes: HABP planning - application objects. . . cont

Commitment control is a good thing! If you have written your applications to take advantage of commitment control, make sure that you set up your high availability solution to take advantage of this. Also, make sure that you determine whether or not the before image journal receiver entries need to be sent to the secondary system.

Constraints and referential integrity normally need to be evaluated as to whether you can leave them in place or need to remove them until you actually use the database on the secondary system as your primary. If constraints and referential integrity need to be removed, you will best be served by developing a program to automatically reinstate them. Also, consider that the business partner synchronization checking tools may recognize the absence of a constraint and point you to nonexistent out-of-sync conditions.

Triggers may be left in place if you can set up filters within your HA solution to ignore transmitting the actions of the triggers. In other words, if the triggers could execute on both the primary and the secondary systems, you may want to let things operate that way. Again, erroneous out of sync indications may occur when triggers are removed.

A V4R4 feature in **DB2/400 UDB that you may be using is data links or URL**s. A difficulty with using this feature is the need to tightly integrate the contents of the IFS with the contents of individual records within database. Discuss your alternatives with your solution provider.

You'll want to look into your use of **RGZPFM** versus taking advantage of reuse deleted records. Remember that the partner solution will send the RGZPFM action to the backup. You could end up with both systems having a file unavailable due to the use of this command. And that is something you're trying to avoid! All partner packages can support reuse deleted records and maybe your application can too.

Recreate sequences - Does your application deploy what we call recreate sequences? That is, does it do a move or rename, create and copy, followed by a delete? These actions can sometimes create the need to perform resynchronization. This series of events also increases your communication bandwidth requirements for data replication. There are alternatives. Discuss them with your solution provider.

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

General rules (summary)

- Usage:
  - All planned outages
  - All unplanned outages
  - Disaster recovery
- Typical attributes:
  - Experienced technology
  - Capable of fast switchover times (5 - 60 minutes)
    - No IPL required, HA server and resources available immediately
    - Very good automation capabilities
  - Requires least LAN/WAN bandwidth of all replication technologies
  - Requires good coordination between application and operations organizations
- Skills:
  - Local consulting and skills usually available via HABP or BP
  - Application staff awareness required
  - Requires slightly more operations staffing than other technologies

- Education
  - HABP specific
  - COMMON
- Consulting/support
  - HABP, IBM ITS, 3rd party

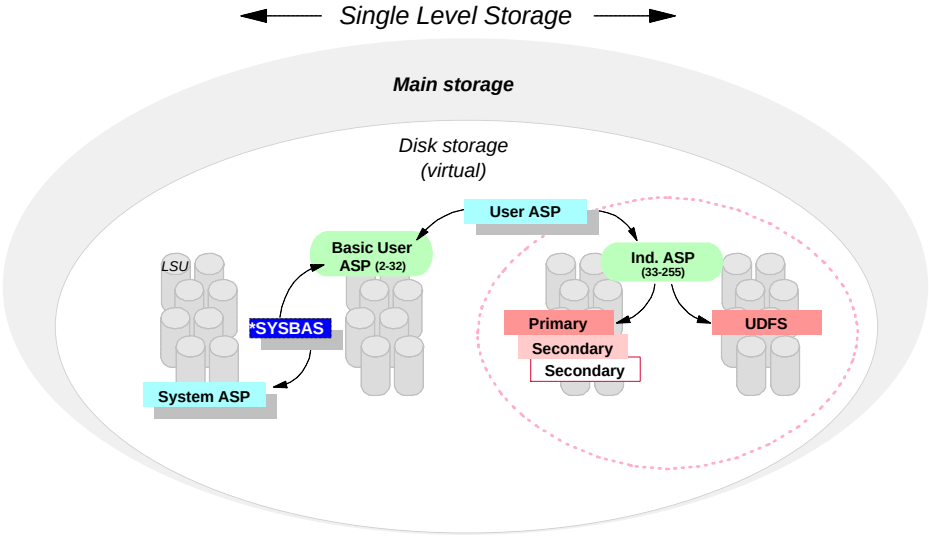
IBM Systems and Technology Group University 2005

Auxiliary Storage Pools - IASPs

Switched Disk Clusters

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

iSeries Storage Layout with OS/400-i5/OS Version 5



Apply new thinking when designing iSeries storage

- More storage design granularity
- More robust storage availability
- Enablement for new cluster solutions
- Mission critical ASPs require disk mirroring

IBM Systems and Technology Group University 2005

Notes: iSeries Storage Layout

In the chart you can see the full breadth of single level storage (SLS) spanning main storage and disk storage. Within the disk storage the disk are allocated to either the system ASP or a user ASP. Both these spaces are part of SLS. The user ASP portion can be further subdivided into Basic ASPs or Independent ASPs
*SYSBAS is the system ASP and all Basic User ASPs

Basic ASP = Auxiliary Storage Pool (Disk Pool)

- An ASP is a software definition of a group of disk units on your system. An ASP provides a way of organizing data to limit the impact of storage-device failures and to reduce recovery time.
- A basic disk pool is used to isolate some objects from the other objects that are stored in the system disk pool. Data in a basic disk pool is always accessible whenever the server is up and running.

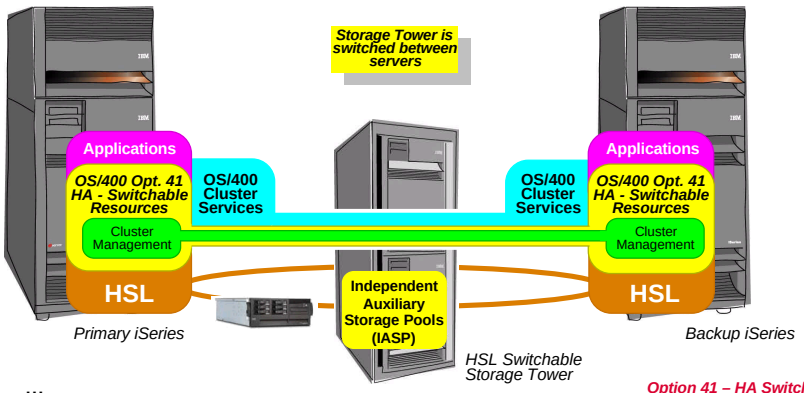
IASP = Independent ASP (Independent Disk Pool)

- An independent ASP can be made available and made unavailable to the server without restarting the system. When an independent ASP is associated with a switchable hardware group, it becomes a switchable ASP and can be switched between one server and another server in a clustered environment.
- Differences from Basic User ASP
  - Introduced in R510
  - Identified by device name
  - Overflow not allowed
  - VRYCFG used to independently bring online or take off-line

Type of ASP	OS/400 release supported	ASP numbers	Maximum quantity supported on the system
System ASP	All	ASP 1	1
Basic User ASP1	V4R5 and earlier	ASP 02 - ASP 16	15
Basic User ASP1	V5R1	ASP 02 - ASP 32	31
Independent ASP	V5R1	ASP 33 - ASP 99	67
Independent ASP	V5R2	ASP 33 - ASP 255	223

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

iSeries Switched Disk Clusters



- Device resiliency
  - IASP is switched between primary and backup server
  - Addresses some planned and unplanned outages
  - Not for save window reduction or disaster recovery
  - Limited distances

Option 41 – HA Switchable Resources (Included with Enterprise Edition)

IASP Feasibility Study

Clustering and IASPs for Higher Availability on the IBM eServer iSeries Server - SG24-5194

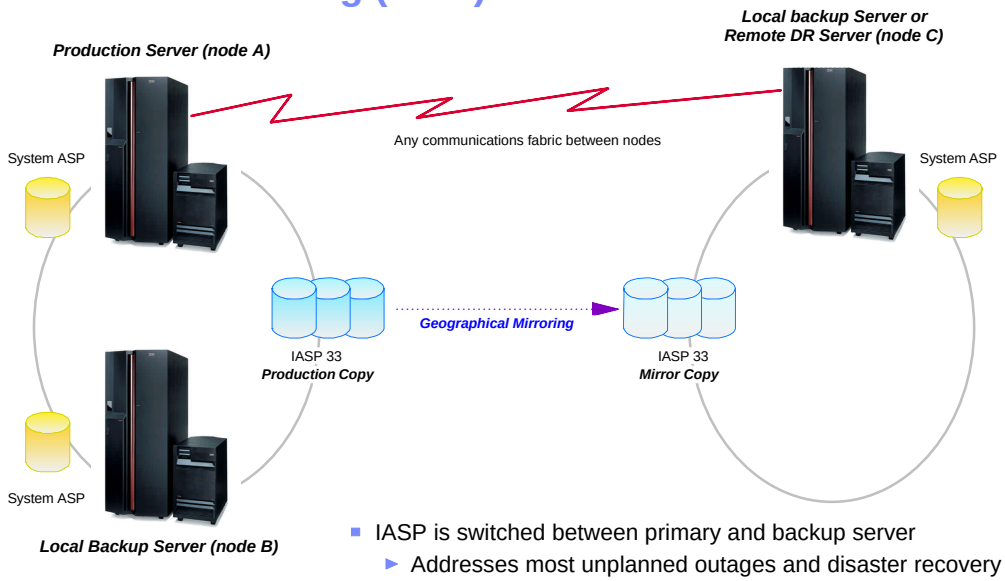
General rules (summary)

- Usage:
  - Some planned outages (CEC maint, some upgrades)
  - Not for backups to tape
  - Some unplanned outages (not for disk failures, maint. or damaged objects)
  - Not applicable for disaster recovery
- Typical attributes:
  - New technologies
    - Requires new education for customer staff
    - Proof of Concept and/or benchmark may be prudent
  - Capable of good switchover times
    - Vary on of IASP required (10 - 60 minutes)
    - Very good automation capabilities
  - Requires participation and commitment of both application and operations orgs
- Skills:
  - Limited skills available - but growing
  - Application staff commitment required
  - Reduced operations staffing than other technologies

- Education
  - IBM iTC's IASP and Cluster courses
  - COMMON
  - IBM iSeries Technical Conference
  - ITSO Forum - Rochester
- Consulting
  - IBM iTC, IBM CTC, IBM ITS

Cross Site Mirroring - XSM

Cross-Site Mirroring (XSM)



- IASP is switched between primary and backup server
  - Addresses most unplanned outages and disaster recovery
  - Not for save window reduction at this time
  - Capable of long geographical distances

Planning for XSM

- Requires V5R3, configured cluster, & Option 41
- Disk capacities of IASP copies have to be similar (but don't require equivalent physical configurations)
- When geographic mirroring is active, cannot access mirror copy
- Mirror copy can only be made available when detached
- If detach mirror copy, a reattach will require a full resynch
  - Use, e.g., for save processing only if have done vary off and can accept resynch time
  - Better choice for save might be:
    - Save-While-Active with BRMS
    - HABP product
    - ESS PPRC with iSeries Copy Services for ESS toolkit
- Typically avoid suspend or detach due to resynch time
  - Geographic mirroring is optimized for large files. A large number of small files will produce a slower synchronization rate.
- If suspend, resume normally requires resynch
- During resynch, disk performance is very important

General rules (summary)

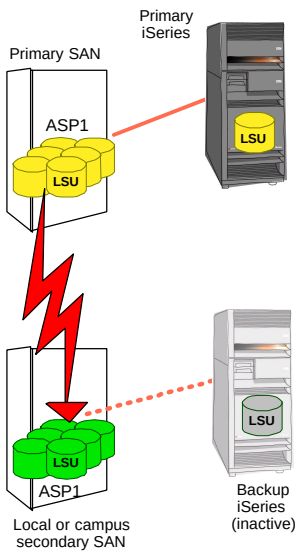
- Usage (today - V5R3):
  - Most planned outages
  - Most unplanned outages
  - Not viable for tape backups
  - Disaster recovery
  - Should be limited to basic HA/DR
- Typical attributes:
  - Same considerations as iSeries Switch Disks Clusters, plus
  - Very new technology
  - Resych between IASP copies can results in outages and performance issues
    - Must understand the issues before proposing
  - Capable of good switchover times
  - Should be limited to smaller or less mission critical environments
  - Proof of Concept, review by XSM expert and Solutions Assurance a must
- Skills:
  - Same considerations as iSeries Switch Disks Clusters

- Education
  - iTC IASP and Cluster courses
  - COMMON
  - IBM iSeries Technical Conference
- Consulting
  - IBM iTC
- Review
  - Contact IBM for assistance with XSM design

ESS and DS Copy Services in an iSeries/i5 environment

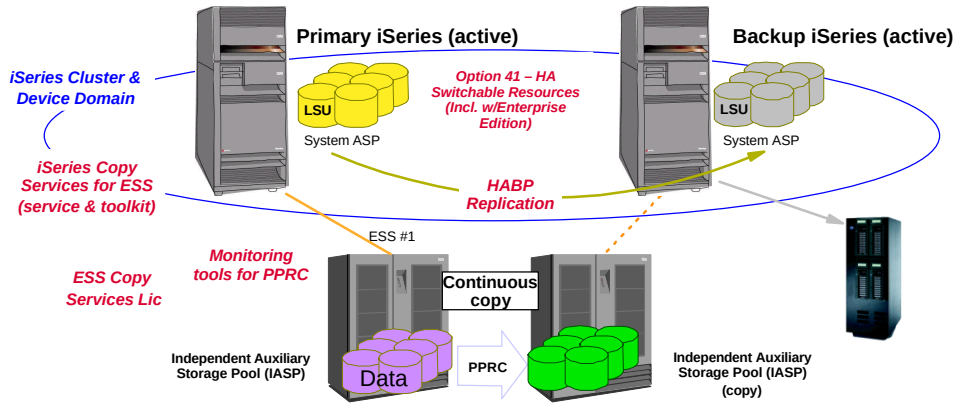
Continuous, synchronous copy

- Desired benefits:
  - Campus/city disaster recovery (up to 103km)
  - Multiple platforms supported
- Considerations:
  - Requires aggressive use of journaling and SMAPP
    - Does not protect data in server memory - risk of data loss
  - Unplanned failover can be many hours
    - IPL required, plus any DB recovery
  - New complexities
    - Manual processes to switch or recover
    - Will replicate physical data damage
  - Higher communication costs (remote)
  - Second SAN disk copy *and* second iSeries are unavailable for other uses
  - Not viable for masking outages from software and hardware upgrades and maint.



Basic Copy Services

iSeries Cluster/ESS Copy Solutions - PPRC



Advanced Copy Services

- HA and DR solution
- No IPLs on source server with IASP
  - No load source recovery required
  - Controlled availability iSeries 8xx
  - 2005 for i5 models
- Considerations
  - Requires IASPs and Clusters
  - Must coordinate with iTC before architecting and selling
  - Requires consulting and customer education

General rules (summary)

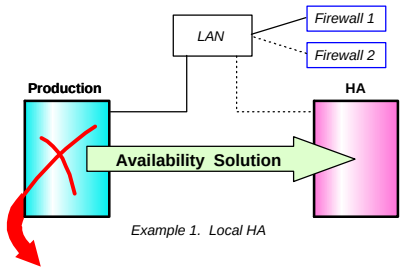
- Usage (ESS with IASPs and Clusters):
  - Most planned outages
  - Most unplanned outages
  - May require 3rd copy of disk for Flsahcopy and tape backup
  - Disaster recovery possible, but no examples today
- Education
  - iTC IASP and Cluster courses
  - COMMON
  - IBM iSeries Technical Conference
- Consulting
  - IBM iTC, IBM CTC
- Typical attributes:
  - Same considerations as iSeries Switch Disks Clusters, plus
  - Good resync capabilities
  - Capable of good switchover times - similar to IASPs
  - Second copy of disk not available for other uses
  - Proof of Concept prudent, review by experts and Solutions Assurance a must
    - Must contact the iTC or CTC before selling this to a customer
    - "iSeries Copy Services Toolkit for ESS" is an offering that must be approved
- Skills:
  - Same considerations as iSeries Switch Disks Clusters, plus
  - ESS sizing and copy services skills
  - "iSeries Copy Services Toolkit for ESS" Skills

Switch times

IBM Systems and Technology Group University 2005

Switch times

Anatomy of a failover using redundant servers - HA is more than just technology



Example 1. Local HA

- Outage occurs . . . now . . .
- Decision must be made to failover (switch)*
- Prepare sever and related hardware *
- Switch communications
- Prepare data as needed *
- Prepare application as needed *
- Bring users-access back online
- Data currency (verify and re-enter) *
- Business back online

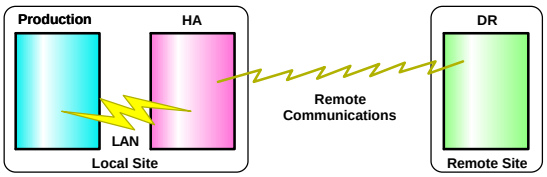
* Huge design and customer preparedness variable

Common timing examples

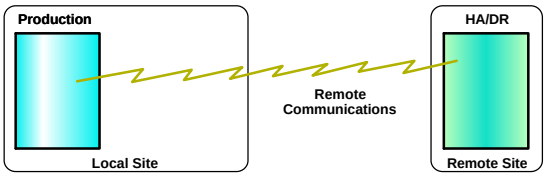
- **Customer switchover or failover decisions**
  - May be automatic or manual (manual is most common)
  - Human variable - minutes to hours
- **Server and hardware preparation**
  - *HABP*
    - Role swap - short minutes to 60 minutes
    - Activate journaling on target - short minutes to hours
  - *Switch Disks, XSM*
    - Role swap - short minutes
    - Vary on IASP - 10 to 60 minutes
  - *Basic SAN Copy Services*
    - Role swap and prep ESS - 5 to 60 minutes
    - IPL server - 60 minutes to many hours
    - Minor config changes (comm, IO, etc.)
  - *Advanced SAN Copy Services*
    - Role swap and prep ESS - 5 to 60 minutes
    - Vary on IASP - 10 to 60 minutes
- **Communications**
  - IP Address takeover - short minutes - 60 minutes
  - SNA - 10 to 30 minutes
- **Data preparation (DB, IFS, etc.)**
  - *Verify data status, repair as needed - none to hours*
- **Applications preparation**
  - *Restart application(s) as needed - minutes to small hours*
- **Bring users back online**
  - *Start subsystems, IP access, etc. - minutes to small hours*
- **Data currency**
  - *Verify, re-enter lost or new data - none to many hours*

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Consolidation and confusion of HA and DR



Example 1. Local HA, separate remote DR



Example 2. Remote HA and DR combined

- HA is local - addresses local outages
  - Very fast recovery times usually expected (RPO/RTO critical)
- DR is remote - addresses site or facility loss
  - Longer recovery times, more data loss is usually assumed

- **HA**
  - The highest availability is usually obtained with a local HA solution
    - Best performance
    - Fastest recovery
    - Easiest to maintain
    - Local skills more apt to be available
- **Active DR**
  - Requires significant planning
    - Comm must be sized for
      - DB, plus IFS, objects, admin workloads, etc.
      - Peak/batch workloads or DR server replication will fall hours behind
    - Remote site skills
  - DR and HA RTO/RPO should be determined separately
- **Combining HA and DR**
  - Combining for the sake of cost reduction can reduce HA levels to that of
- **Consider proposing HA and CBU (DR) models to meet both HA and DR and drive down cost**

IBM Systems and Technology Group University 2005

Application resiliency

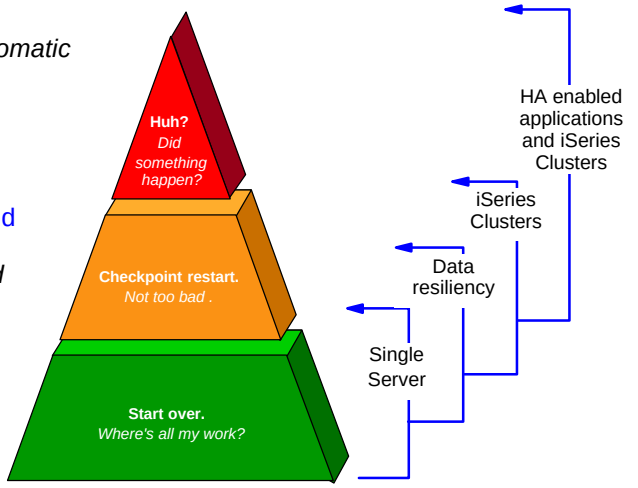
Application service continuity

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Application resiliency

Application service continuity

- Full application resilience - *automatic restart and transparent failover*
- Automatic application restart and automatic recovery - *to last transaction boundary*
- *Automatic application restart* and semi-automatic recovery - *user returned to last architected application "restart" point*
- Automatic application restart - *user manually repositions within application*
- No application recovery – *manual restart with resilient data*



Cluster Enabled Application *Application recovery design* *GUI client/server*

Multi-tiered client- servers *Commitment Control*

IBM Systems and Technology Group University 2005

Notes: Application Resilience

Application resilience can be classified according to the following categories:

No application recovery - After an outage the end users must manually restart their applications. Based on the state of the data, they determine where to restart processing within the application Automatic application restart and then users manually repositioning within applications

Automatic application restart and then users manually repositioning within applications - Applications that were active at the time of the outage are automatically restarted. However, the user must still determine where to resume within the application based on the state of the data.

Automatic application restart and semi-automatic recovery - In addition to the applications automatically restarting, the end users are returned to some predetermined "restart" point within the application. The restart point could be, for example, a primary menu within the application. This is normally consistent with the state of the resilient application data, but the user may have to advance within the application to actually match the state of the data.

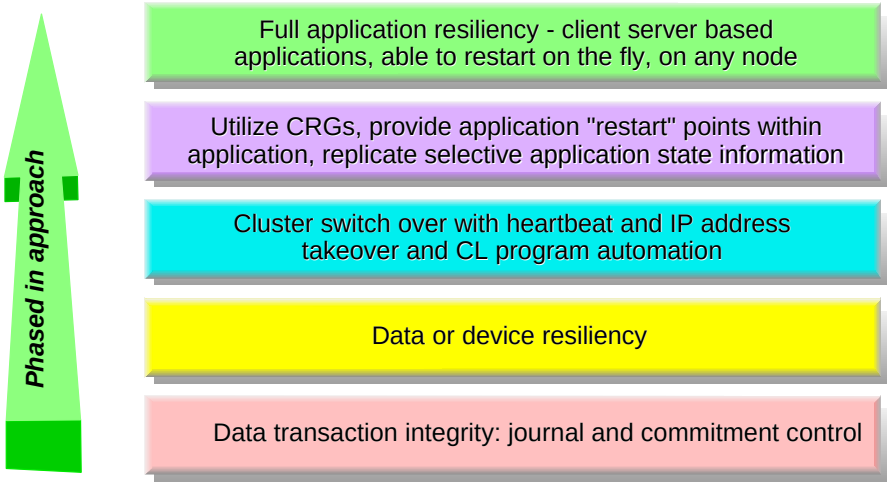
Automatic application restart and automatic recovery to last transaction boundary - The user is repositioned within the application to the processing point that is consistent with the last committed transaction. The application data and the application restart point match exactly.

Full application resilience with automatic restart and transparent failover - In addition to being repositioned to the last committed transaction, the end user continues to see exactly the same screen with the same data as when the outage occurred. There is no data loss, sign on is not required, and no perception of loss of server resources. The user only perceives a delay in response time.

It is important to note that any of the above application resilience mechanisms can be combined with the data resilience mechanisms described below to provide a complete solution.

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

Combing application, data and device resiliency



IBM Systems and Technology Group University 2005

Notes: Combing application and data and device resiliency

IBM's role: To implement clustering, OS/400 delivers integrated system services and a set of APIs to create, change, delete, and manage clusters, nodes and CRGs.

A critical clustering system service is activation of the exit program whenever an event occurs for a recovery domain. Calling the exit program is much like calling a trigger program when a database event occurs. The programmer determines what happens when the program is called, but OS/400 initiates the program automatically on all the nodes in the affected recovery domain. When any change to the cluster environment occurs, OS/400 calls the exit program with information such as the current role of the node and the action code - addition of a node, for instance. Most of the clustering APIs also allow developers to specify up to 256 bytes of information to be passed to the exit program.

Other clustering system services, while less visible, are still important because some functions can be implemented more efficiently at the operating-system level than they could be with any third-party product. IP address takeover (or IP takeover) makes it possible for multiple nodes in the recovery domain to have the same IP address at different times. (Two nodes can never have the same IP address at the same time). IP takeover facilitates transparent switchover in a TCP/IP environment.

Perhaps the most critical clustering function in OS/400 V4R4 is heartbeat monitoring, which constantly checks communications between the nodes in a cluster. If communication between two nodes is lost, heartbeat monitoring attempts to reestablish communications. If a node fails, heartbeat monitoring reports the node failure to the rest of the cluster. Although the HABPs can implement a form of heartbeat monitoring, IBM's system-level implementation of heartbeat monitoring consumes fewer resources and provides a more accurate view of node status. Heartbeat monitoring, in conjunction with other integrated cluster services, ensures that all nodes have a consistent view of the cluster.

HABP's role: OS/400 APIs are available to define resilient resources, and initiate planned switchovers. HABPs have used these APIs to deliver interfaces and tools to manage clusters. These tools complement the existing high-availability offerings from the HABPs and insulate application developers from coding directly to the clustering APIs. Instead of delivering an interface to manage clustering objects, IBM defined a standard for cluster management utilities. Some HAPBs have provided a graphical tool with its own unique personality that satisfies this standard and integrates OS/400 clustering functions with functions in the HABP's high-availability product.

Application developers' role: Application developers handle all application-specific clustering tasks. Some of these tasks work in conjunction with the high-availability solutions. For example, an application developer defines resources such as files and program objects for replication in the automated installation data area, but the HABP usually handles the task of replicating information. Application developers also provide an exit program to restart applications on a backup system after a switchover. For example, the exit program can be used to ensure that the required CRGs are available before an application is restarted.

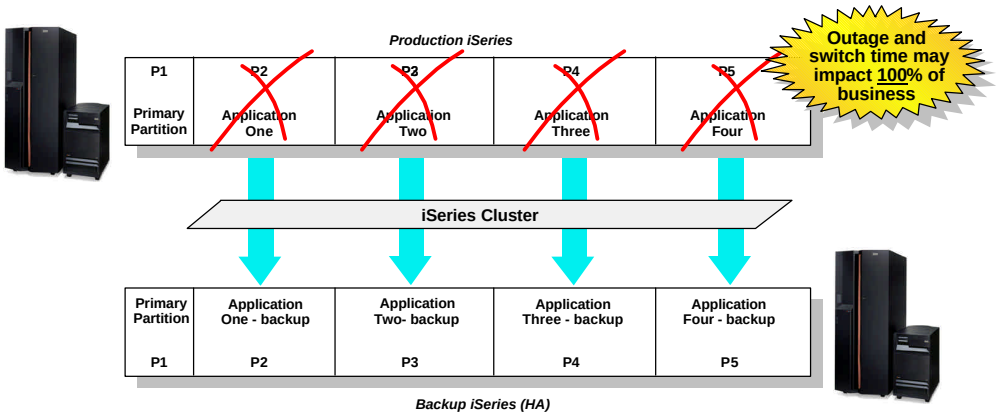
Ideally, an application could be transferred from one iSeries to another and the user could be repositioned at precisely the same spot in a transaction as he was before the switchover began. In pursuit of this ideal, application developers should consider adding checkpoint facilities to their applications. An application checkpoint is usually implemented using commitment control. In conjunction with an appropriate exit program, a checkpointed application has the ability to restart at the last complete transaction.

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

SCON and HA

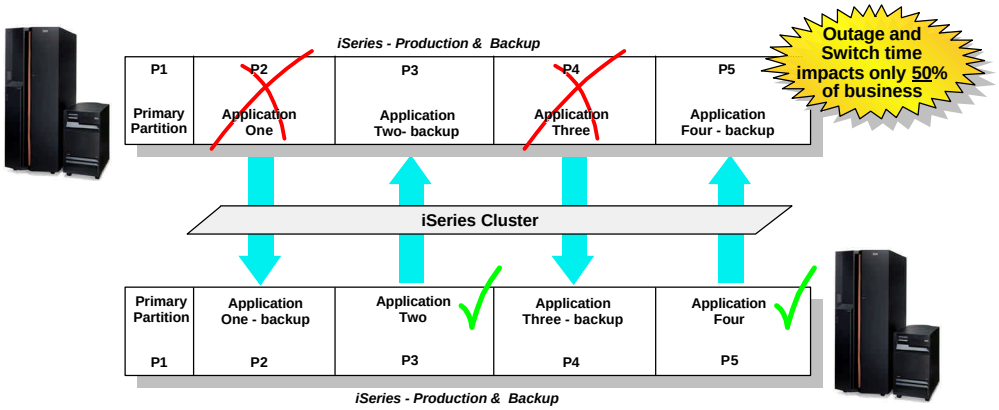
Making them work together

Single Points of Failure in an LPAR environment



- Common SCON and LPAR design: ***Retrofitted high availability***
  - Full SCON benefits, but outage from SPOF impacts 100% of business
  - Beneficial for customers with very good, practiced HA implementations
    - Outage may require that all LPARs switch to backup server - potentially Impacts all users
    - Allows use of new CUoD, 'iSeries for High Availability' and 'Capacity Backup' models
    - May not work well for customer with bad habits, or those having problems with replication

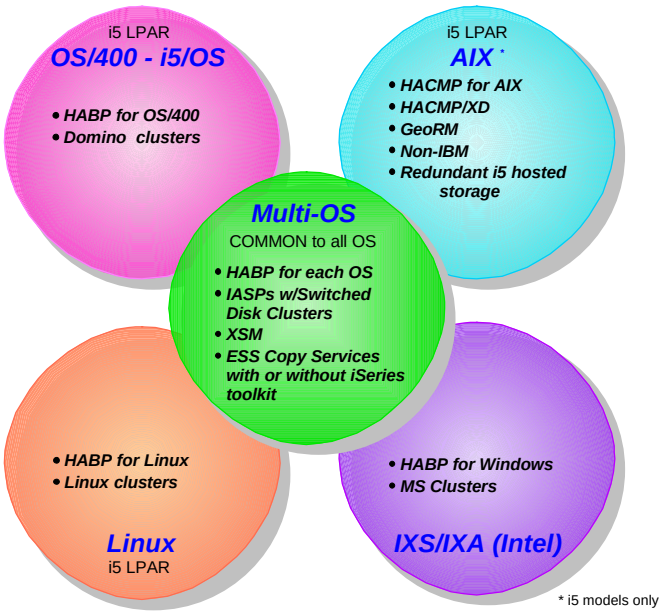
Single Points of Failure in an LPAR environment



- SCON, LPAR and HA solution: ***Integrated high availability***
  - Best of both worlds - reduced impact from SPOF, with benefits of LPAR and SCON
  - Full SCON benefits, but outage only impacts 50% of business
    - Outage requires that only 1/2 of all LPARs switch to backup server
    - Impacts less users, offers more OS and hardware upgrade flexibility
    - Should be considered during any major upgrade activity

i5 availability technologies - multi-OS and LPAR

- SCON design questions
  - What HA solution is used for each OS?
  - Use a common solution?
  - Use an OS unique solution?
- Considerations
  - What was being used before?
  - What are the business availability expectations and requirements
  - What skills are available?
- Complexity and costs vs. availability expectations
  - Easier, single solution?
  - Specialized OS availability solution to maximize it's availability?



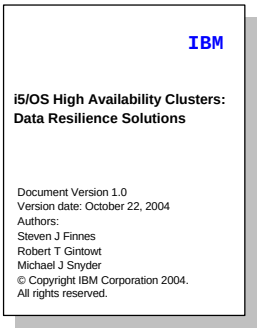
* i5 models only

Session Summary

IBM Systems and Technology Group University 2005

HA Technology selection criteria checklist

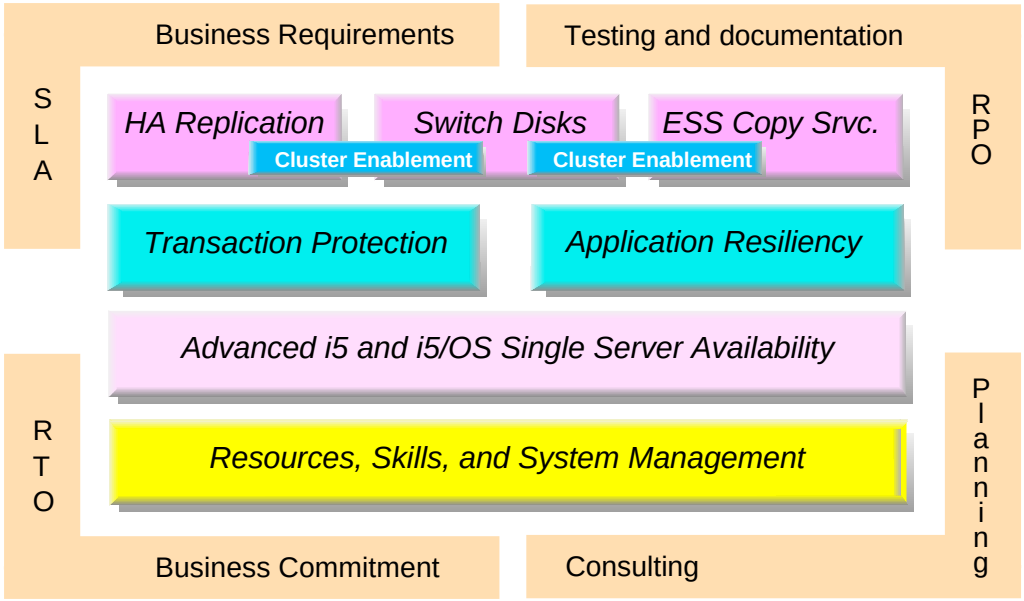
- __ Up time requirements
- __ Recovery time objective
- __ Recovery point objective
- __ Resilience requirements
- __ Concurrent access requirements
- __ Geographic dispersion requirements
- __ Tolerance for end user disruption
- __ Outage type coverage
- __ Cost - initial, ongoing, staffing
- __ Service and support



www.iseries.ibm.com/ha
- or -
www.redbooks.ibm.com

© 2005 IBM Corporation - This educational piece is intended for your use in sales and support It is NOT a deliverable for your customers

HA Solution Summary



The End, Thank You!

Appendix - Related Links

- **iSeries Technology Center (iTC)**
  - http://www-1.ibm.com/servers/eserver/iseries/service/itc/technicalmarketing.htm
- **iSeries Storage and SAN solutions**
  - http://www.iseries.ibm.com/storage
- **iSeries SAN Interoperability Support (current iSeries SAN hub/switch and server support matrix)**
  - http://www.iseries.ibm.com/storage
- **iSeries Benchmark Center**
  - http://www.as400.ibm.com/developer/cbc/index.html
- **Request IBM technical assistance (TechXpress) for IBM America teams only (Americas Partners use PartnerLine)**
  - http://w3.ibm.com/support - or by calling 1-877-707-2727 for US and Canada; or 770-858-5451 for Latin America
- **iSeries High Availability Solutions**
  - http://www.iseries.ibm.com/ha
- **iSeries Performance and Capacity Planning tools**
  - iSeries performance tools/pubs: http://www.iseries.ibm.com/developer/performance/
  - Tools: http://www-912.ibm.com/supporthome.nsf/document/23393024#Planning
  - WLE: http://www-912.ibm.com/supporthome.nsf/document/16533356
  - BMC PATROL® for iSeries (AS/400) - Predict: http://www.bmc.com/products
- **iSeries Storage: Performance Data Collection (planning paper)**
  - IBM: http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100357
  - Partners: http://partners.boulder.ibm.com/src/atsmastr.nsf/WebIndex/WP100357
- **Solutions Assurance**
  - IBM: http://w3-1.ibm.com/support/assure/assur30i.nsf/Web/SA
  - Partners: access Solutions Assurance via PartnerWorld
- **iSeries Systems Management Tools and Solutions**
  - Http://www.iseries.ibm.com/software/#sys
- **Redbooks**
  - http://www.redbooks.ibm.com

Appendix - Related Publications

- **iSeries Clusters and High Availability:**
  - **IBM eServer iSeries Independent ASPs: A Guide to Moving Applications to IASPs - Redbook, SG24-6802-00**
  - **InfoCenter**
  - **Striving for Optimum Journal Performance - Redbook - SG24-6286**
  - **IASP Performance Study - Redpaper, redp3771**
  - **Clustering and IASPs for Higher Availability on the IBM iSeries - Redbook, SG245194**
  - **LPAR Configuration and Management Working with IBM eServer iSeries Logical Partitions - Redbook, SG24-6251-00**
  - **Direct Attach xSeries for the IBM eServer iSeries Server: A Guide to Implementing xSeries Servers in iSeries, SG24-6222-00**
  - **Seven Tiers of Disaster Recovery - Redbook Tip, TIPS0340**

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml: AS/400, DBE, e-business logo, ESCO, eServer, FICON, IBM, IBM Logo, iSeries, MVS, OS/390, pSeries, RS/6000, S/30, VM/ESA, VSE/ESA, Websphere, xSeries, z/OS, zSeries, z/VM

The following are trademarks or registered trademarks of other companies

Lotus, Notes, and Domino are trademarks or registered trademarks of Lotus Development Corporation
Java and all Java-related trademarks and logos are trademarks of Sun Microsystems, Inc., in the United States and other countries
LINUX is a registered trademark of Linux Torvalds
UNIX is a registered trademark of The Open Group in the United States and other countries.
Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.
SET and Secure Electronic Transaction are trademarks owned by SET Secure Electronic Transaction LLC.
Intel is a registered trademark of Intel Corporation
* All other products may be trademarks or registered trademarks of their respective companies.

NOTES:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput 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 improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

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

Any proposed use of claims in this presentation outside of the United States must be reviewed by local IBM country counsel prior to such use.

The information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.