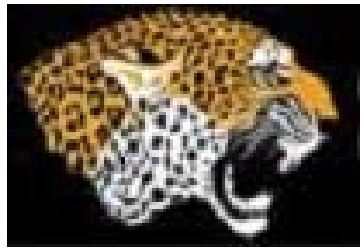
Informix Dynamic Server



Continuous Availability in IDS11

Madison Pruet

IDS Replication Architect

June 28, 2007

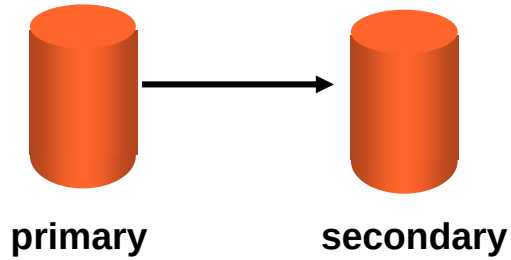
Agenda

- **Why Replicate?**
- **Availability solutions prior to IDS11**
- **Business cases lending to new features**
- **Characteristics of new functionality**
- **Highest degree of availability solutions**

Why Replicate?

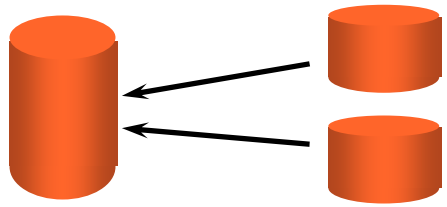
High Availability

Provide a backup site for failover.
Also provide some additional reporting from secondary.



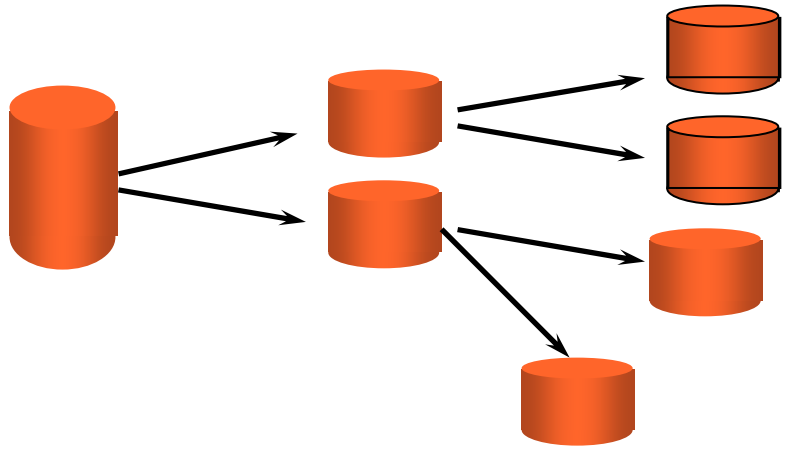
Data Consolidation

Consolidate remote data into a central repository.

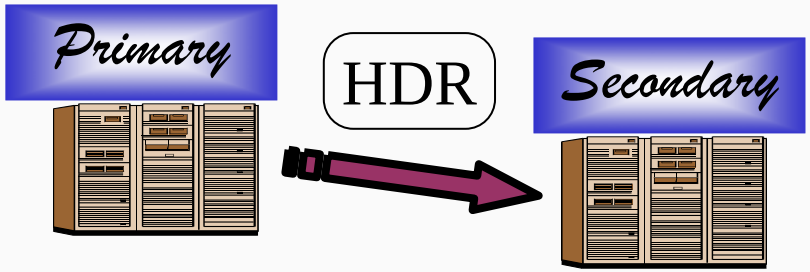


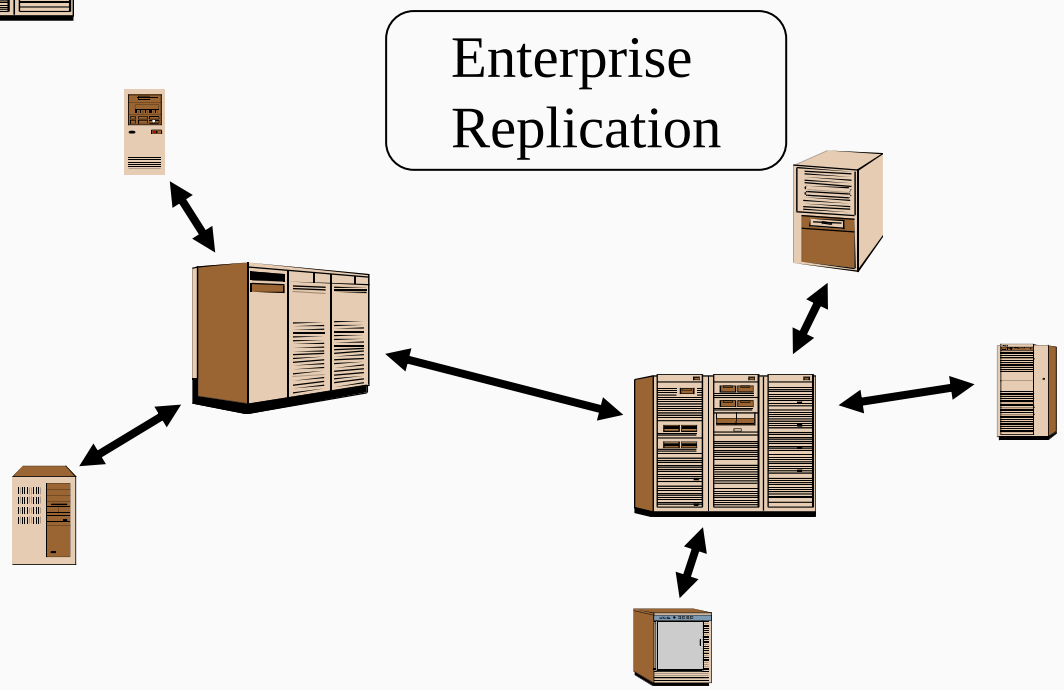
Distributed Availability

Distribute information from a central server.



Availability Offerings prior to IDS11



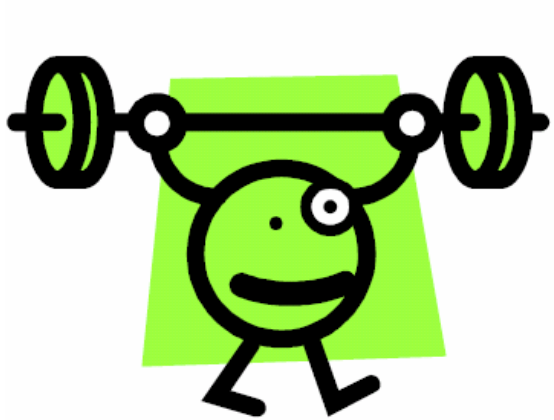


Strengths of HDR

- **Easy setup**
  - Just backup the primary and restore on the secondary
  - No significant configuration required
- **Secondary can be used for dirty reads**
- **Provides failover to secondary**
  - Automatic failover when DRAUTO is set
- **Stable code**
  - Has been part of the product since IDS6
- **Integrates easily with ER**

ER Strengths

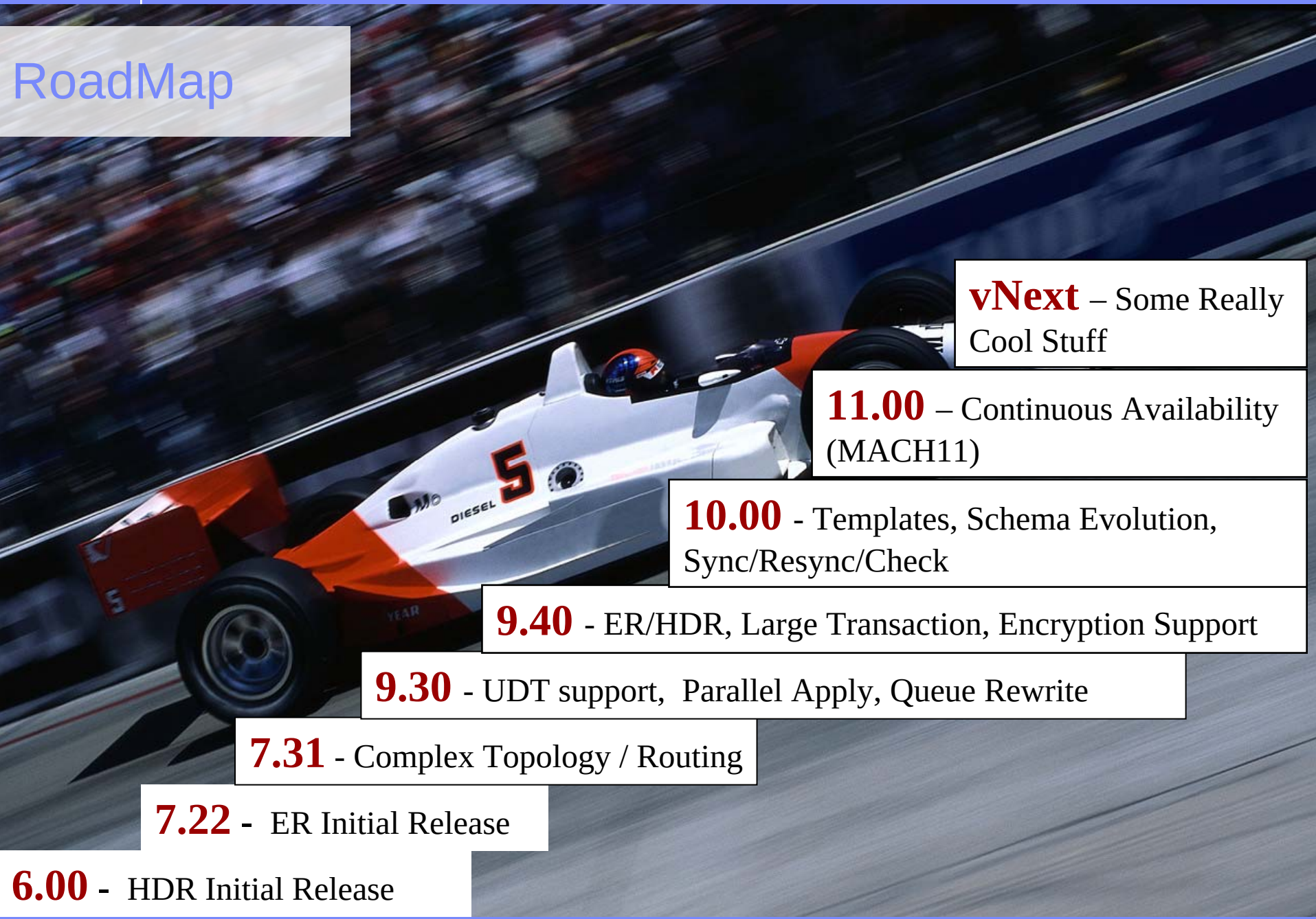
- **Flexible**
  - Chose Columns/Rows to Replicate
  - Chose where to Replicate
- **Supports Update anywhere**
  - Conflicting updates resolved by:
    - Timestamp, Stored Procedure, Always Apply
- **Completely implemented in the Server**
  - No additional products to buy
- **Based on log snooping**
- **Supports heterogeneous OS and IDS versions (rolling upgrades)**



Differences between HDR and ER

HDR	ER
Provides single primary and single secondary	Allows configurable source(s)/target(s) and supports peer-to-peer
Primary and secondary must run the same executables and have similar disk layout – they are mirror images	Source/target do not have to be the same – they do not have to be mirror images
Secondary restricted to ‘dirty read’ report processing	Allows full transactional usage of both source and target
Simple to set up and administer	Setup and admin. more complex
Does not support blobspace blobs	Supports blobspace blobs
Replication can be synchronous	Replication is asynchronous
Primary purpose is for **high availability**	Primary purpose is for **data distribution**

RoadMap



vNext – Some Really Cool Stuff

11.00 – Continuous Availability (MACH11)

10.00 - Templates, Schema Evolution, Sync/Resync/Check

9.40 - ER/HDR, Large Transaction, Encryption Support

9.30 - UDT support, Parallel Apply, Queue Rewrite

7.31 - Complex Topology / Routing

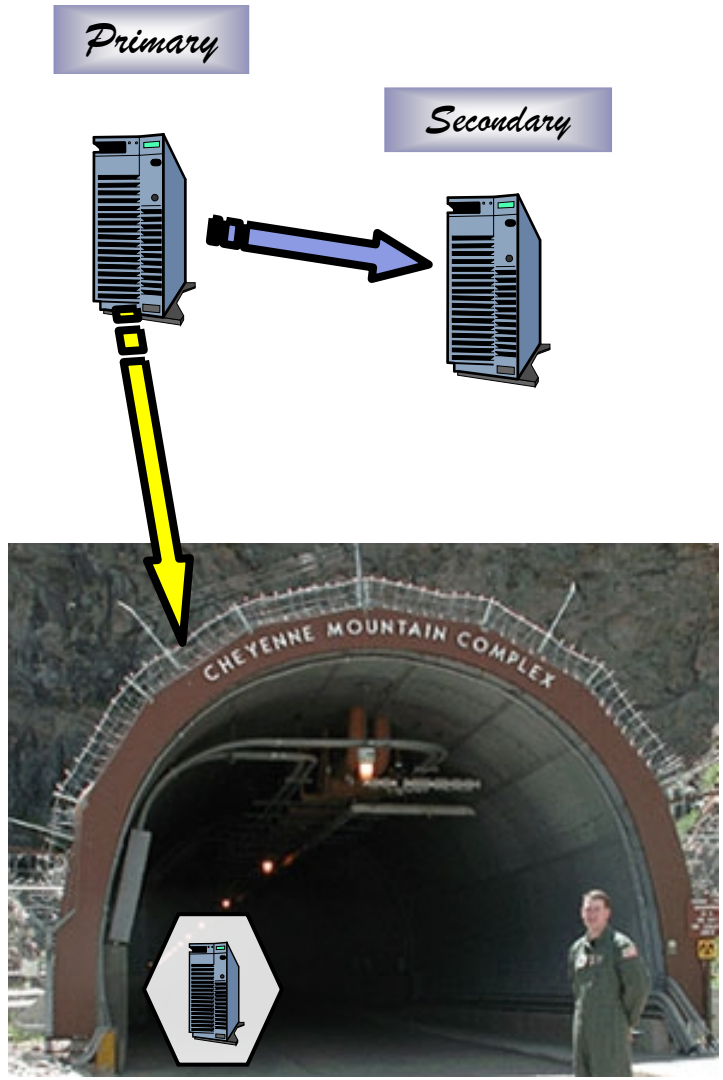
7.22 - ER Initial Release

6.00 - HDR Initial Release

Problems with pre-IDS11 Availability

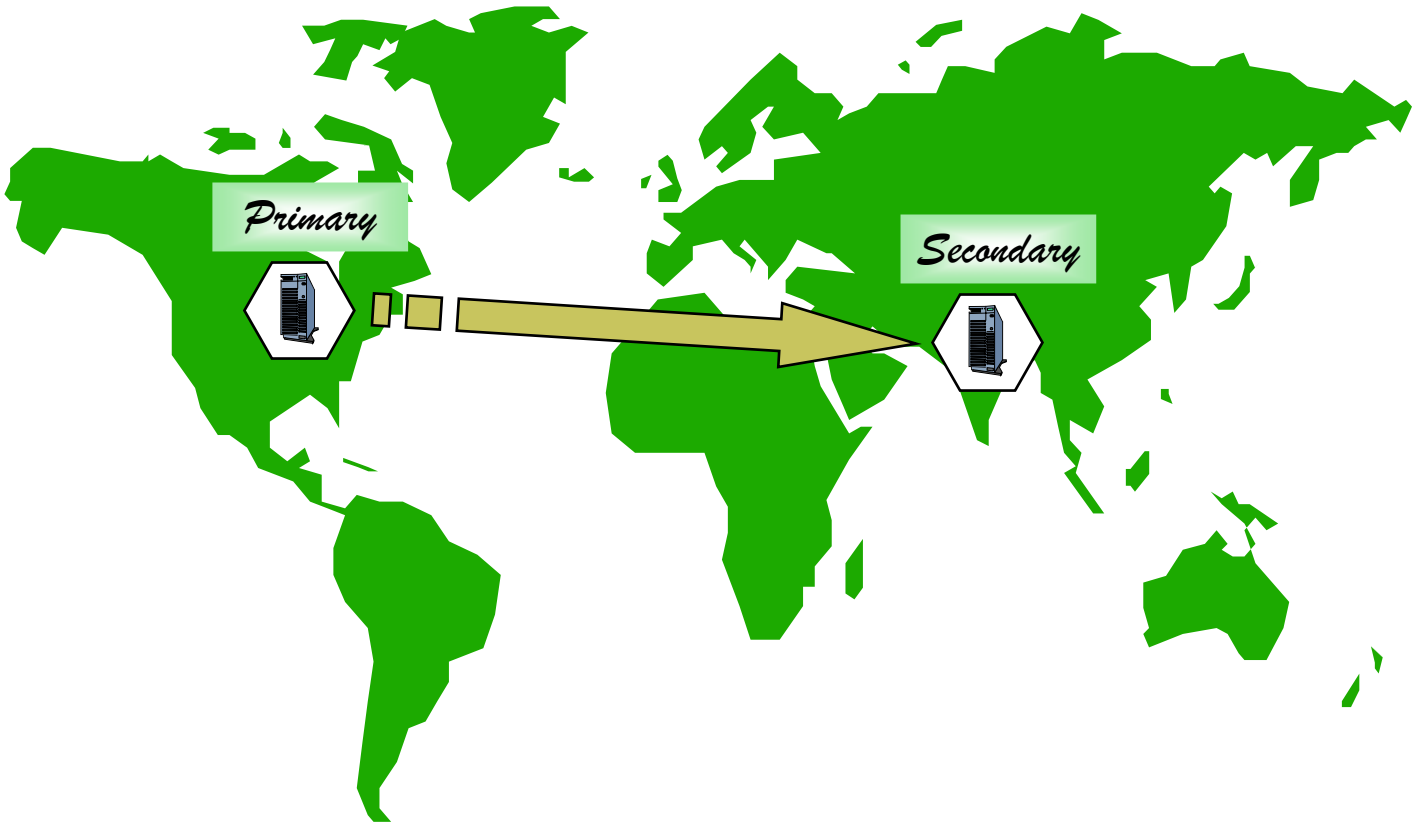
Problem - Need Additional Availability

- **Need a bunker site in addition to a hot backup**
- **Need a backup to my secondary**
- **Need additional nodes for reporting purposes**
- **Need to be able to utilize additional sites for a disaster**
- **Would like to have primary/secondary readily available, but need other backup sites.**



Problem – Want to use HDR on a WAN

- **Want to place my Primary and Secondary in different parts of the world. However it seems like the secondary is unable to keep up with the primary.**



Problem – Lots of Data

- **HDR requires a complete copy of the database**
- **Clones require a complete copy of the database**
- **I have 8 TB of data!!!**
- **I'm getting more and more data all the time.**
- **"I don't want to have to set up so much data storage to use HDR!"**

In September 1956 IBM launched the 305 RAMAC, the first computer with a hard disk drive (HDD). The HDD weighed over a ton **and stored 5MB of data**.

Today the IBM TotalStorage DS8000 series is designed to …

- Scale up to **192TB** of physical capacity and support storage sharing and consolidation for a wide variety of operating systems and mixed-server environments

What's New With IDS11 (Cheetah)

- **Continuous Log Restore**
- **Continuous Availability (MACH11)**
  - Remote Standalone Secondary (RSS)
  - Shared Disk Secondary (SDS)
  - HDR
- **Supporting Features**
  - Server Multiplexer (SMX)
  - Index Page Logging
- **Works with ER**

Continuous Log Restore

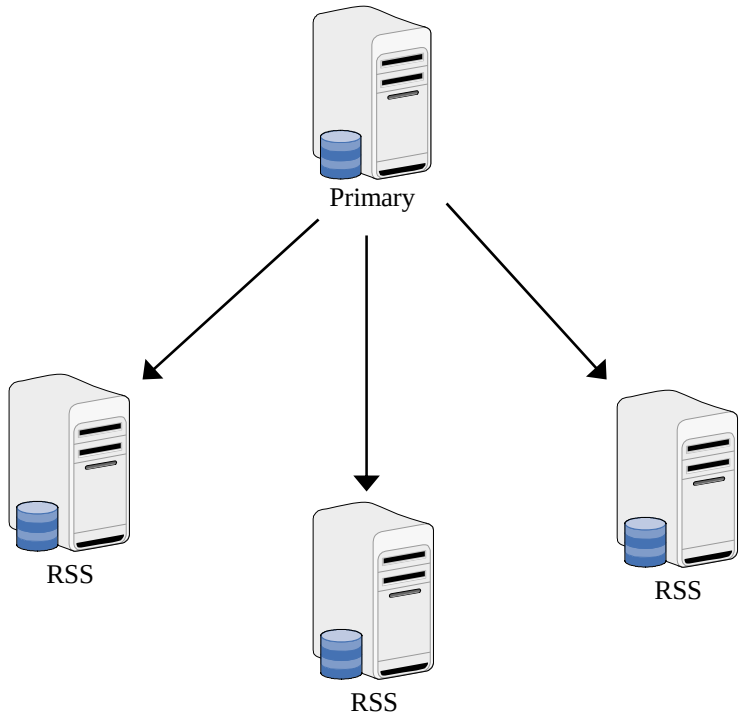
- **Allows logical recovery to span multiple ‘ontape/onbar’ commands**
- **Provides a secondary instance with ‘log file granularity’**
- **Does not impact primary server**
- **Can co-exist with MACH11 (HDR/RSS/SDS) as well as ER**
- **Can be automated by scripting the log backup alarms**
- **Useful when backup site is totally isolated (i.e. no network)**

Using Continuous Log Restore

- **Perform a full backup of Source and restore on backup instance**
- **As logs are backed up on the source, they are applied on the backup by running –**
  - ontape –l –C
  - onbar –r –l –C
- **When ready to terminate recovery mode run**
  - ontape –l
  - onbar –r -l

Remote Standalone Secondary

- **Similar to HDR**
- **Maintains a full disk copy of the database**
- **Created by performing a backup / restore of the instance**
- **Can be used for**
  - Additional Backup
  - Report processing
  - Load balancing



Remote Standalone Secondary

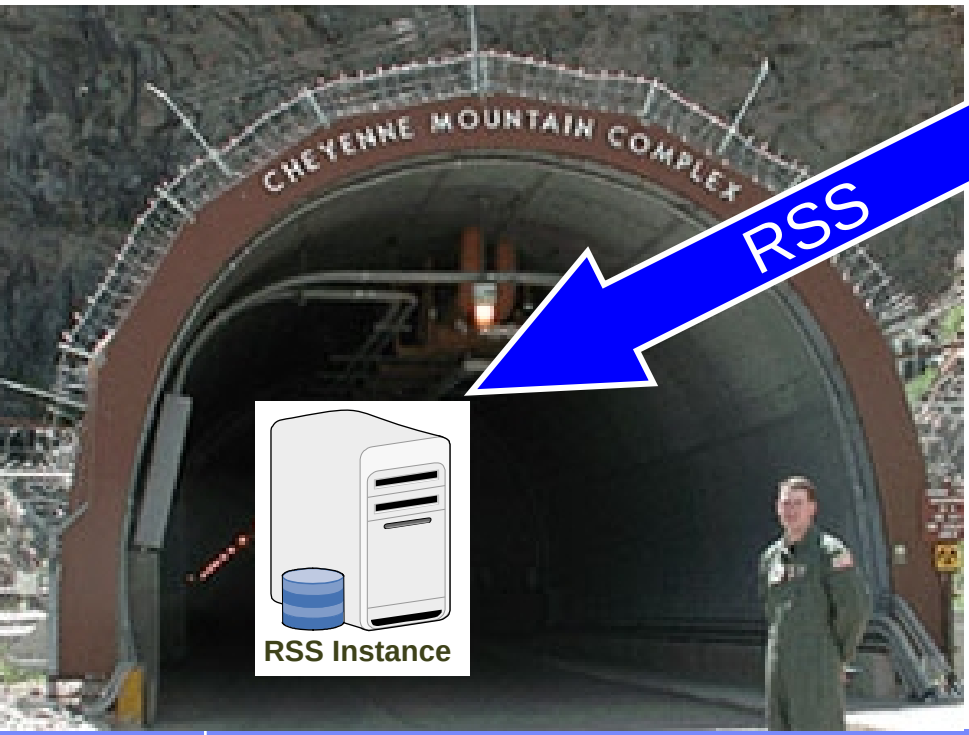
- **Distinct from HDR**
  - Uses full duplex communication (SMX)– better throughput over slower lines
  - Does not support SYNC mode, not even for checkpoints
  - Can not currently be ‘promoted’ to primary – but can be promoted to HDR secondary
    - Focus is on Disaster Recovery, not HA
  - There can be any number of RSS instances
  - Requires Index Page Logging be turned on.

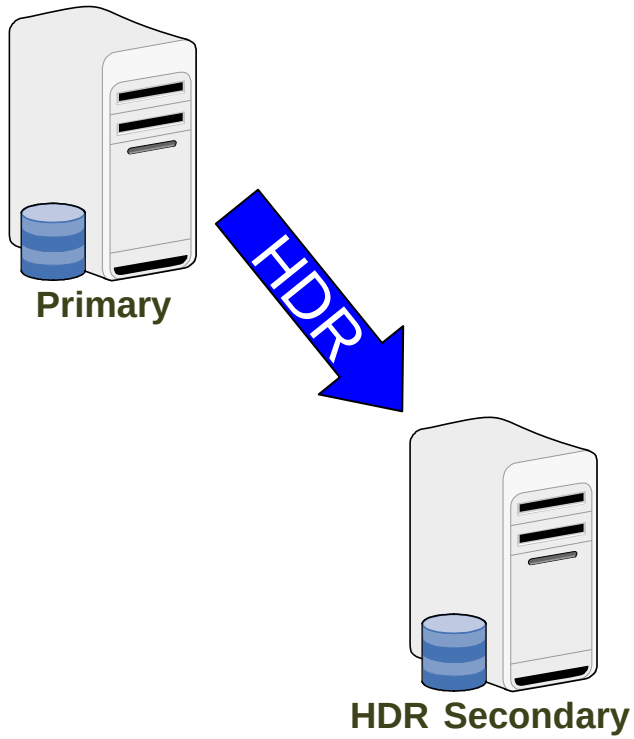
- **RSS can be used in combination with HDR secondary**
  - RSS can be converted into HDR secondary
  - HDR secondary can be converted into RSS

Usage of RSS – Bunker Backup

Customer is currently using HDR for high availability but would like to have an additional backup of their system in the event of a disaster in which both the primary and the secondary should be lost.

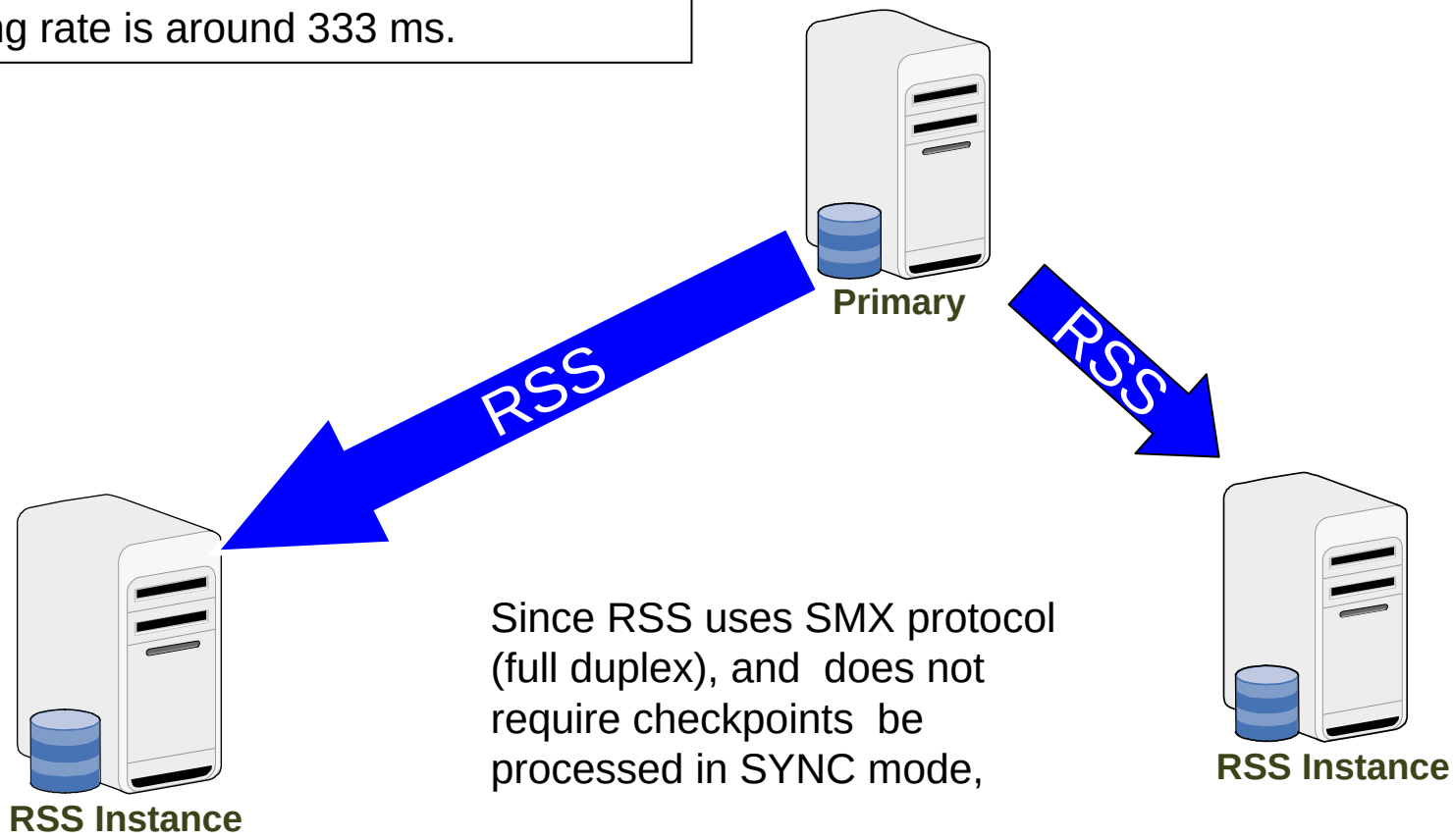
Using HDR to provide High Availability is a proven choice. Additional disaster availability is provided by using RSS to replicate to a secure 'bunker'.





Availability with Poor Network Latency

Customer is replicating instance in remote locations, but was tested and knows that the ping rate is around 333 ms.

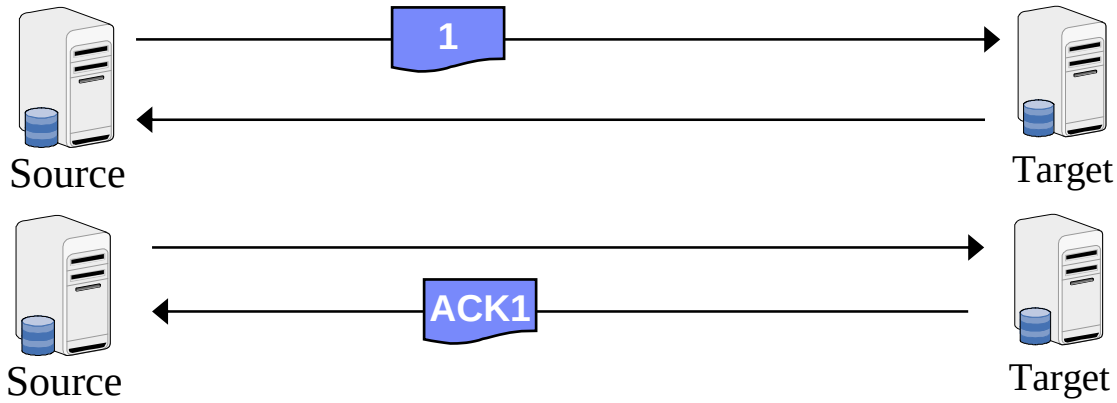


Since RSS uses SMX protocol (full duplex), and does not require checkpoints be processed in SYNC mode,

Fully Duplexed???

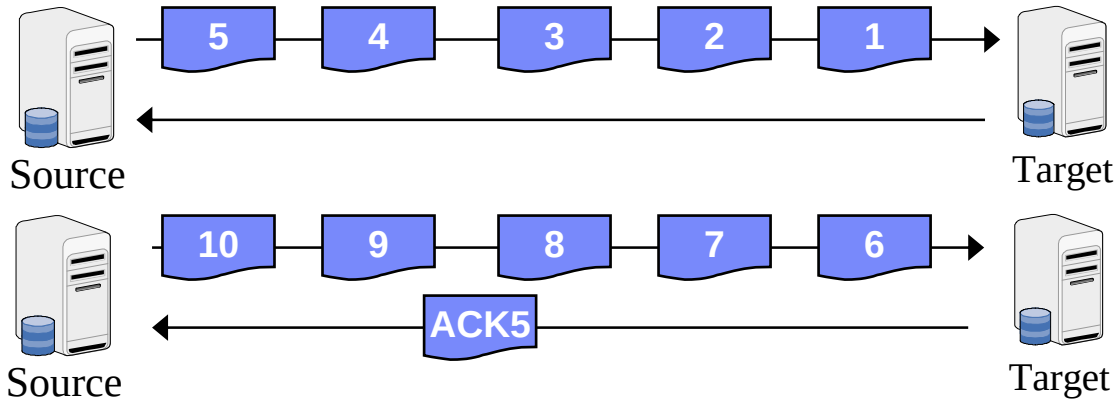
Half Duplexed

HDR Requires a half duplexed connection

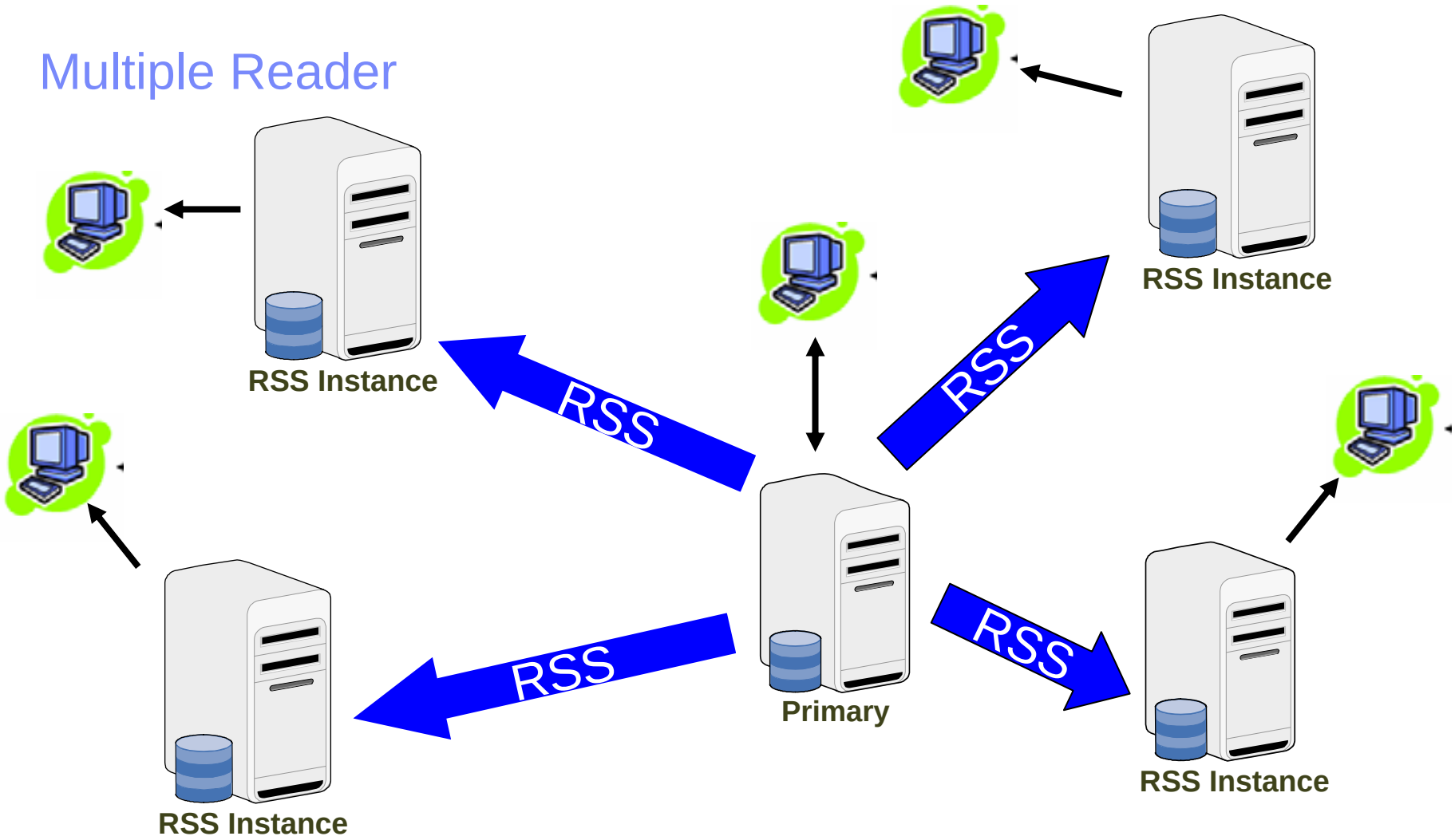


Fully Duplexed

RSS utilizes a fully duplexed connection



Multiple Reader



Setup of RSS

- **Similar setup process as with HDR**

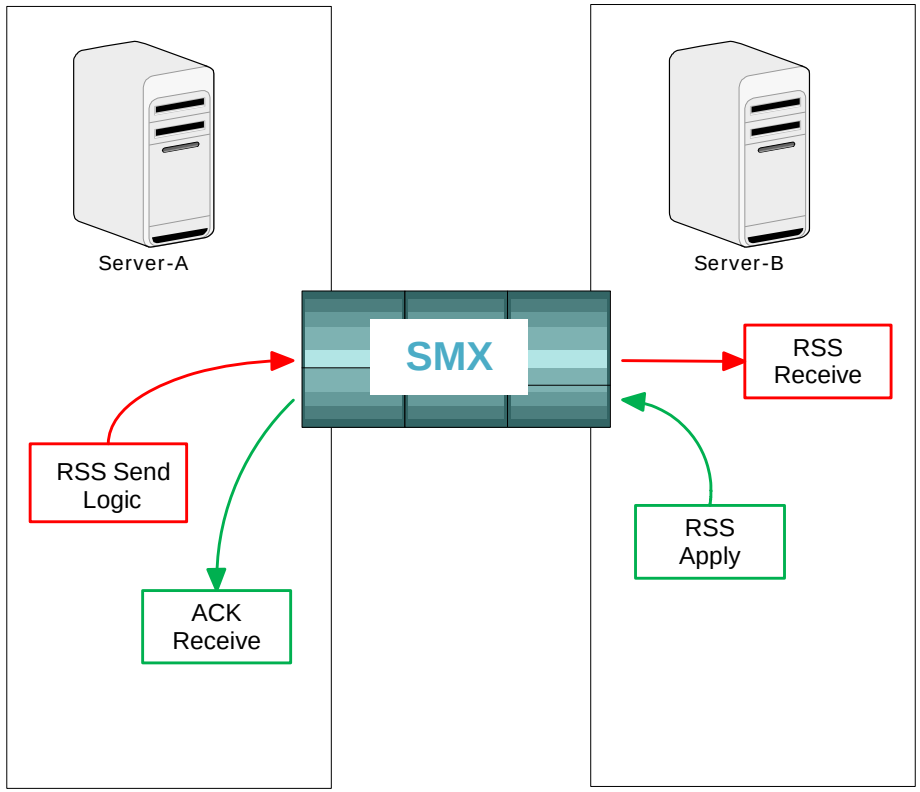
Action	Primary	RSS Secondary
Set Index Page Logging on Primary	LOG_INDEX_BUILDS 1	
Register the new secondary on primary	onmode –d add RSS <secondary-node>	
Take full backup of primary	ontape –s –L 0	
Physical restore on secondary		ontape –p
Connect to primary		Onmode –d RSS <primary-node>

Index Page Logging

- **Eliminates the create index transfer as done by HDR**
  - Caused index pages to be copied to the logical log when initially creating the index
  - REQUIRED for RSS
  - The logging of the index can be split into multiple transactions and is not part of the original user transaction
  - Control not returned to user until logging is performed
  - Can be use by HDR as well as with RSS
  - Activated by the LOG_INDEX_BUILDS parameter in onconfig

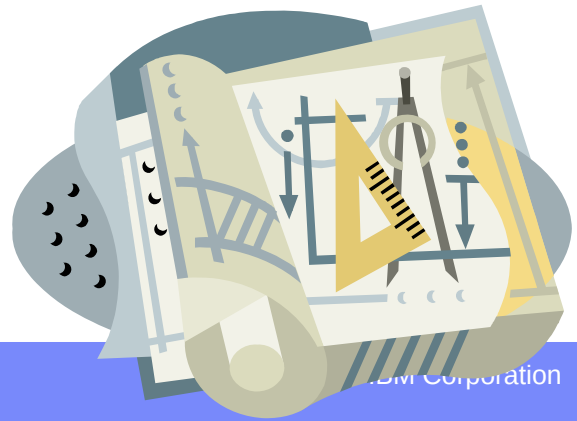
Server Multiplexer (SMX)

- **Creates a multiplexed network connection between two servers**
- **Multiple internal routines can be communicating over a single SMX connection**
- **Supports encryption**
- **Simplifies inter-instance communication**
- **Automatically activated**
- **Requires no configuration other than encryption**
- **Can support either Half or Fully duplexed protocols**



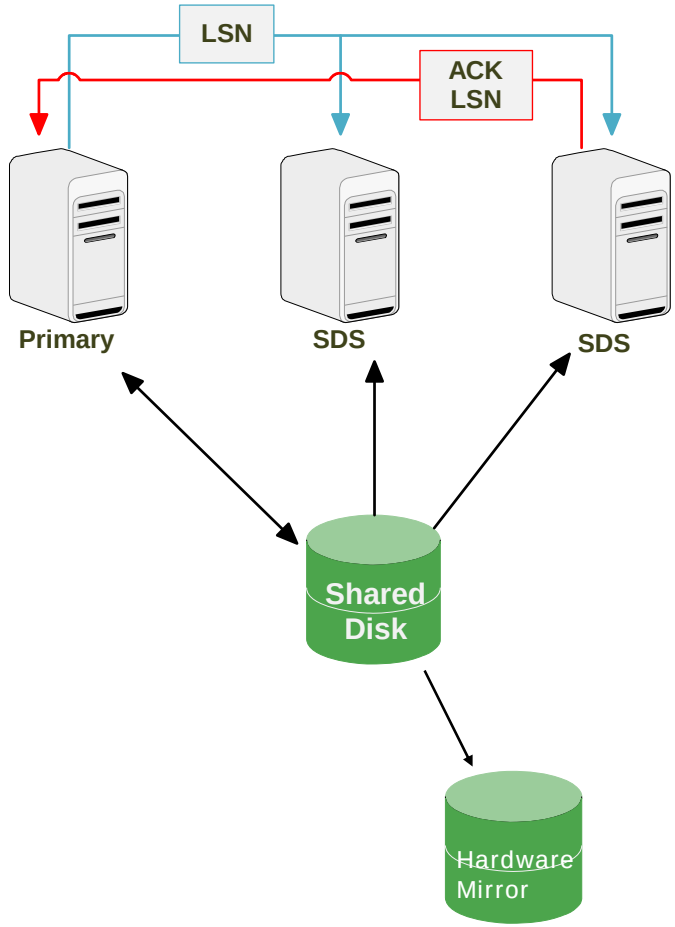
RSS: Failover

- **Rules**
  - The RSS instance can not currently be swapped with the primary
  - DRAUTO does not work with RSS
  - The RSS node can be converted into an HDR secondary
  - The HDR secondary can be converted into an RSS node
  - The RSS instance can be converted into a standard instance



Shared Disk Secondary

- **HDR on top of a shared disk subsystem**
- **Works nicely with blade servers**
- **Minimal Setup Time – only a checkpoint is required to start the SDS**
- **Primary role can shift to any of the SDS nodes**
- **Provides additional read capacity without requiring additional disk**
- **Works by coordination of page flushing to disk**
- **Network exchanges log LSN, not log pages**



Setting up Shared Disk Secondary

- **Map the shared disk identically on each of the nodes**
- **Mark the primary as allowing shared disk secondary nodes to connect**
  - onmode –d set SDS primary <listener port>
- **Configure the SDS node (next slide)**
- **Bring up the secondary (oninit)**
  - This will cause a hard checkpoint to be performed on the primary
  - The SDS nodes simply start from that checkpoint – no recovery

SDS Testing Environment

Server	OS	Shared Disk
HPIA64	HP-UX B.11.23	Veritas Clustering and Veritas FileSystem
SOL64	SunOS 5.10 (sparc)	Veritas Clustering and Veritas Filesystem
Linux64	Linux 2.6.9-34.EL	GPFS (IBM)
AIX64	AIX 5.3	GPFS (IBM)

Configuration of the SDS node

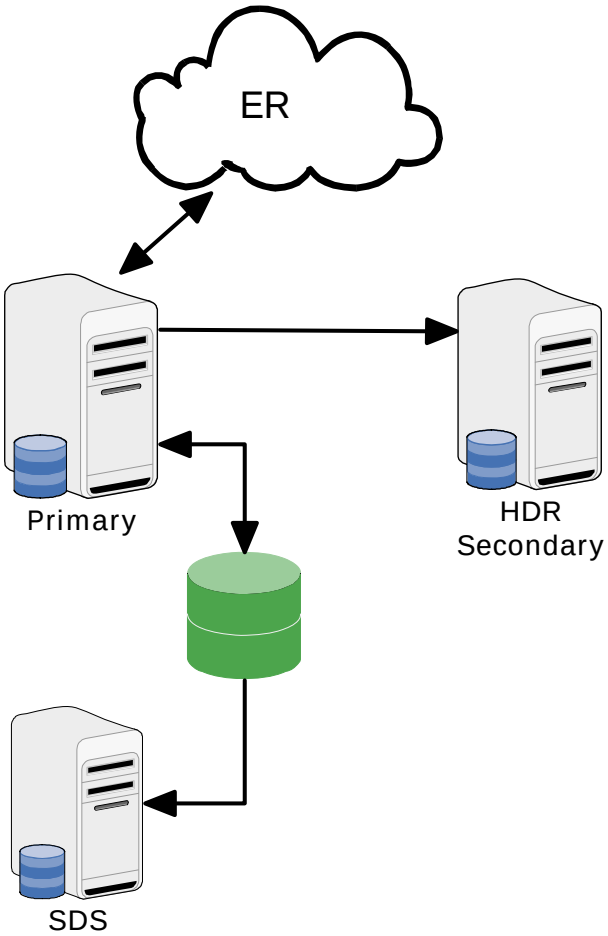
- **Enable SDS**
  - onconfig ENABLE_SDS 1
- **Setup two ‘paging files’**
  - Used for temporary storage of pages in between checkpoints on the SDS node
  - onconfig SDS_PAGING <path1>,<path2>
- **Define a local temporary dbspace**
  - onconfig SDS_TEMPDBS
  - onspaces not used to create

Preferred Failover Order

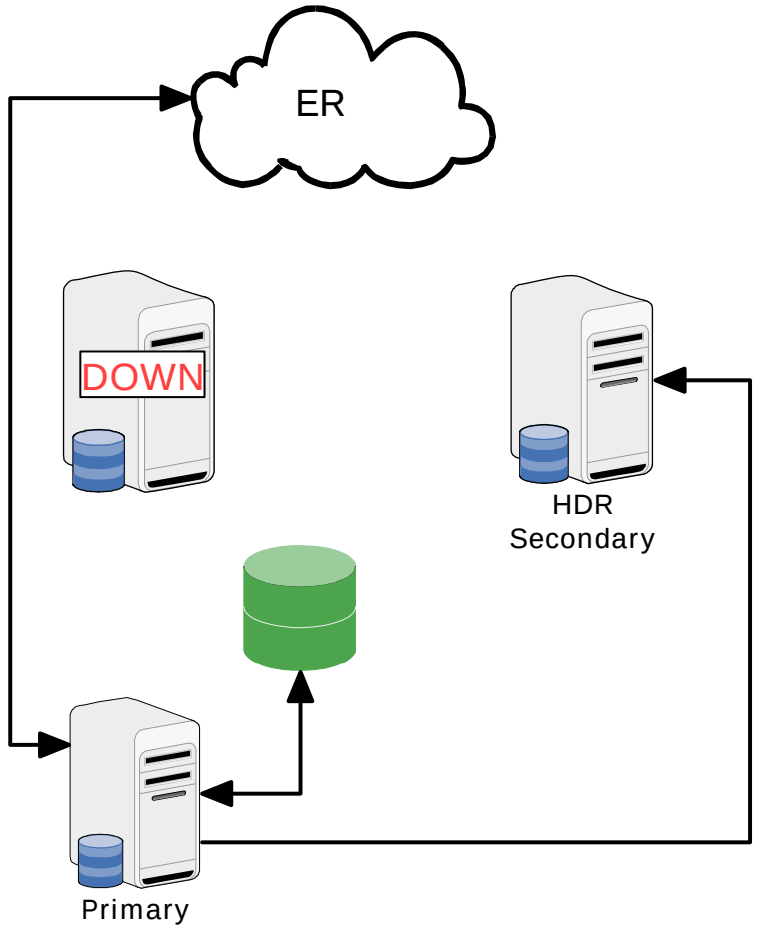
- **First fail to another SDS node**
  - When you failover to the HDR secondary, or RSS node, then all active SDS nodes must be shut down.
- **Then fail to the HDR secondary if it exists**
- **Finally fail to an RSS node**
- **After failover, all MACH11 and ER nodes realign to the new master**

N.B. – If all servers are down in a SDS cluster and the current primary can not be brought back online, then you can start one of the SDS nodes with oninit –SDS=<local_alias>.

Automatic Realignment



After Running onmode –d make primary

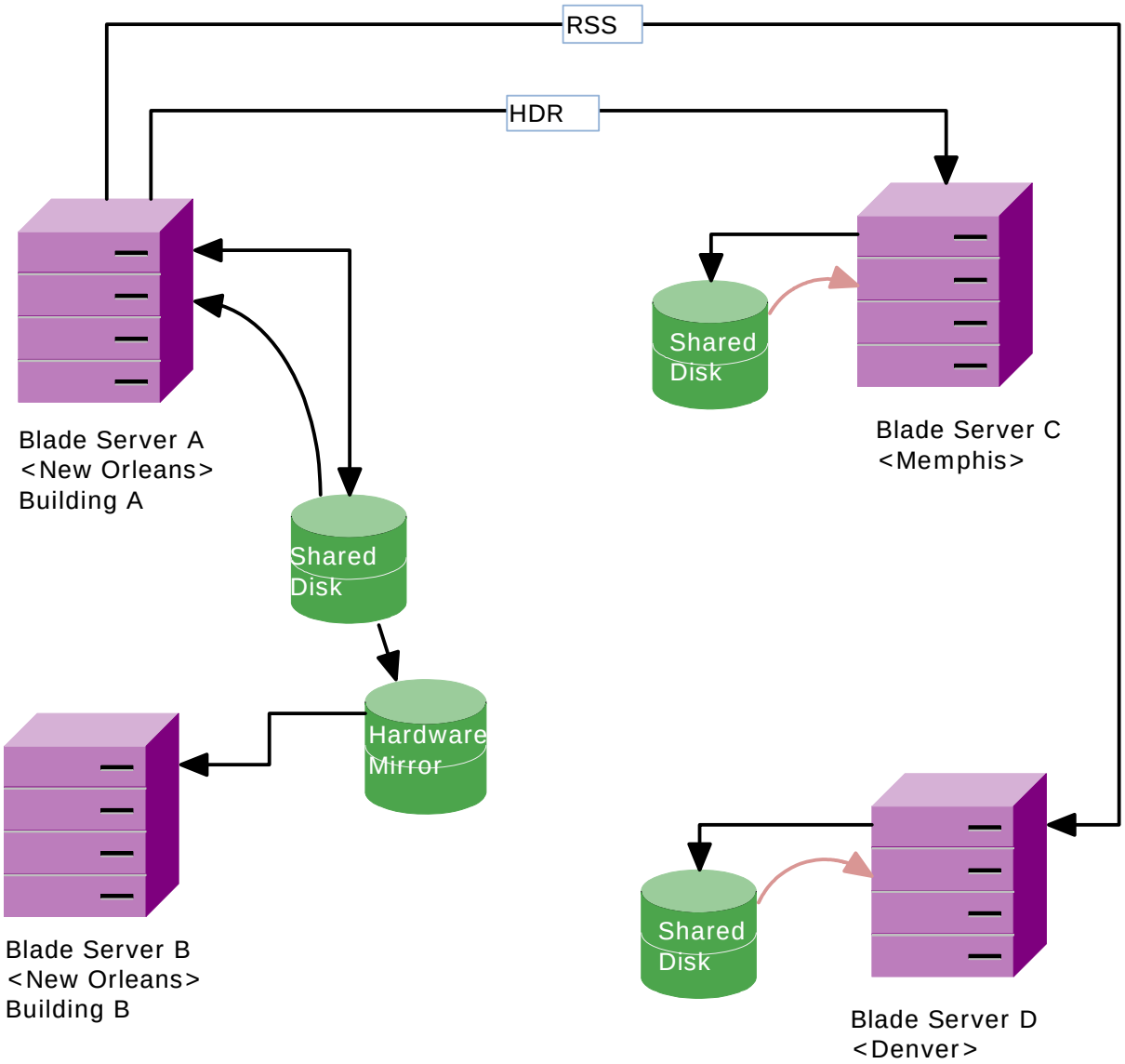


New onmode commands

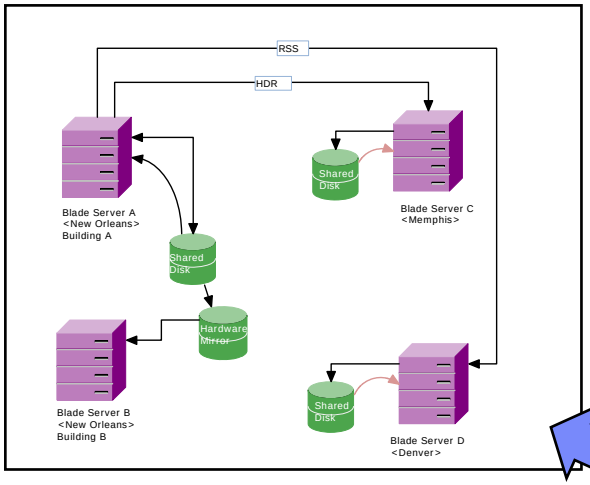
- **onmode –d add/change/delete RSS <RSS_node>**
  - Add, changes, or deletes the RSS node which will attach to the primary
- **onmode –d RSS <primary_node>**
  - Used to connect the newly defined RSS node to the primary
  - Also used to demote an HDR secondary to an RSS node
- **onmode set/clear SDS primary <local alias> [force]**
  - Used to identify the local alias
- **onmode make primary <local alias> [force]**
  - Used to convert the local node to the primary node, *regardless of the secondary type*

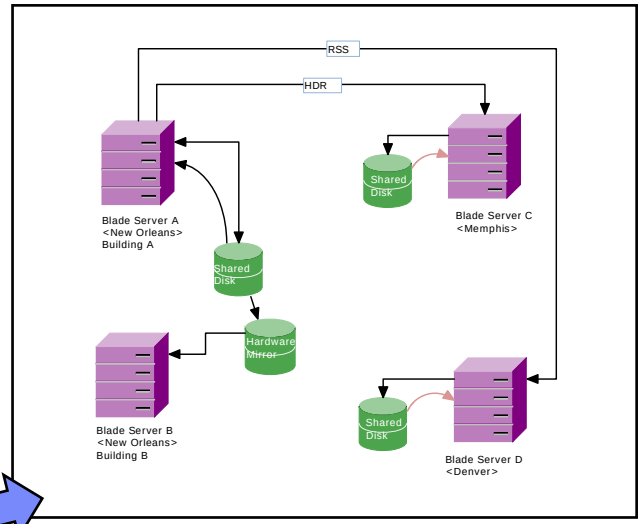
N.B. – The force option is used to convert the local node when the current primary is offline.

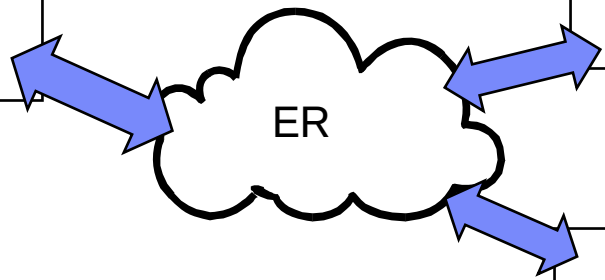
Multi-Site Failover

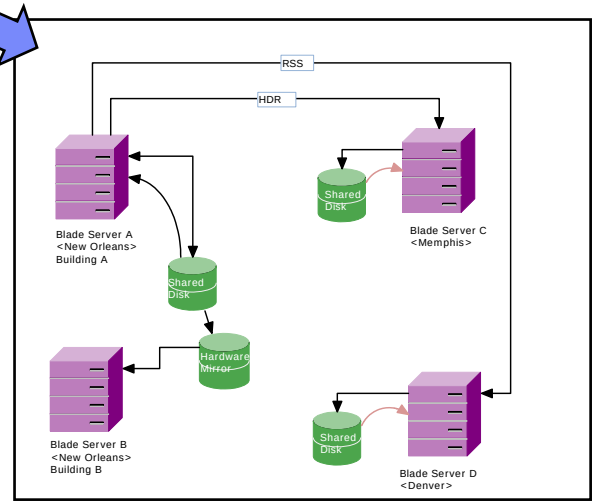


Don’t forget about ER









Any Node within the ER domain can also be a MACH11 cluster.

Useful Links

- **White paper describing availability solutions in IDS11**
  - ftp://ftp.software.ibm.com/software/data/informix/ids11-availability-wp.pdf
- **Informix Product Page**
  - http://www.ibm.com/informix/
- **Informix Platform Roadmap**
  - http://www-306.ibm.com/software/data/informix/pubs/roadmaps.htm
- **Informix Product Platform Availability (Select Download Spreadsheet on Right)**
  - http://www-306.ibm.com/software/data/informix/pubs/roadmaps.html

Questions?

Madison Pruet (mpruet@us.ibm.com)

IDS Replication Architect

Information Management (IDS)