Building enterprise-class and virtually fail-safe database services on the Mac

Session: IT864

Carlton Doe
cdoe@us.ibm.com

Agenda

- Definition of terms
- A deeper look and hands-on demonstrations
  - Data replication -- business continuity
  - Data distribution -- full or partial copies
- Review and Summary

Download Session Presentations

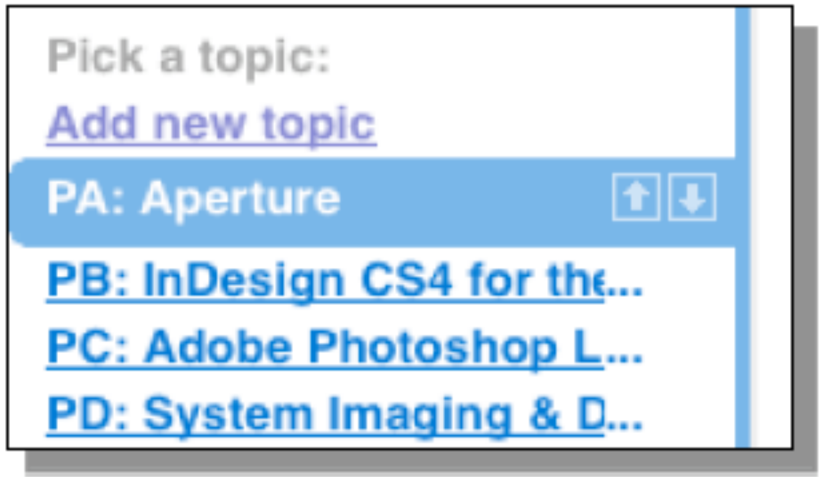
http://macpres09.shownets.net

- All registered conference attendees can go to to access the presentations for sessions they want to download. Each conference program will have a folder, with the corresponding presentations included that speakers have posted.
  - Please refer to the sign outside the conference room if you need information about the Conference Name & Session Number.

Q&A –MacIT® Conference

We are using Google Moderator to take questions for this session.

- Go to http://tinyurl.com/633v6e
- Pick the topic# IT864
- Sign in using a Google Account:
    - User Name: macworldexpo09
    - Password: macworld09
- Submit the questions you want to ask
- Vote on others' questions you want answered



Pick a topic:

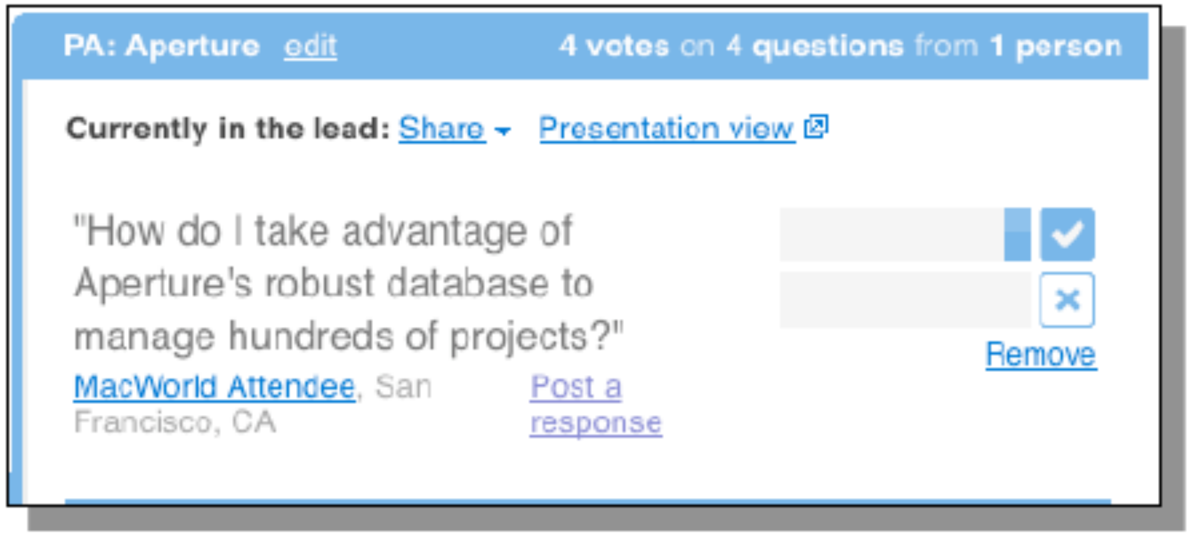
Add new topic

PA: Aperture

PB: InDesign CS4 for the...

PC: Adobe Photoshop L...

PD: System Imaging & D...



PA: Aperture edit 4 votes on 4 questions from 1 person

Currently in the lead: Share ▾ Presentation view

"How do I take advantage of Aperture's robust database to manage hundreds of projects?"

MacWorld Attendee, San Francisco, CA

Post a response

✓

✕

Remove

So what are we talking about??

So what are we talking about?

- High availability computing aka HA
  - Online capability or availability of components in a computer system including
    - Hardware (CPU, memory, disk, backplane etc)
    - Software (O/S, utilities, applications)
- Is NOT a fault tolerant system (FT)
- Is NOT a continuous availability system (CA)

So what are we talking about?

- A Continuous Availability (CA) system
    - A system that is rarely ever unavailable due to planned outages or other reasons
- Is built using high availability (HA) components
    - Specifically designed to protect against single-point-of-failure (SPoF) conditions
- Is NOT a fault tolerant (FT) system

So what are we talking about?

- A Fault Tolerant system (FT) uses
  - Hot standby as well as near-line standby hardware duplicates
  - Software (to a degree)

  to survive a component or service failure

- Result is a system that is highly available and should be continuously available

So what are we talking about?

- A Disaster Recovery system (DR) uses
  - Warm standby hardware duplicates
  - Software

  to survive a component or service failure

- Focus is preserving the ability to do business after some sort of a massive outage
  - Copies are usually NOT maintained in real-time

- Is NOT CA, HA or FT oriented

So what are we talking about?

- The creation of HA, CA, DR or FT system is an exercise in risk vs reward
  - Focused on identifying and protecting against single-points-of-failure (SPoF)
  - Cost is often the major driving factor
    - To protect
    - Risk of not protecting

So what are we talking about?

- Examples:
  - Disk
    - RAID
    - Disk to server connections
    - Enclosures

So what are we talking about?

- Examples (cont):
  - Servers
    - Redundant internal (idle) components
    - Redundant copies

So what are we talking about?

- Examples (cont):
  - Power
    - UPS
    - Redundant internal circuits
    - Generator
      - fuel??
    - Redundant external grid feeds

How does this apply to data services?

- No different than the hardware systems discussed so far
    - Focus in this case is in the database engine aka “data server”
    - Use its HA *and non-HA* technology to build a FT system than can become CA

How does this apply to data services?

- **The key is the data server and the features, functionality, flexibility and ease of use it provides to build the HA/FT/CA system you want!**
  - Does it provide:
    - Data distribution?
    - HA data replication?
      - Are they mutually exclusive?
      - Does it scale to handle total data size, transaction volume, transaction size?
      - How easy is it to design, instantiate, monitor and repair?
  - A “real” Enterprise-class system should be able to provide this functionality rather than requiring third-party components

A deeper look and hands-on demonstrations

In two parts

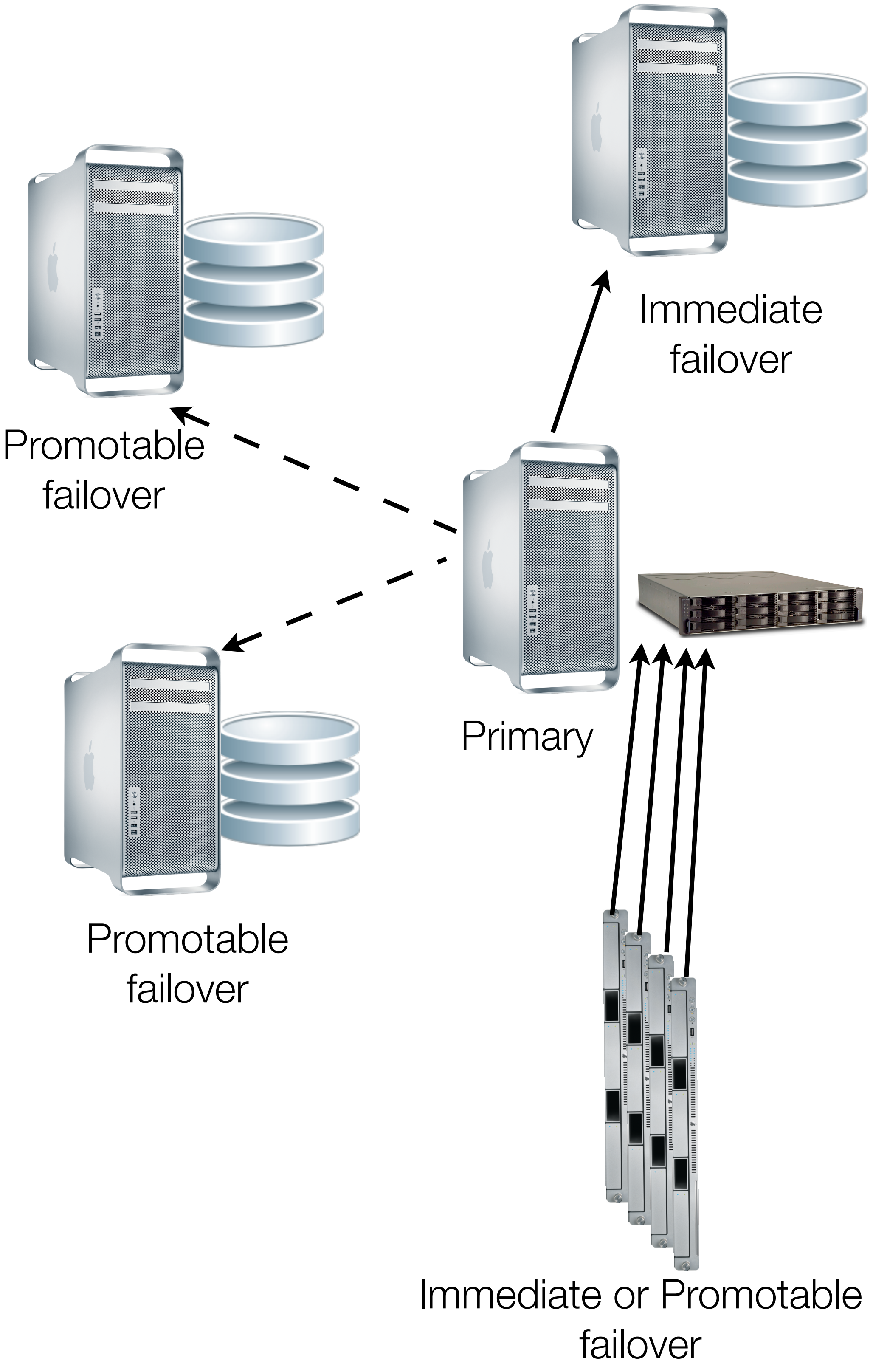
Data Replication --
Continuous availability and business continuity

Data Replication

- Guiding principle: multiple copies of a whole single version of the truth
  - Maintained in real-time
  - Data server driven (no application code)
  - Is data server instance oriented
    - All databases, all objects
  - One to N levels of immediate and promotable fail-over protection
    - Can also provide other business value such as load balancing but not the primary objective

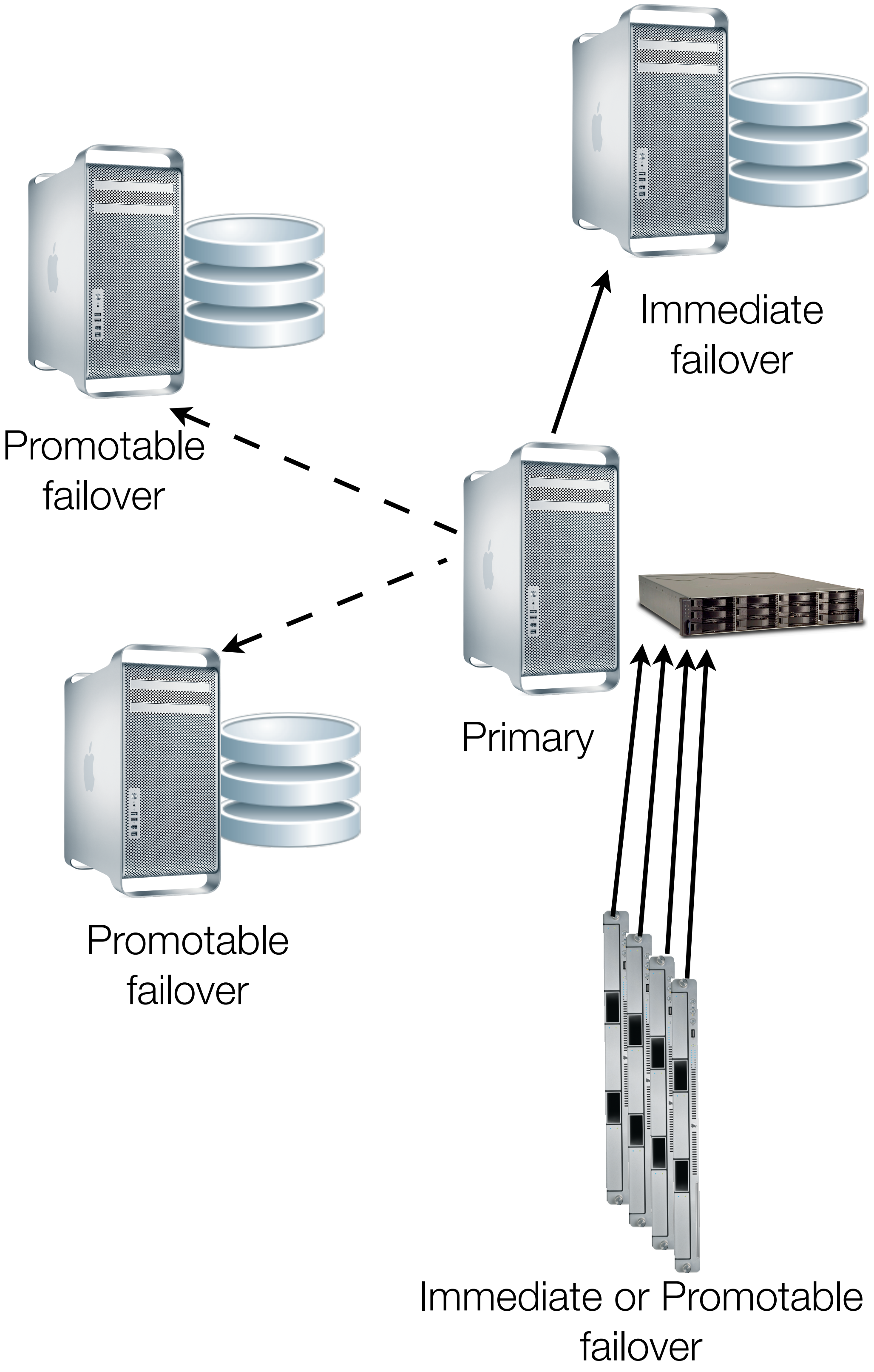
Data Replication

- Focuses on data server SPoF conditions to provide “always-on” capability
- One or more full copies of primary data server for transaction processing failover
    - Synchronous communication to immediate fail-over nodes
    - Asynchronous to promotable nodes with guaranteed data synch
- Flexible server types
    - Can provide protection against different SPoFs
        - Disk
        - Server
        - Network
        - Location



Data Replication

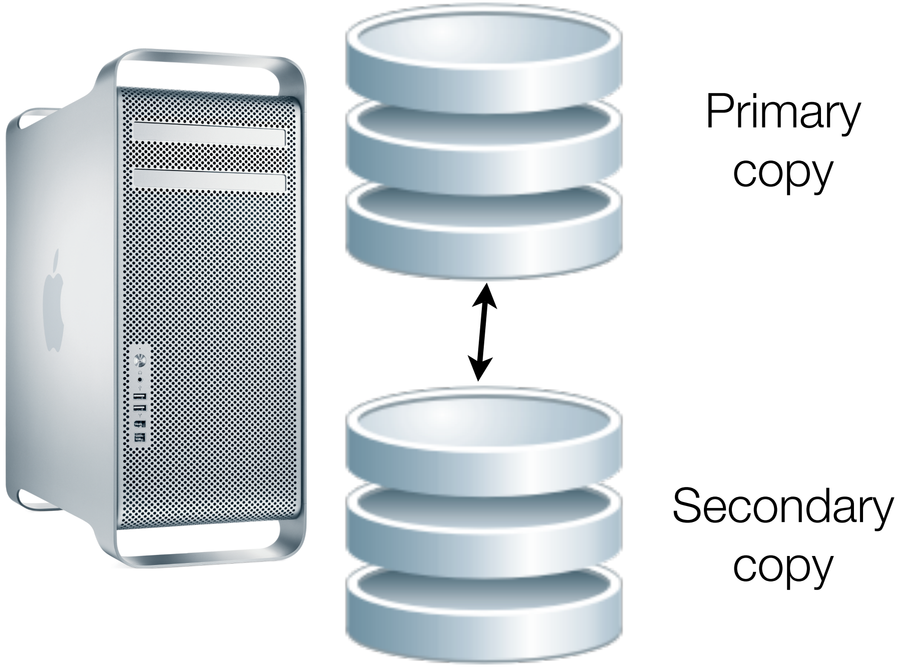
- Can support active/active updates
  - full application support anywhere in the cluster
    - no code changes required to deploy
- Integrated, not add-on
- Compatible and interoperable with data distribution technology
- Can provide secure data communication between servers



Disk Mirroring

Data Replication

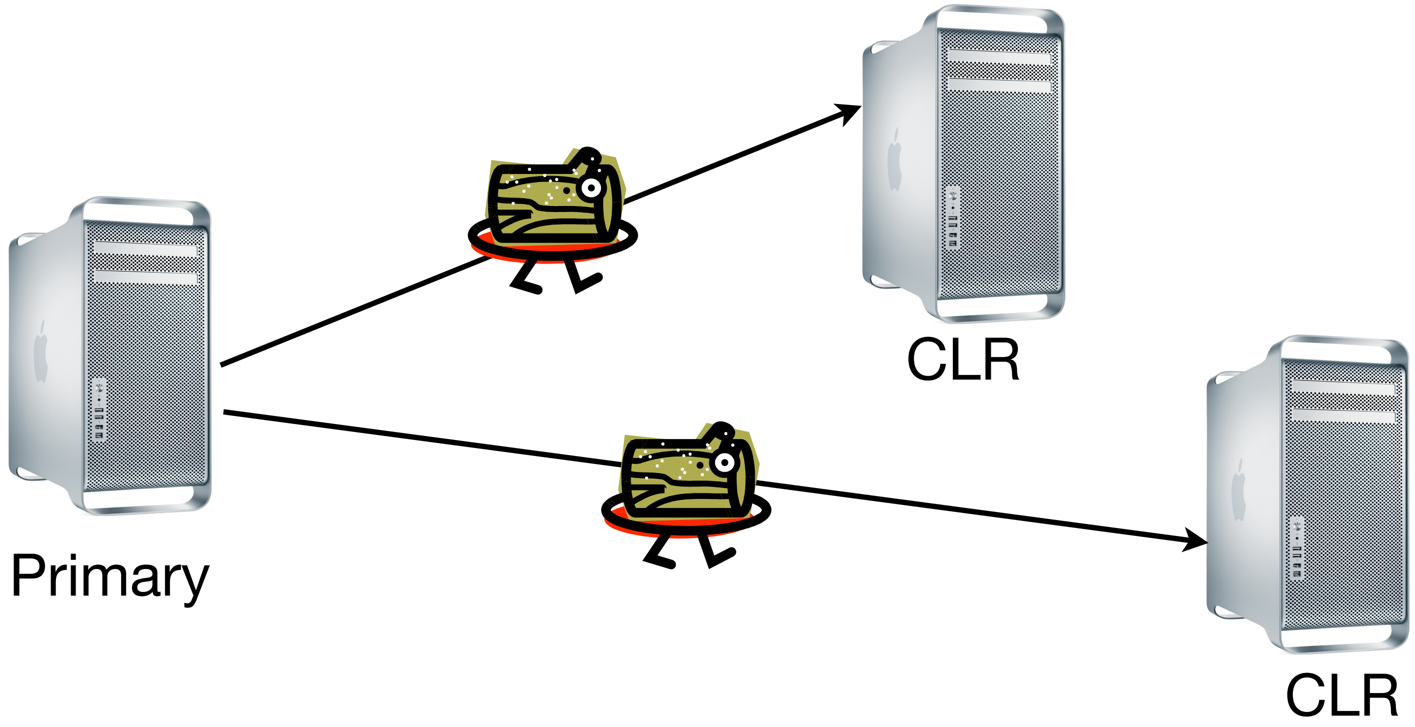
- Technology: disk mirroring
- Data server can create and maintain mirrors of storage objects
    - Protects against the most common type of data server failure
    - Transaction committed after both writes occur
    - Data server can leverage both copies for split read / write operations
    - integrated, not add-on
- Is CA oriented



Continuous Log Restore

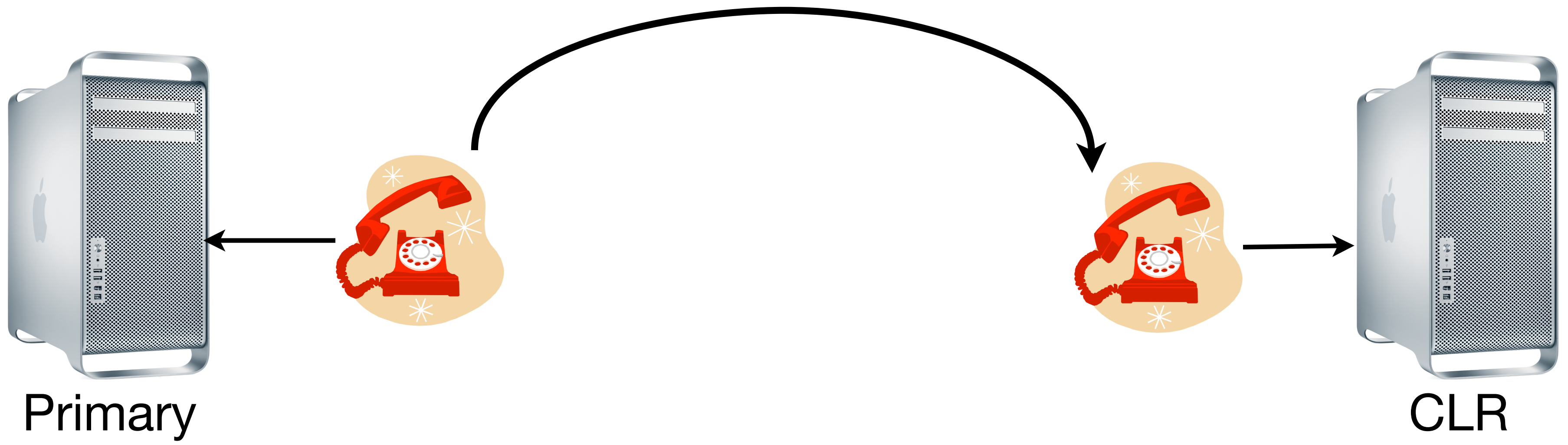
Data Replication

- Technology: Continuous Log Restore server (CLR)
- Is DR oriented
  - Primary replicates groups of transactions to one or more secondaries
  - Delayed capture and apply
- Secondary data server can be on or off-line
  - Use existing but less powerful hardware or servers primarily tasked to a different operation
  - If activated, WILL require recreating some processed data not captured from primary



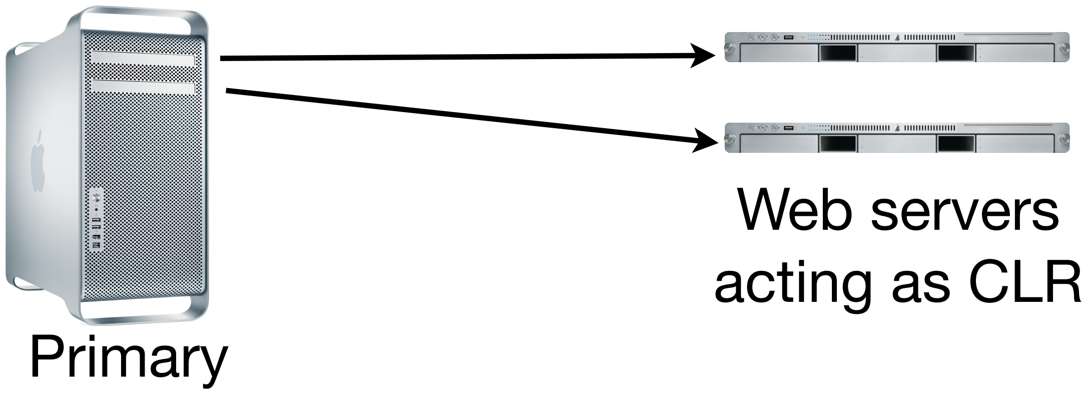
Data Replication

- Technology: Continuous Log Restore server (CLR)
- Deployment scenario: intermittent network connection
  - Server locations lack a consistently available network
  - Business still wants a stand-by copy of transactions
- Transaction record blocks transmitted via FTP or other mechanism when connectivity exists
  - Can either be applied to secondary or stored



Data Replication

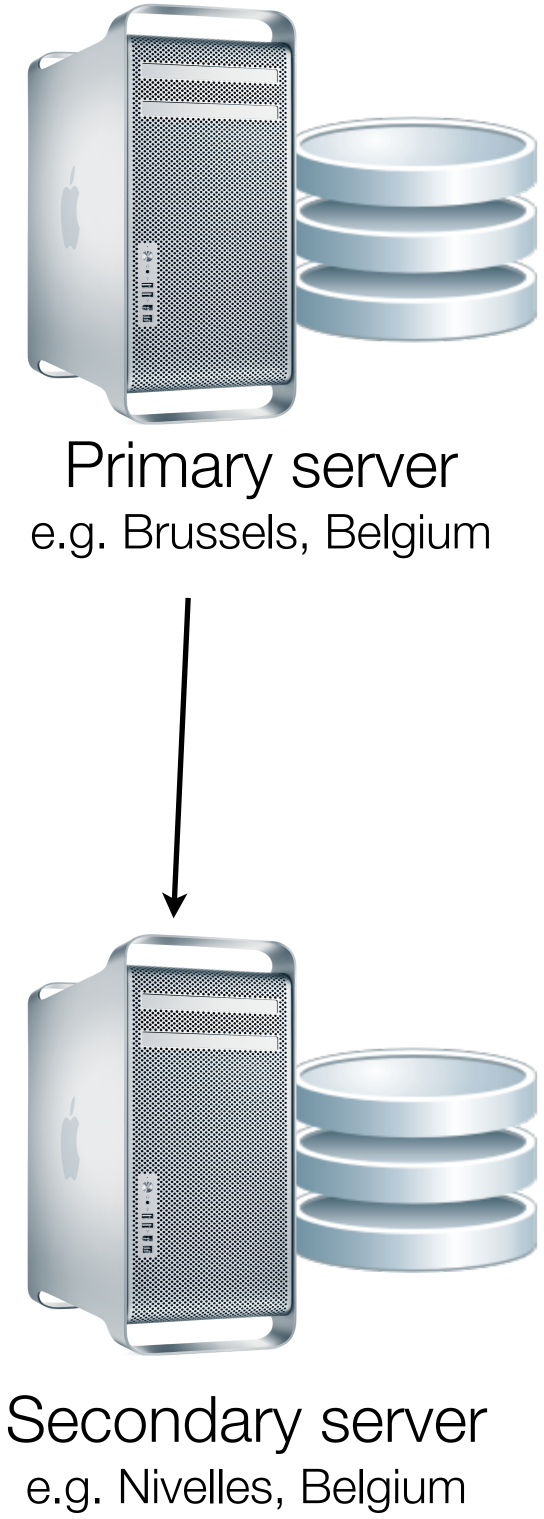
- Technology: Continuous Log Restore server (CLR)
- Deployment scenario: lack of hardware for dedicated HA/CA server
  - Business can not afford to buy separate servers / disk for secondary server(s)
  - Business still wants a stand-by copy of transactions
- Can install / configure data server on existing hardware
  - Data Server is not turned on
  - Blocks of transaction records transmitted from primary and stored on secondary
  - In a failure condition, secondary is activated, transaction blocks applied and limited data services is available
    - Better than nothing at all



HA Server

Data Replication

- Technology: HA server
- Is HA oriented
  - Primary replicated to a single secondary
  - Secondary can be read-only mode OR read/write
- Transaction not committed until change is on the secondary as well
- Secondary could automatically assume primary responsibilities if a failure occurred
- Requires uninterrupted, low latency network connection

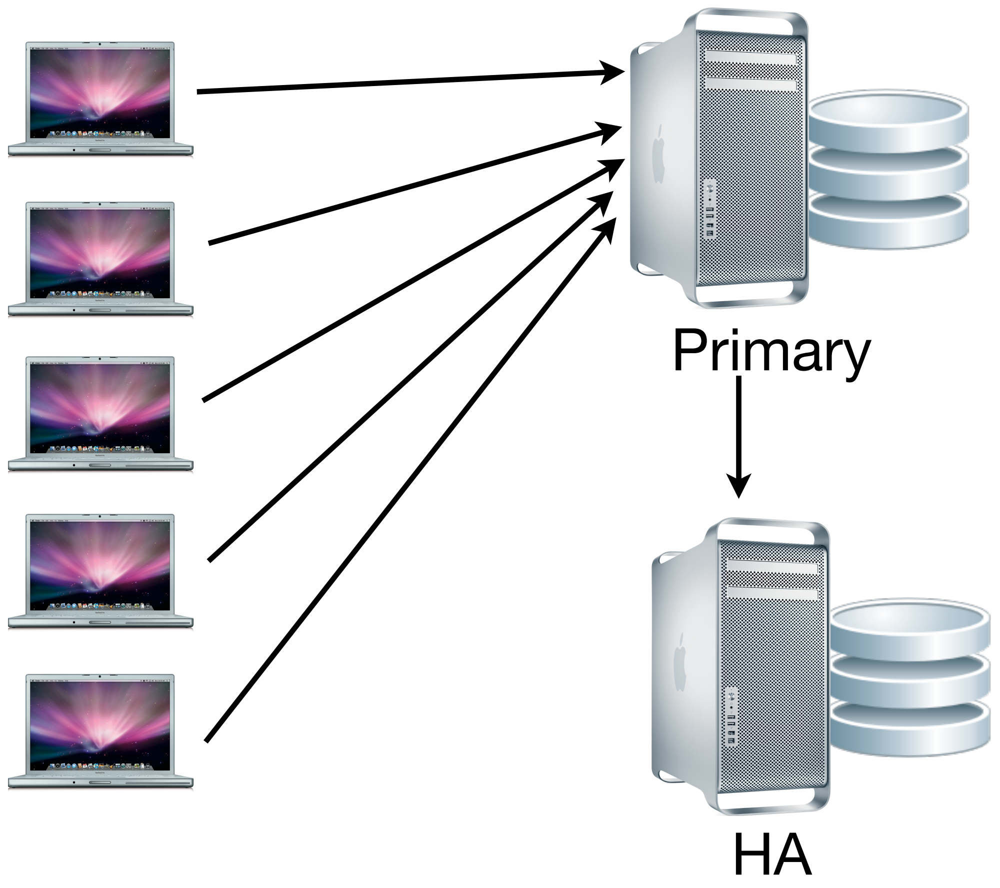


Primary server
e.g. Brussels, Belgium

Secondary server
e.g. Nivelles, Belgium

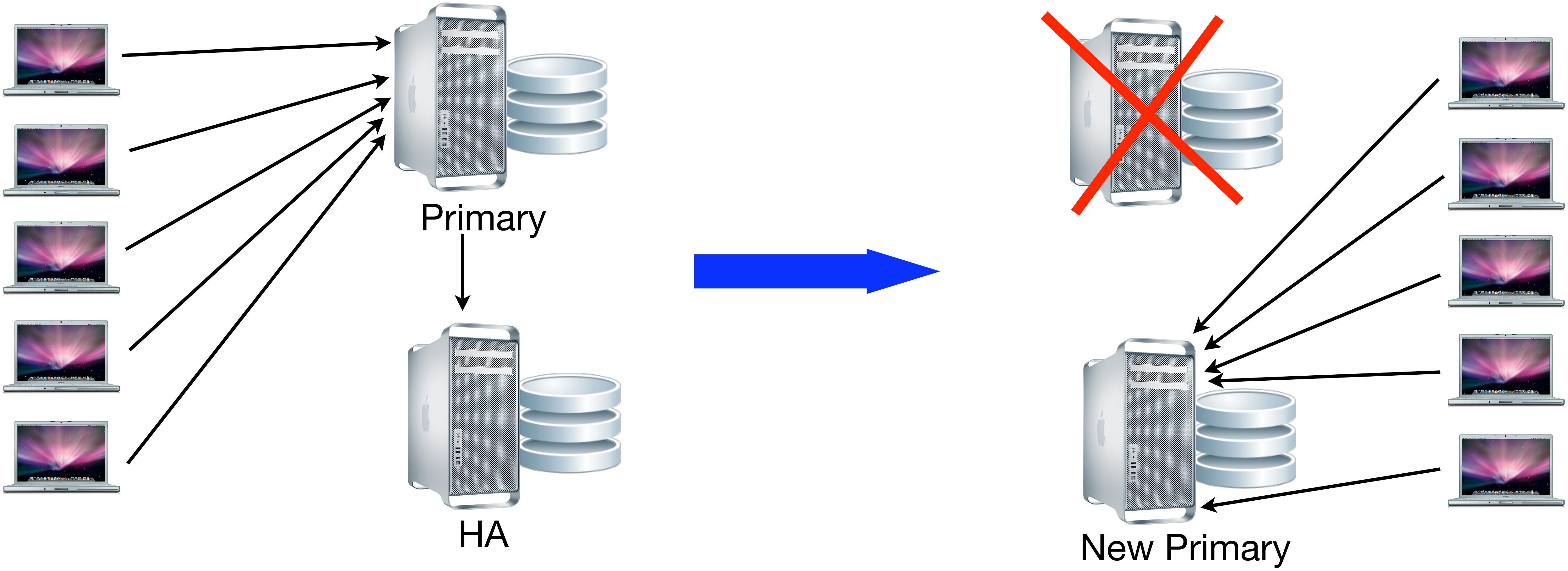
Data Replication

- Technology: HA server
- Deployment scenario: protection against disk or server failure
  - Secondary must be relatively close to primary
- Secondary use -- separation of applications
  - separate read intensive from write-oriented



Data Replication

- Technology: HA server
- Deployment scenario: protection against disk or server failure
  - Secondary must be relatively close to primary
- Secondary use -- separation of applications
  - separate read intensive from write-oriented



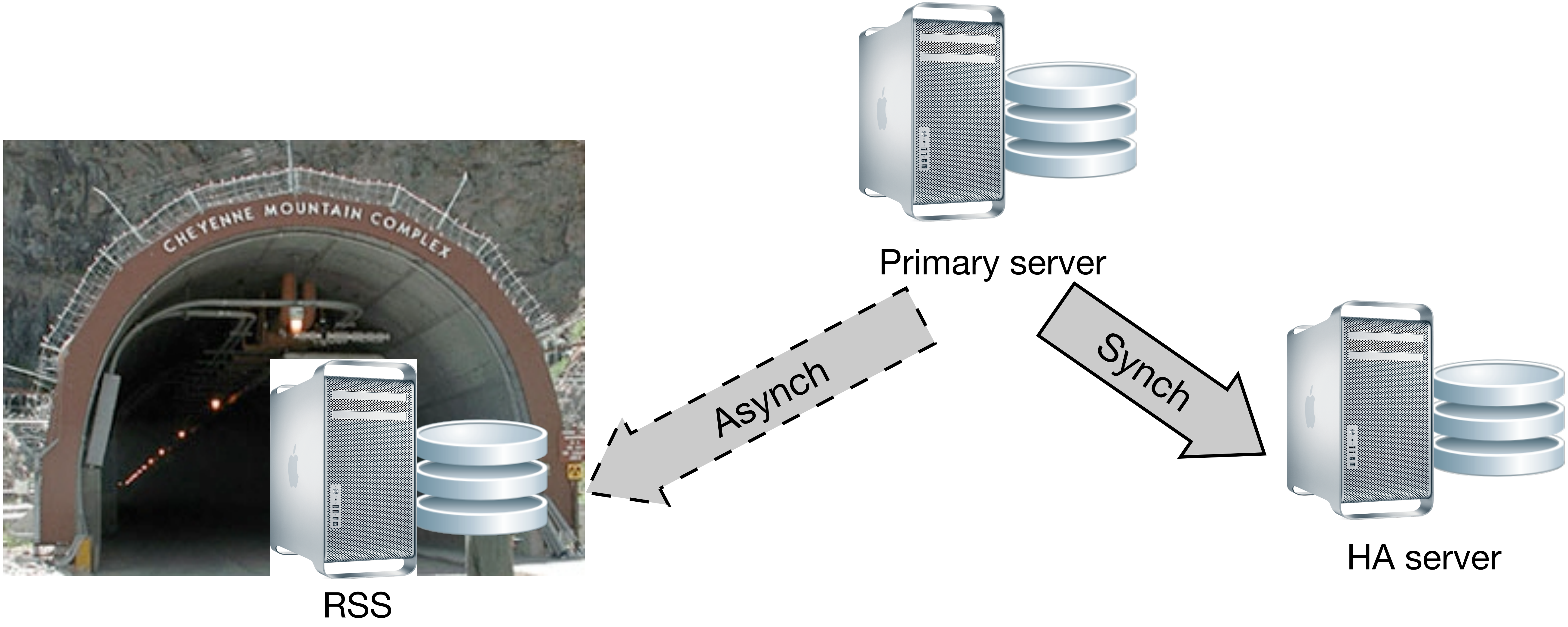
Remote Standalone Server

Data Replication

- Technology: Remote Standalone Server (RSS)
- Similar to HA server :
  - Maintains a full disk copy of the database
  - Can be used for report processing OR full application load
    - depends on data currency needs
- Distinct from HA server:
  - Uses full duplex communication – better throughput over slower lines or over longer distances
  - Can not be directly ‘promoted’ to primary – but can be promoted to HA server – FOR DISASTER RECOVERY (DR), not HA
  - There can be any number of RSS instances
- RSS can be used in combination with HA server
  - RSS can be promoted into HA server
  - HA server can be demoted into RSS

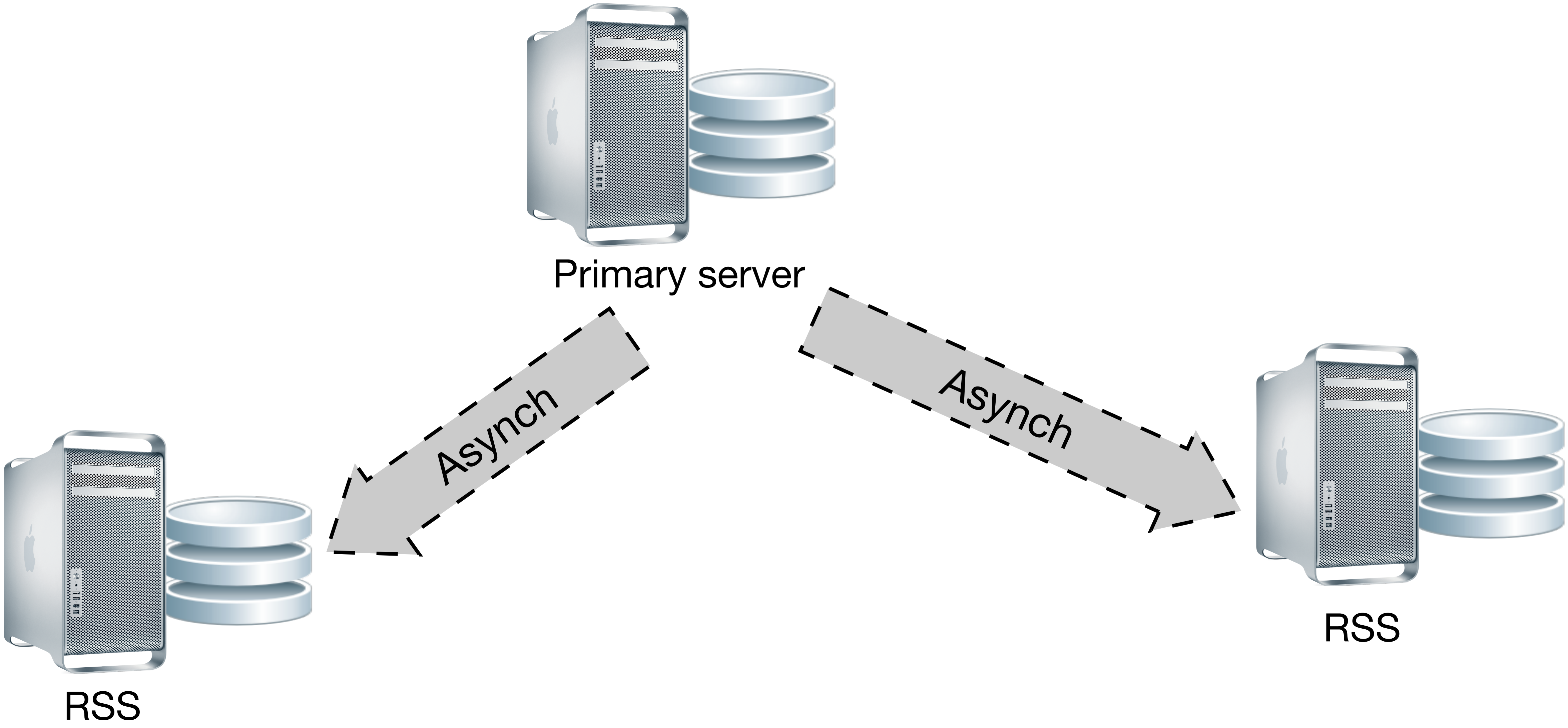
Data Replication

- Technology: Remote Standalone Server (RSS)
- Deployment scenario: bunker backup
  - Using HA server but want an additional backup in case both the primary and the secondary are lost
  - Located in a CA site



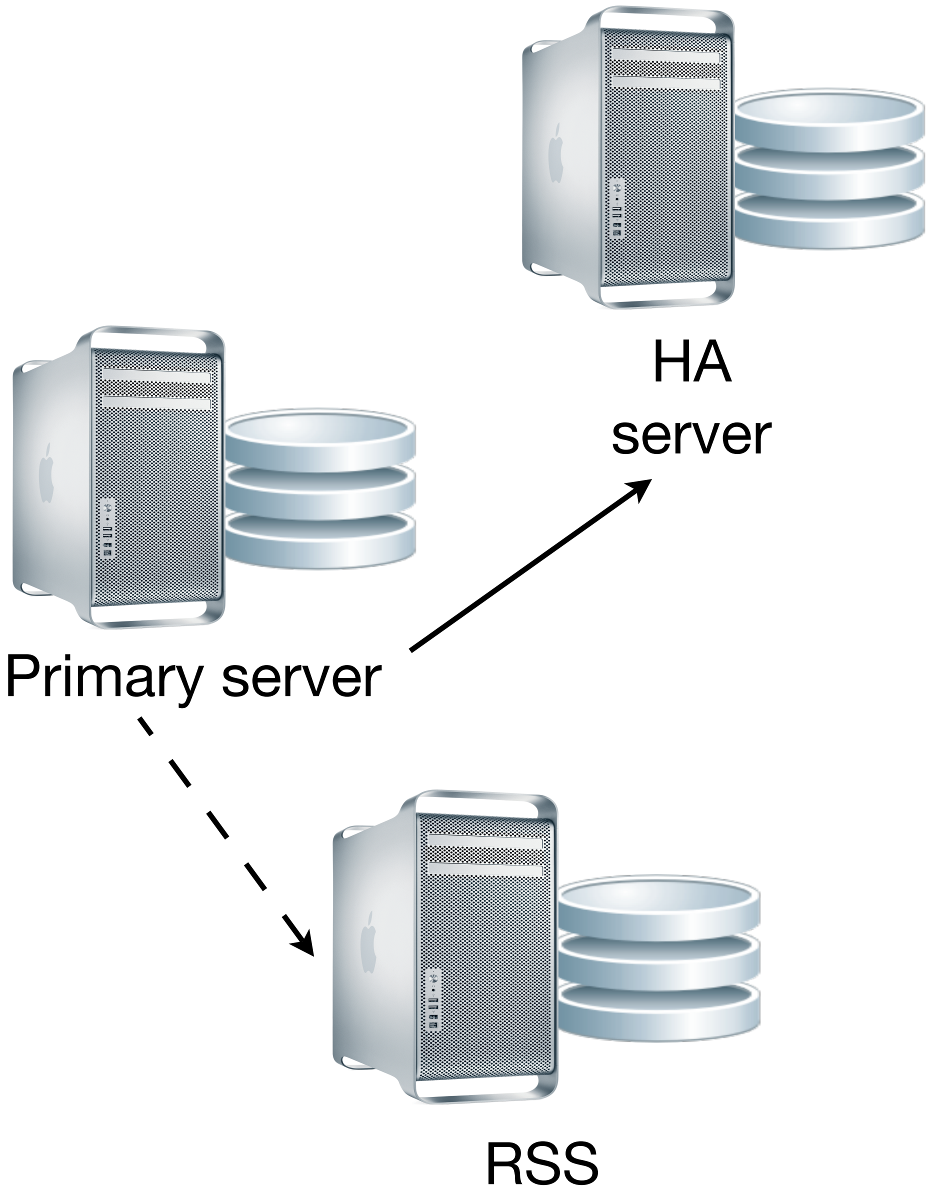
Data Replication

- Technology: Remote Standalone Server (RSS)
- Deployment scenario: poor network latency
  - Throughput tests indicate excessive transport times for an HA server
  - Asynch communication eliminates performance impact on the primary



Data Replication

- Technology: Remote Standalone Server (RSS)
- Deployment scenario: backup to the backup (1:N)
  - Automatic (or manual) promotion so there is always an HA server protecting the primary



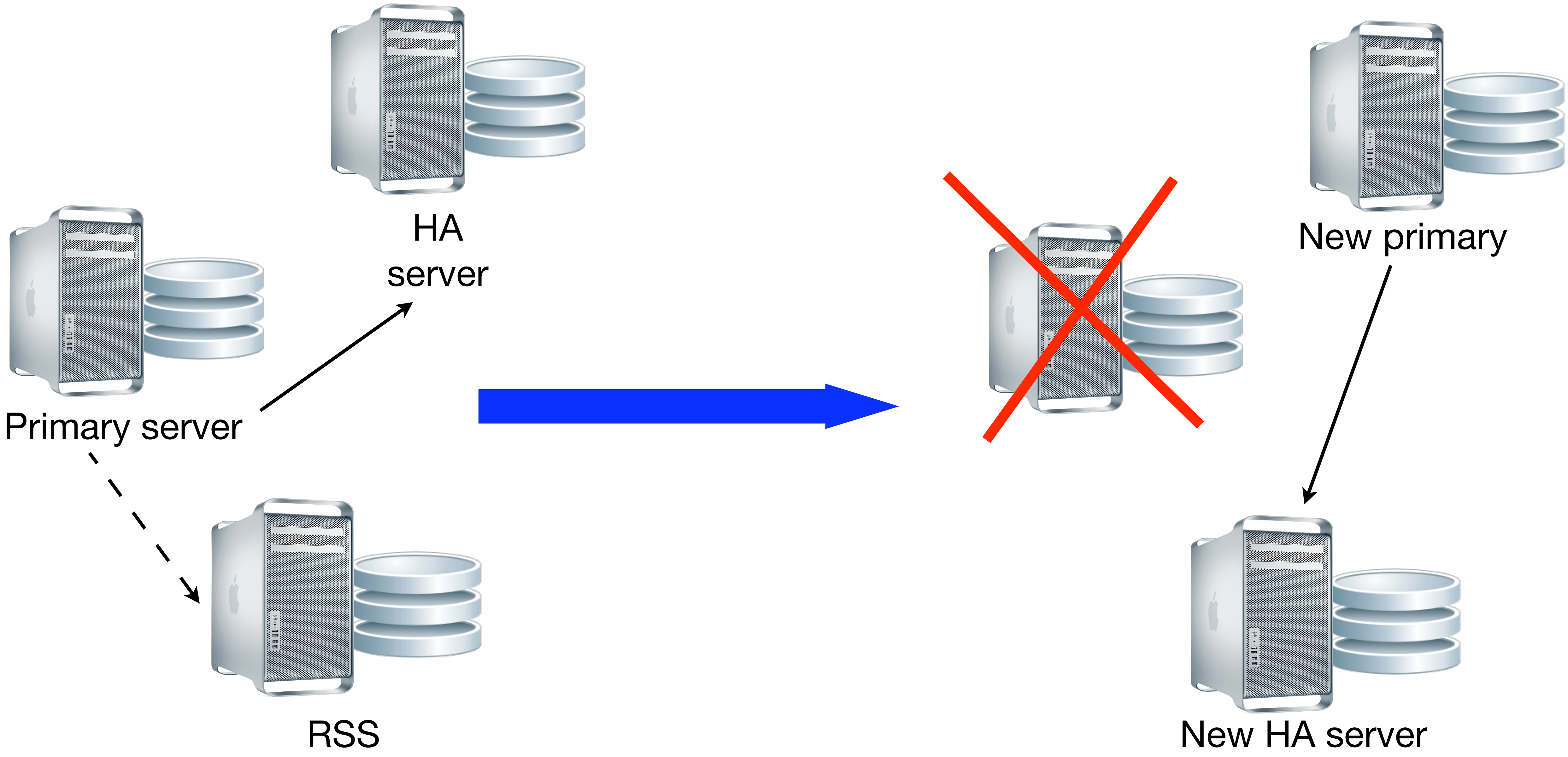
HA server

Primary server

RSS

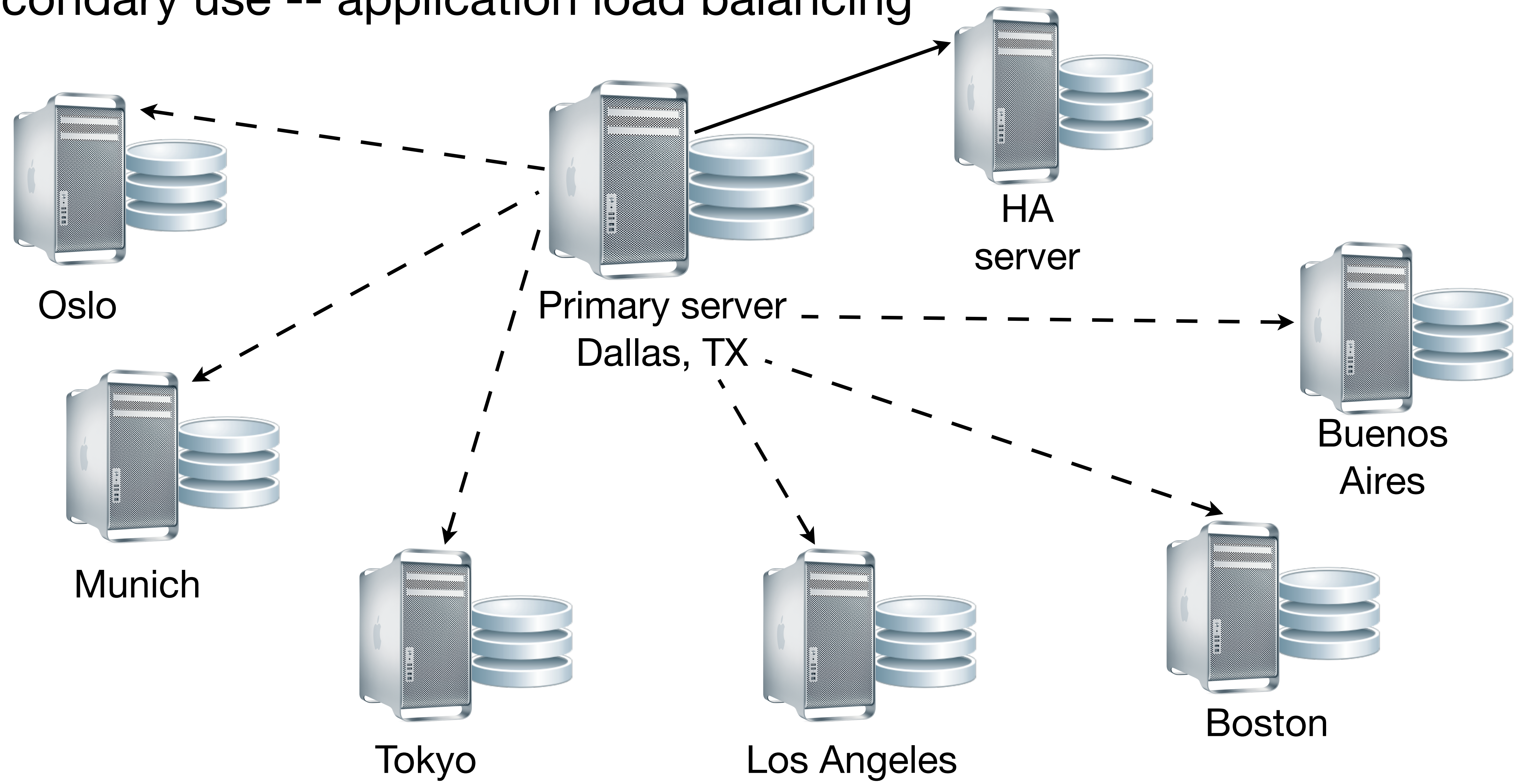
Data Replication

- Technology: Remote Standalone Server (RSS)
- Deployment scenario: backup to the backup (1:N)
  - Automatic (or manual) promotion so there is always an HA server protecting the primary



Data Replication

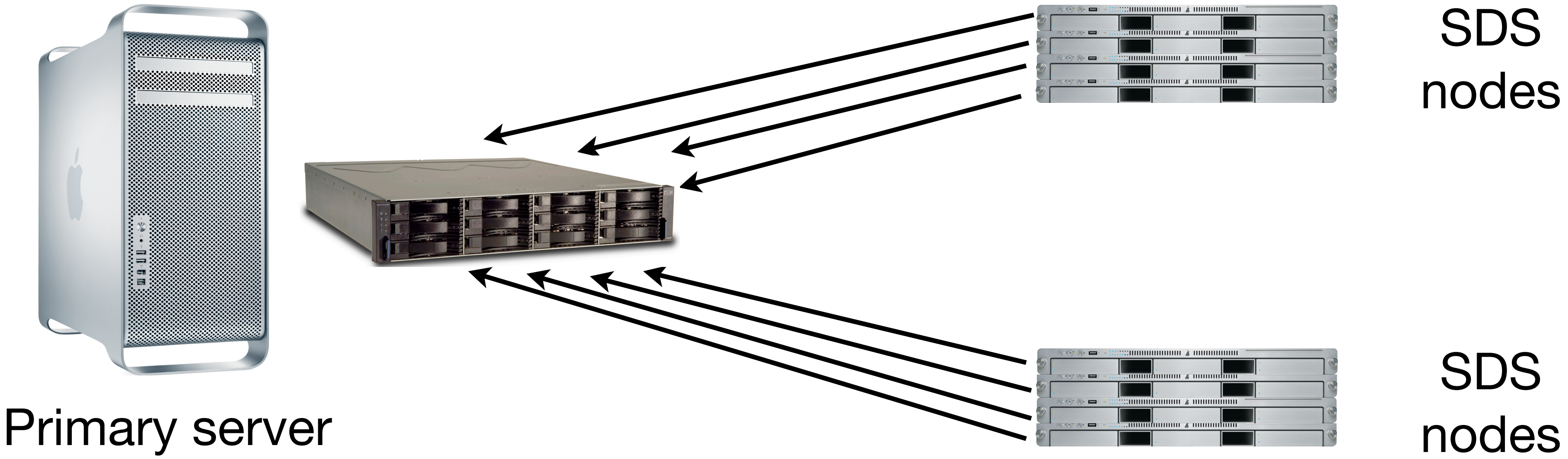
- Technology: Remote Standalone Server (RSS)
- Deployment scenario: disaster recovery (DR) protection
  - Geographically distributed servers to protect against site or system failure
  - Secondary use -- application load balancing



Shared Disk Secondary

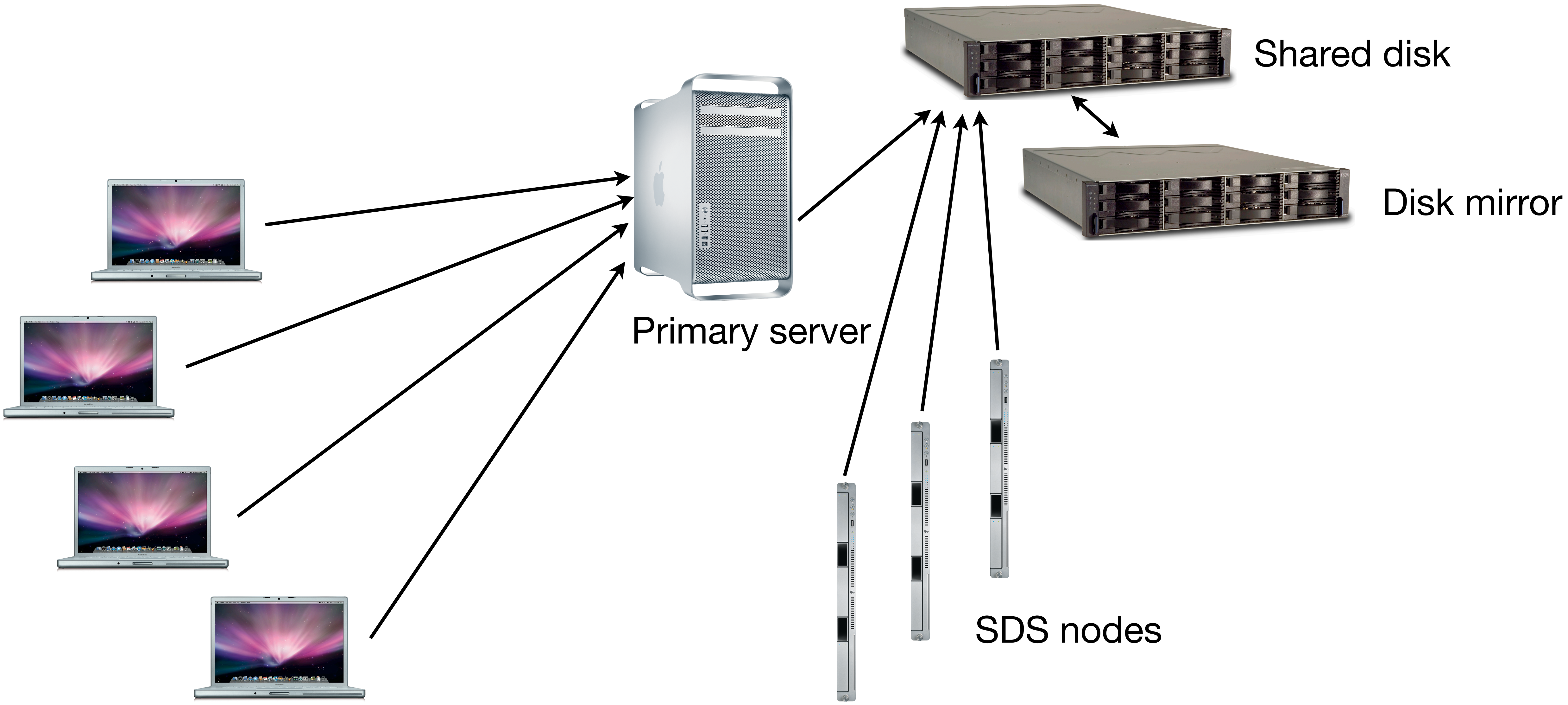
Data Replication

- Technology -- Shared Disk Server (SDS)
- HA server built using single set of shared disk
  - CA and HA oriented
    - Can act as an immediate fail-over



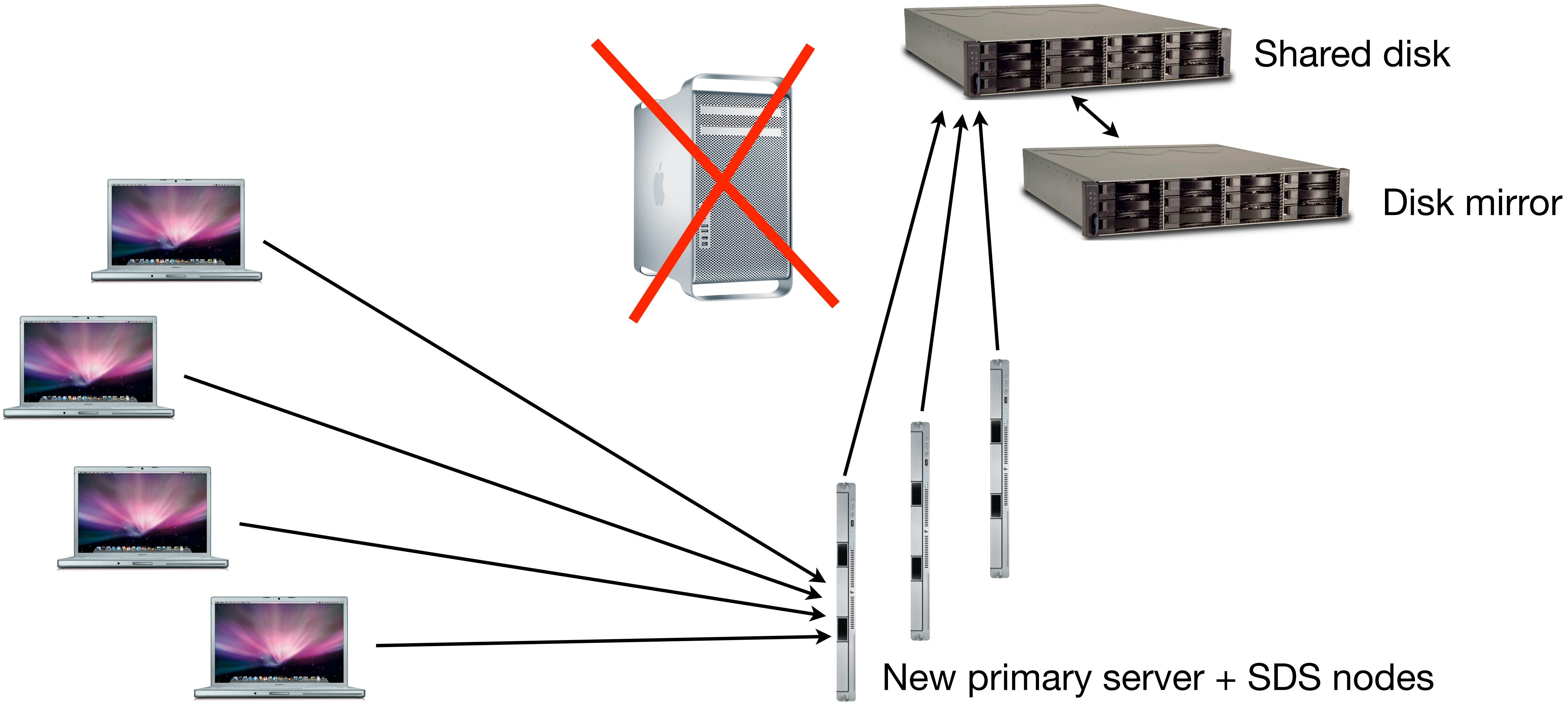
Data Replication

- Technology -- Shared Disk Server (SDS)
- Deployment scenario -- HA protection against server failure
  - Doesn't require an equivalent investment in storage



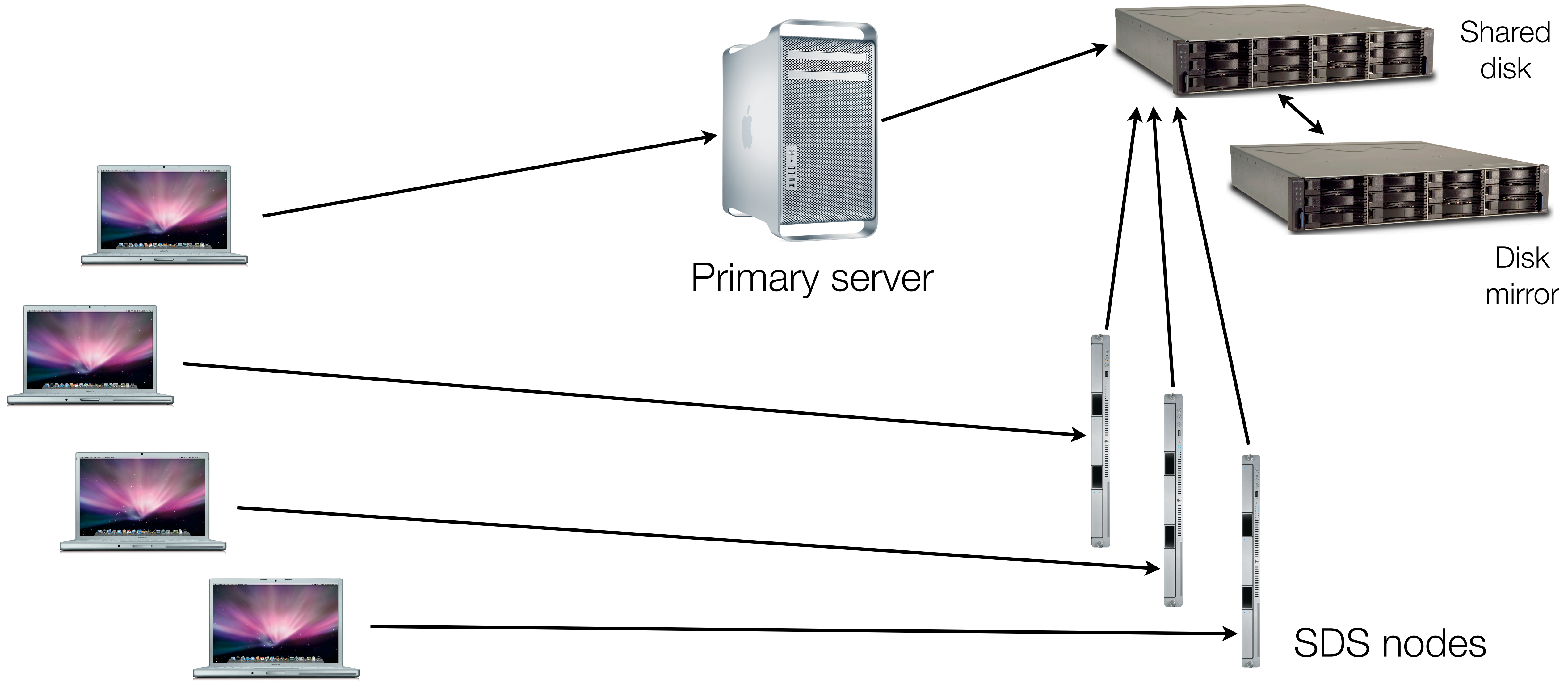
Data Replication

- Technology -- Shared Disk Server (SDS)
- Deployment scenario -- HA protection against server failure
  - Doesn't require an equivalent investment in storage



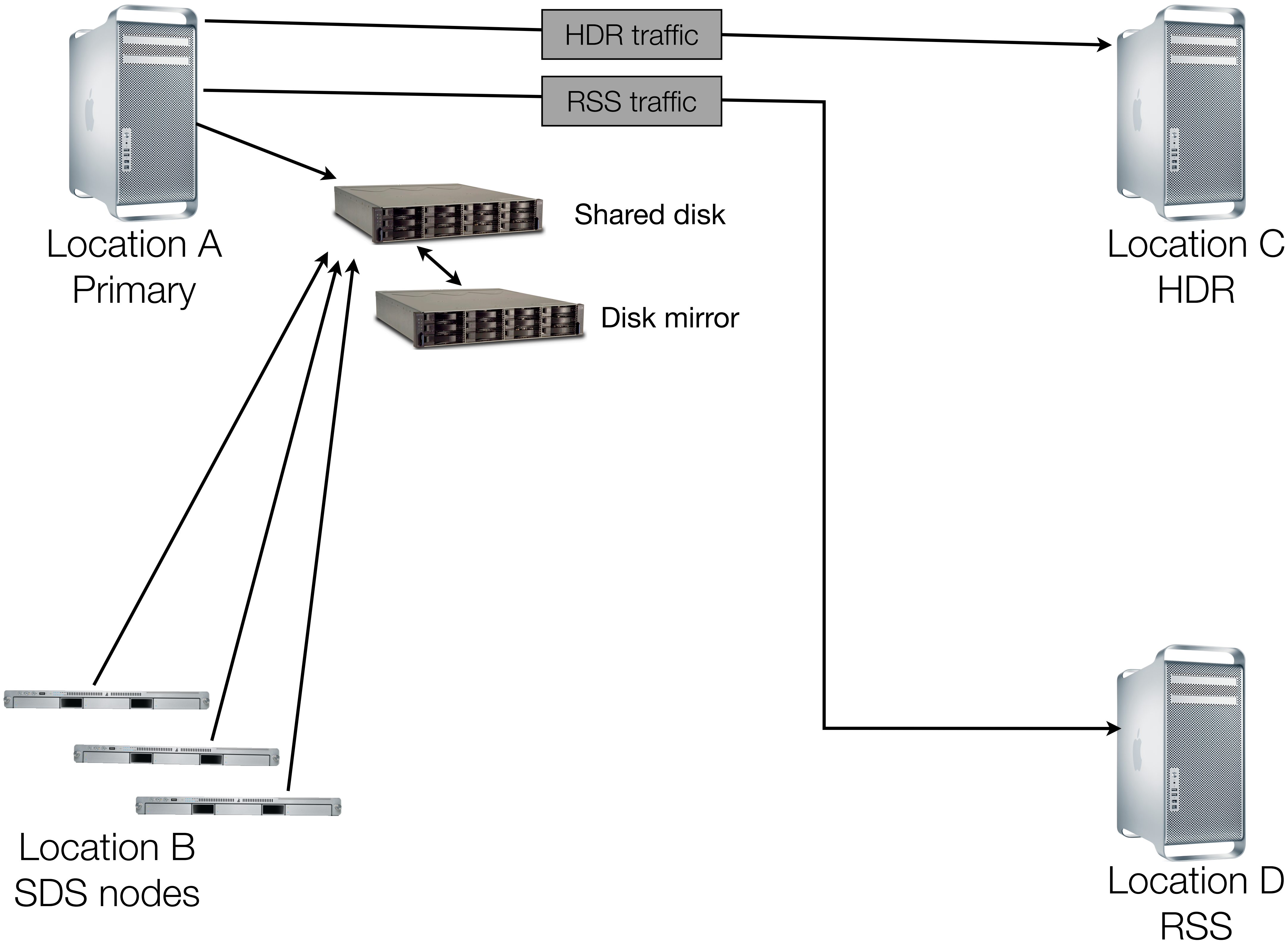
Data Replication

- Technology -- Shared Disk Server (SDS)
- Deployment scenario -- compute intensive application balancing
  - Distribute low to average read/write activity but heavy calculation overhead



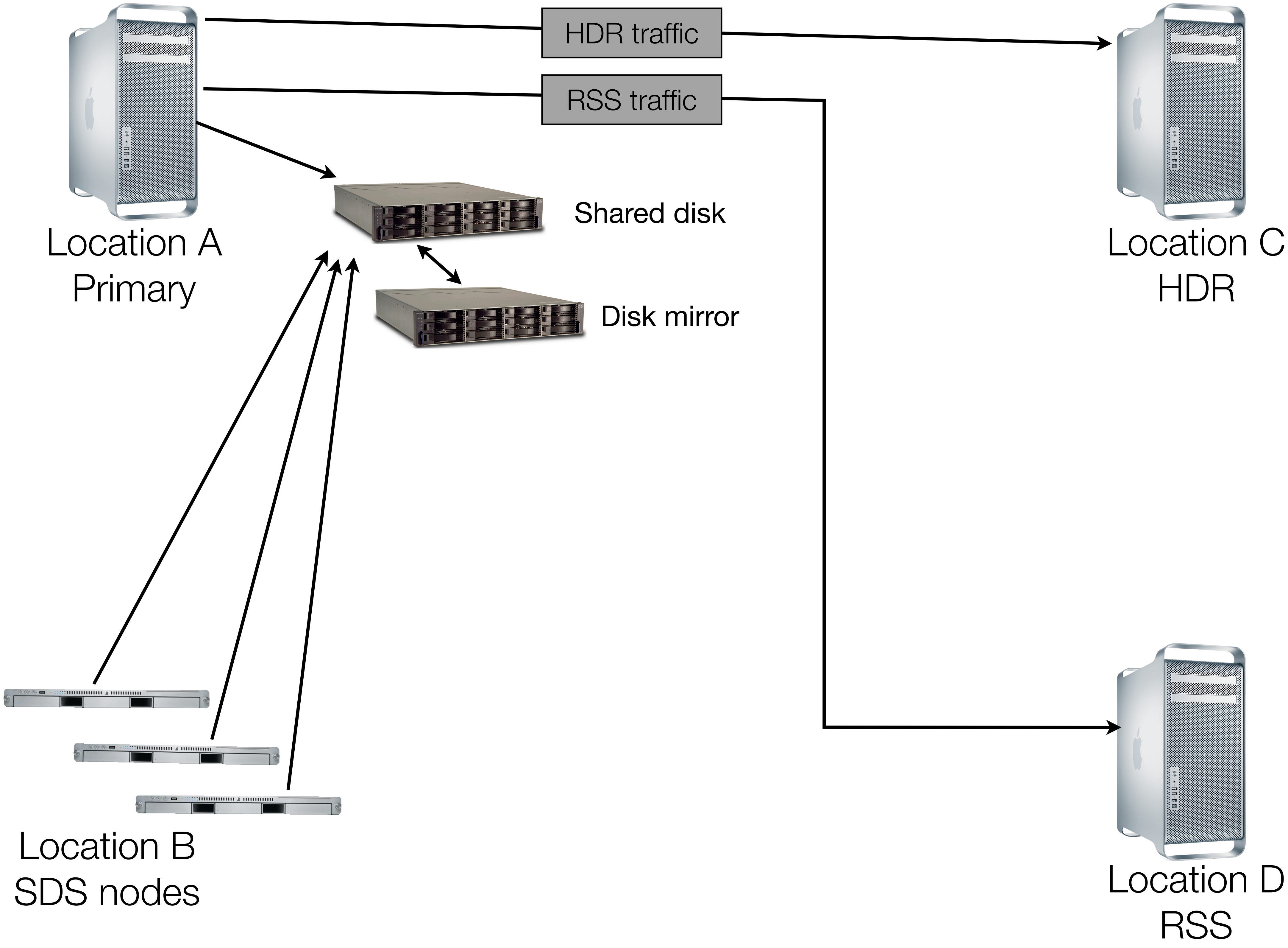
Putting it all together

Putting it all together



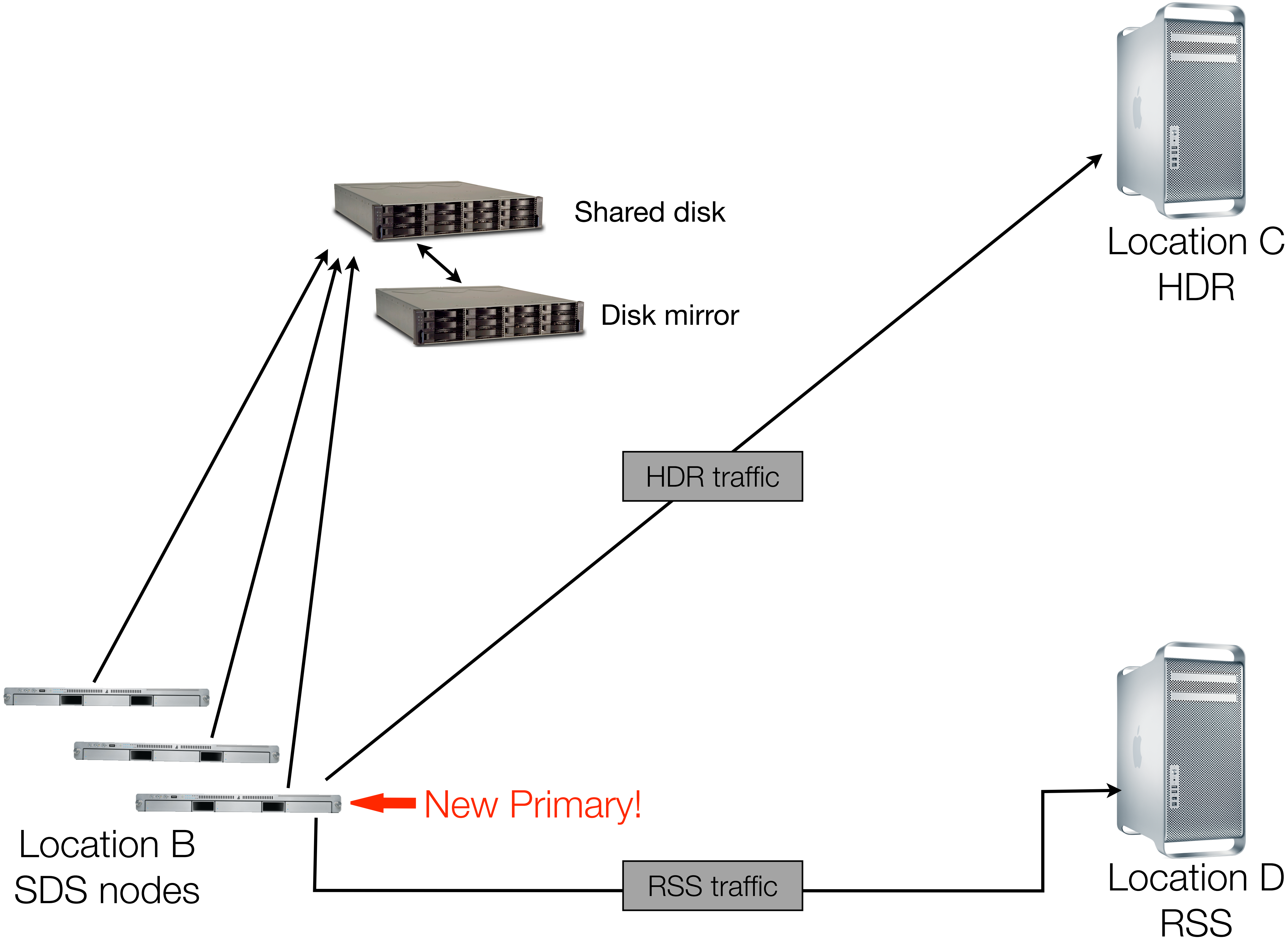
Putting it all together

Location A fails



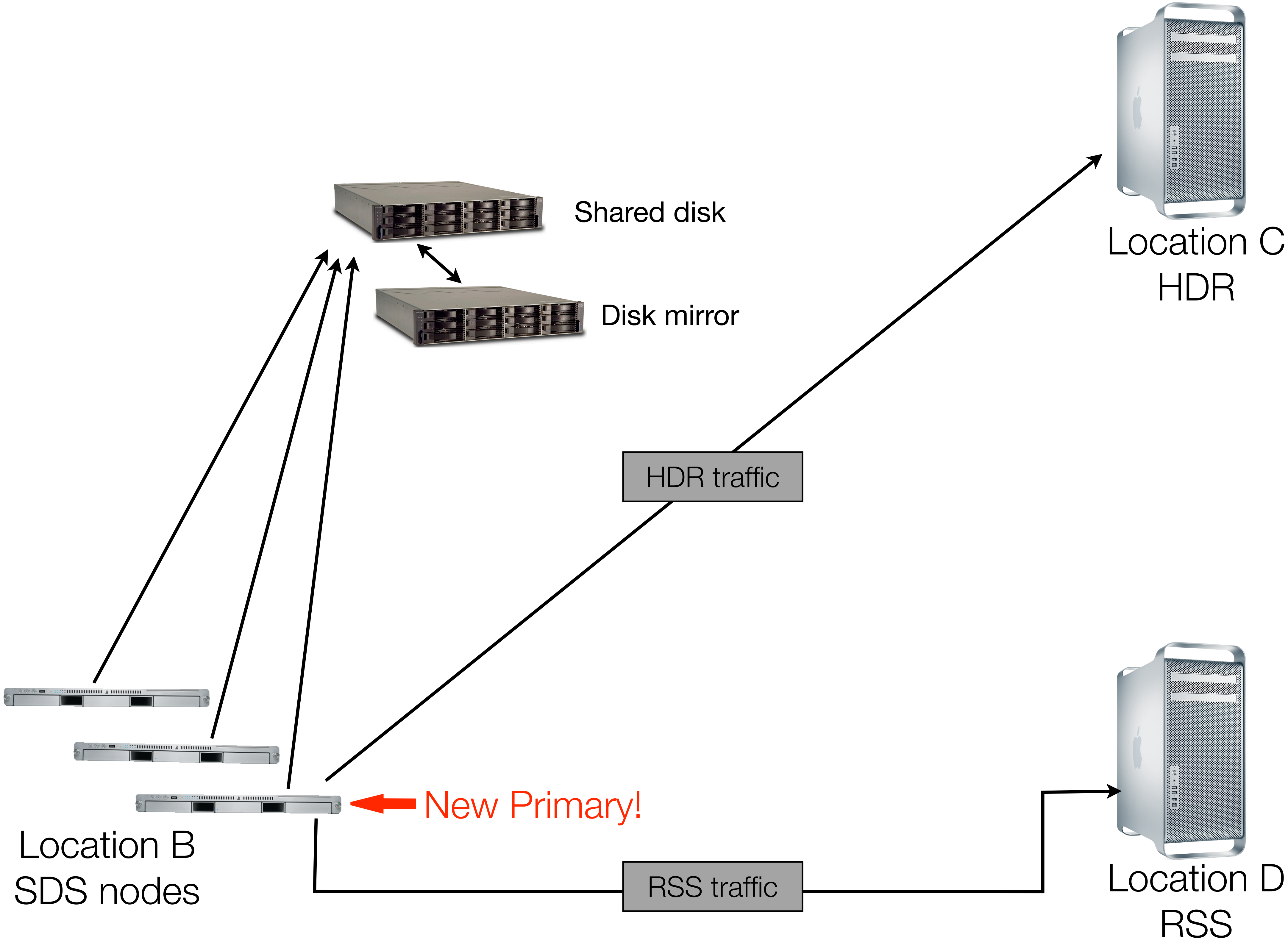
Putting it all together

Location A fails



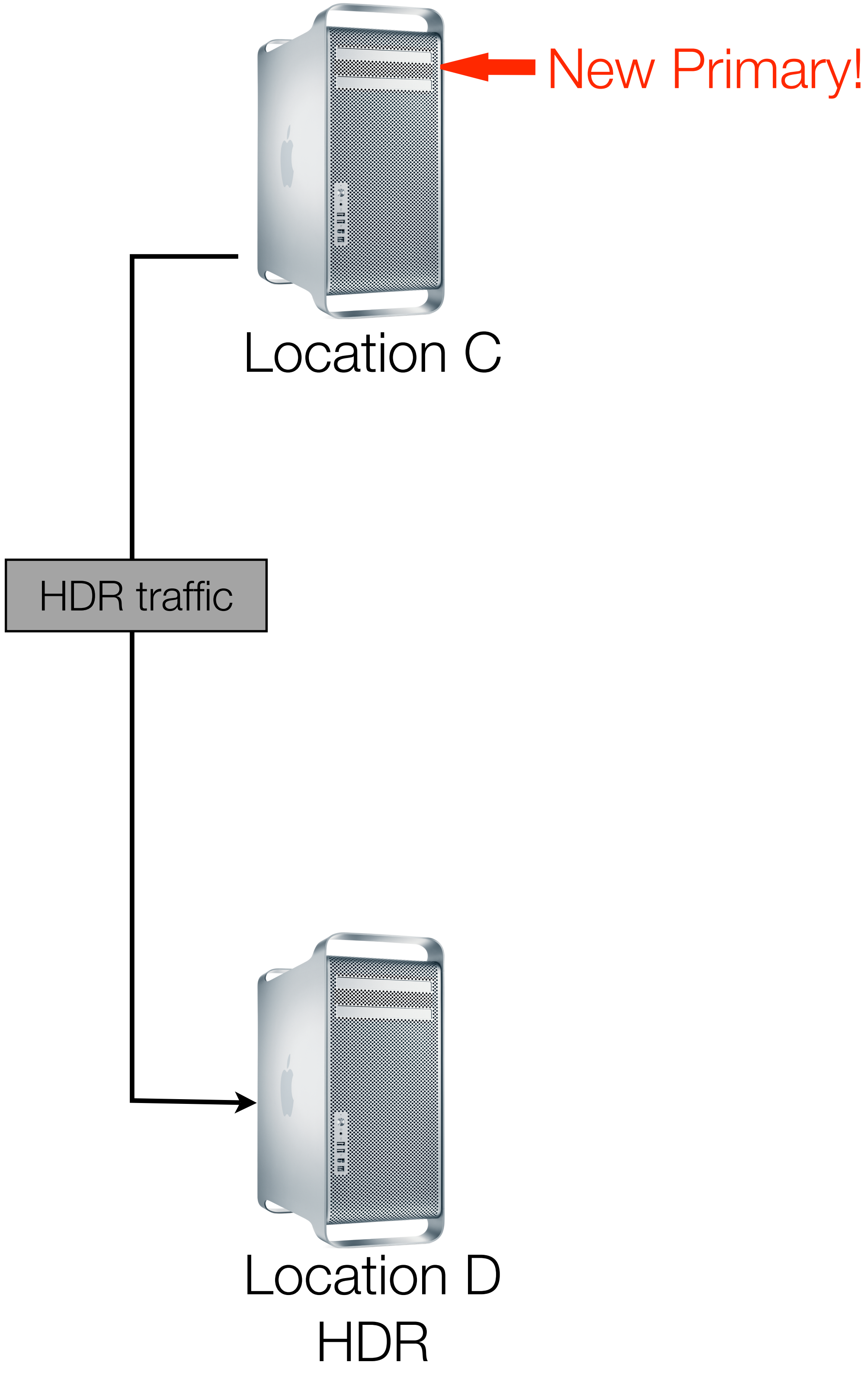
Putting it all together

Location B fails



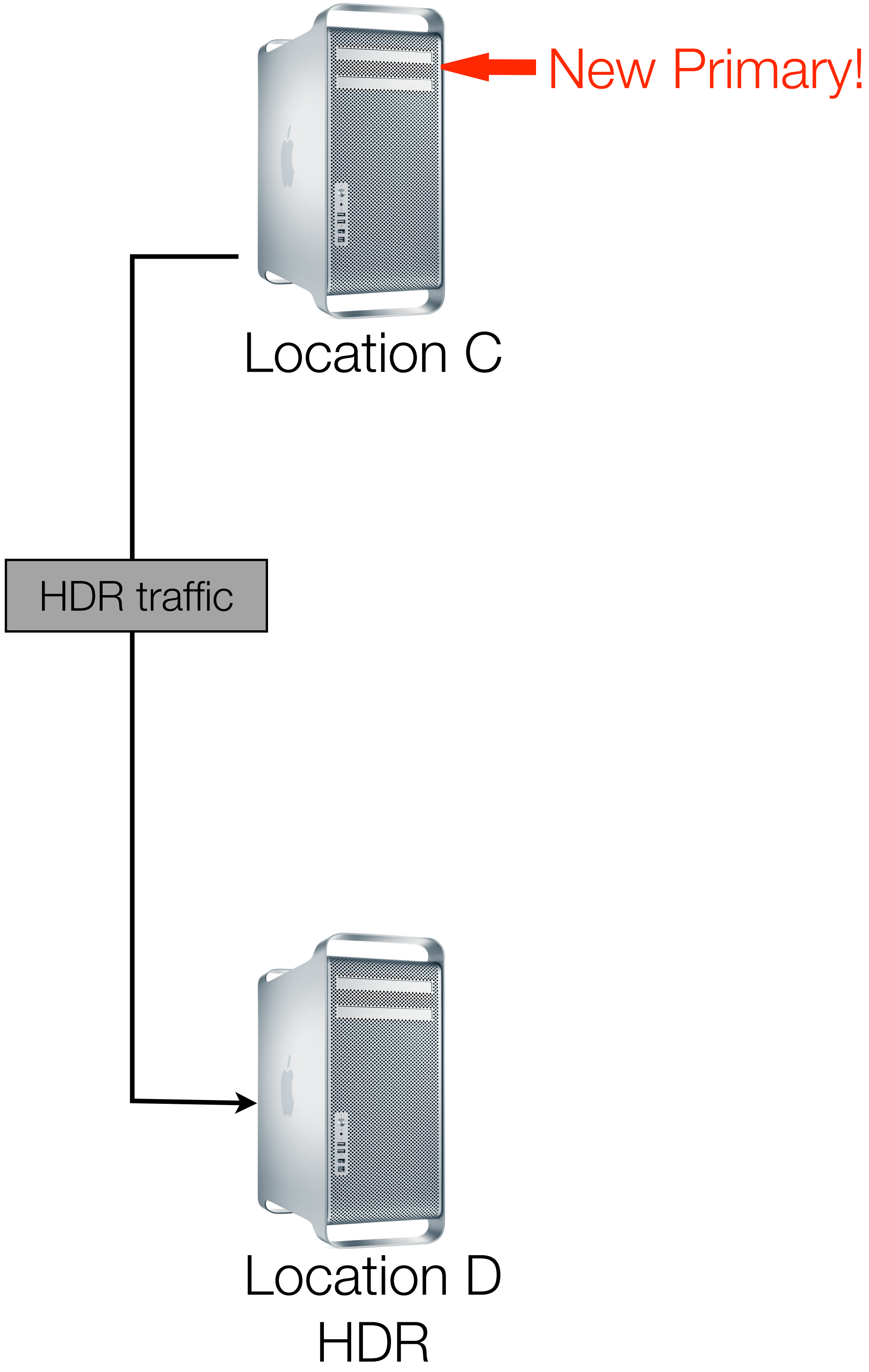
Putting it all together

Location B fails



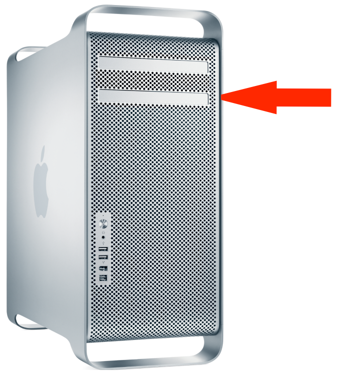
Putting it all together

Location C fails



Putting it all together

Location C fails

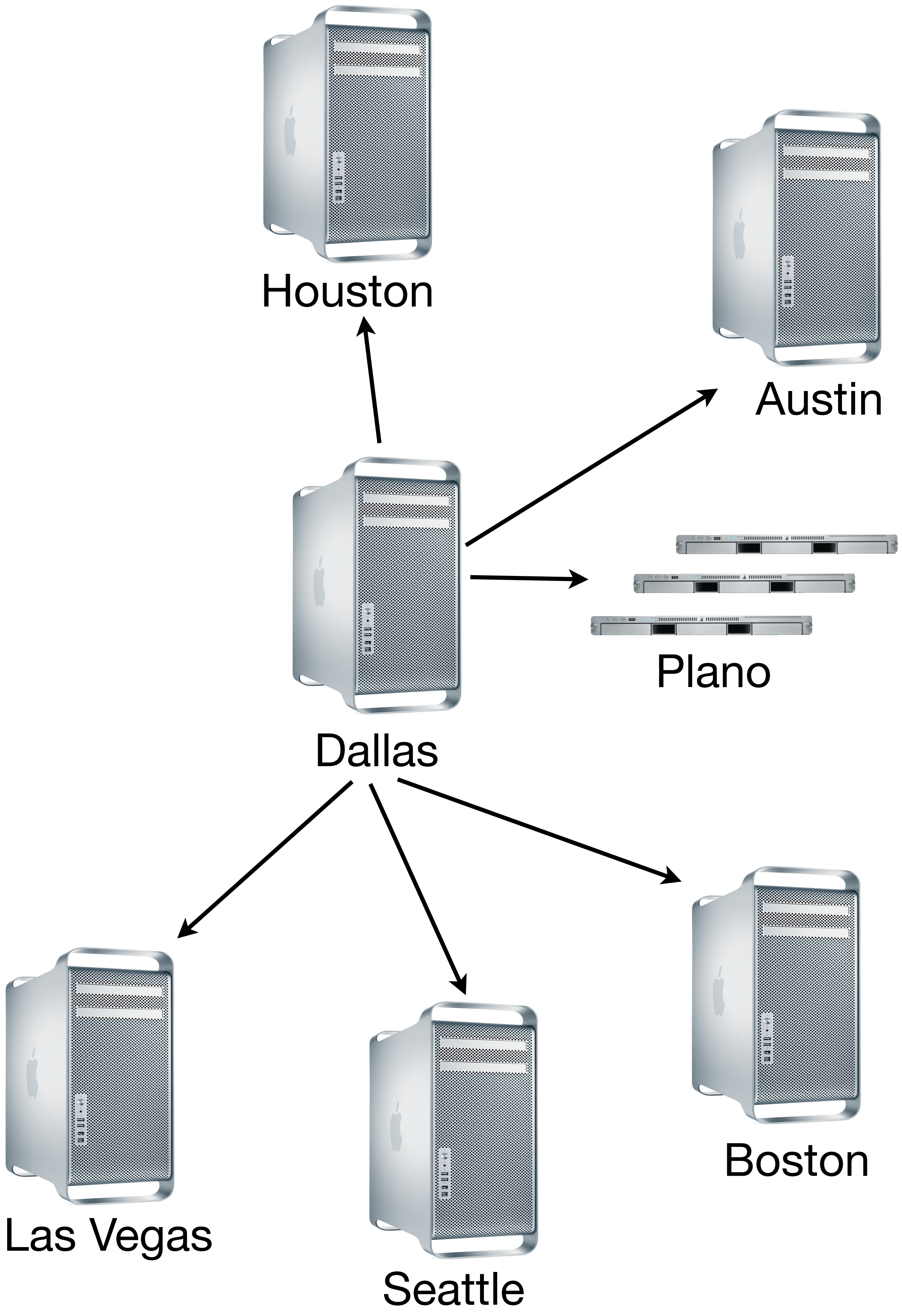


New Primary!

Location D

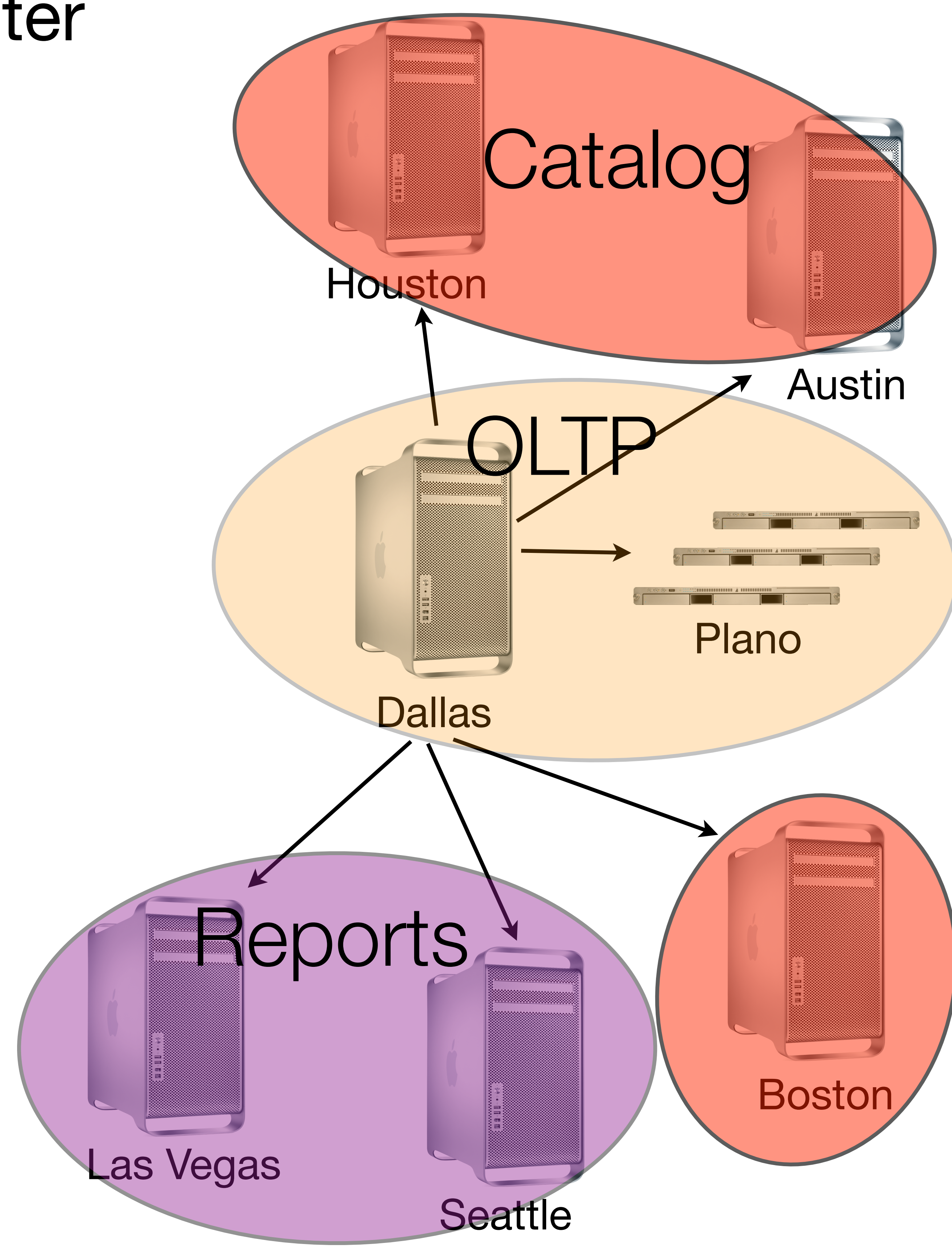
Deriving additional business value

- Load balancing across the cluster



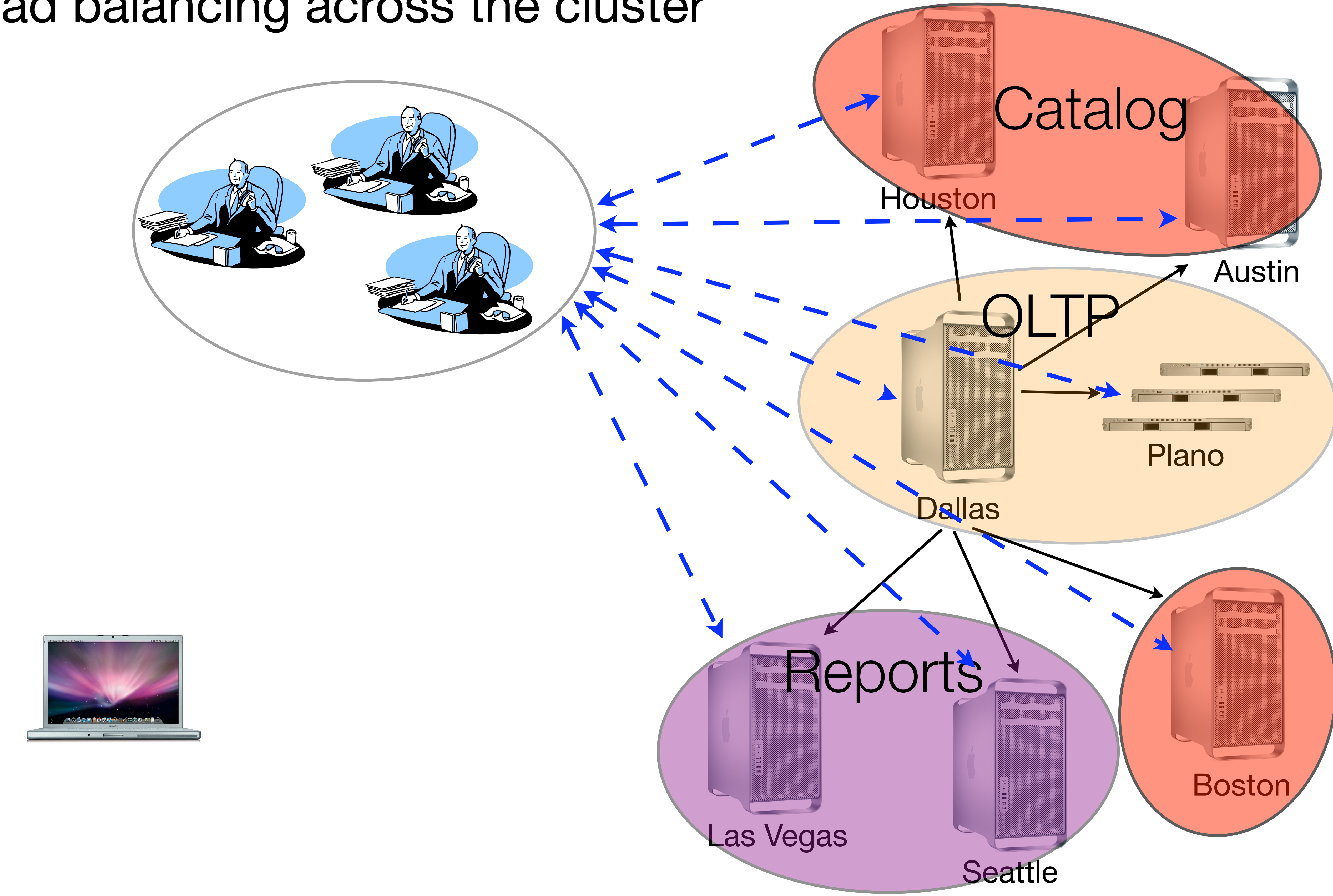
Deriving additional business value

- Load balancing across the cluster



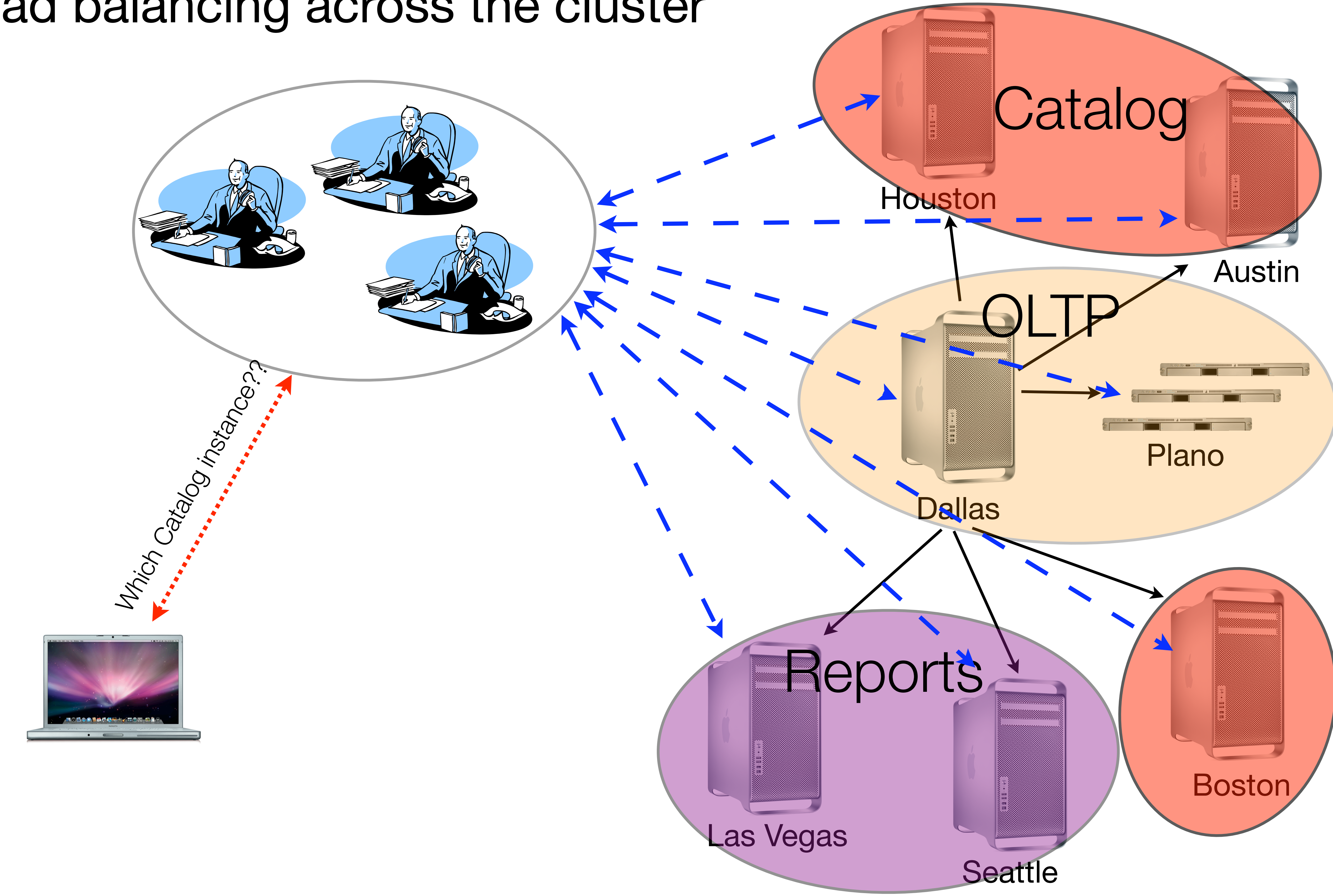
Deriving additional business value

- Load balancing across the cluster



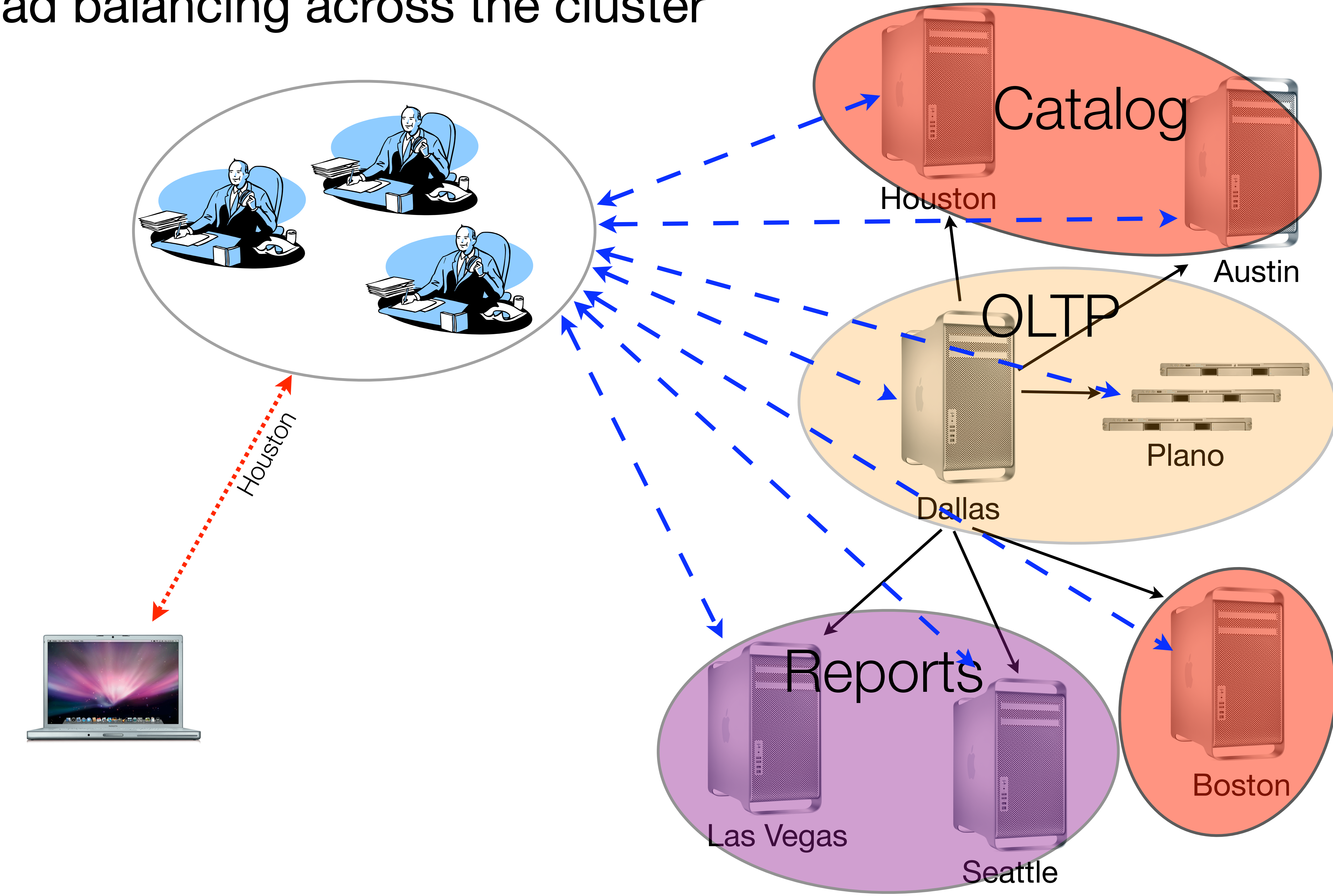
Deriving additional business value

- Load balancing across the cluster



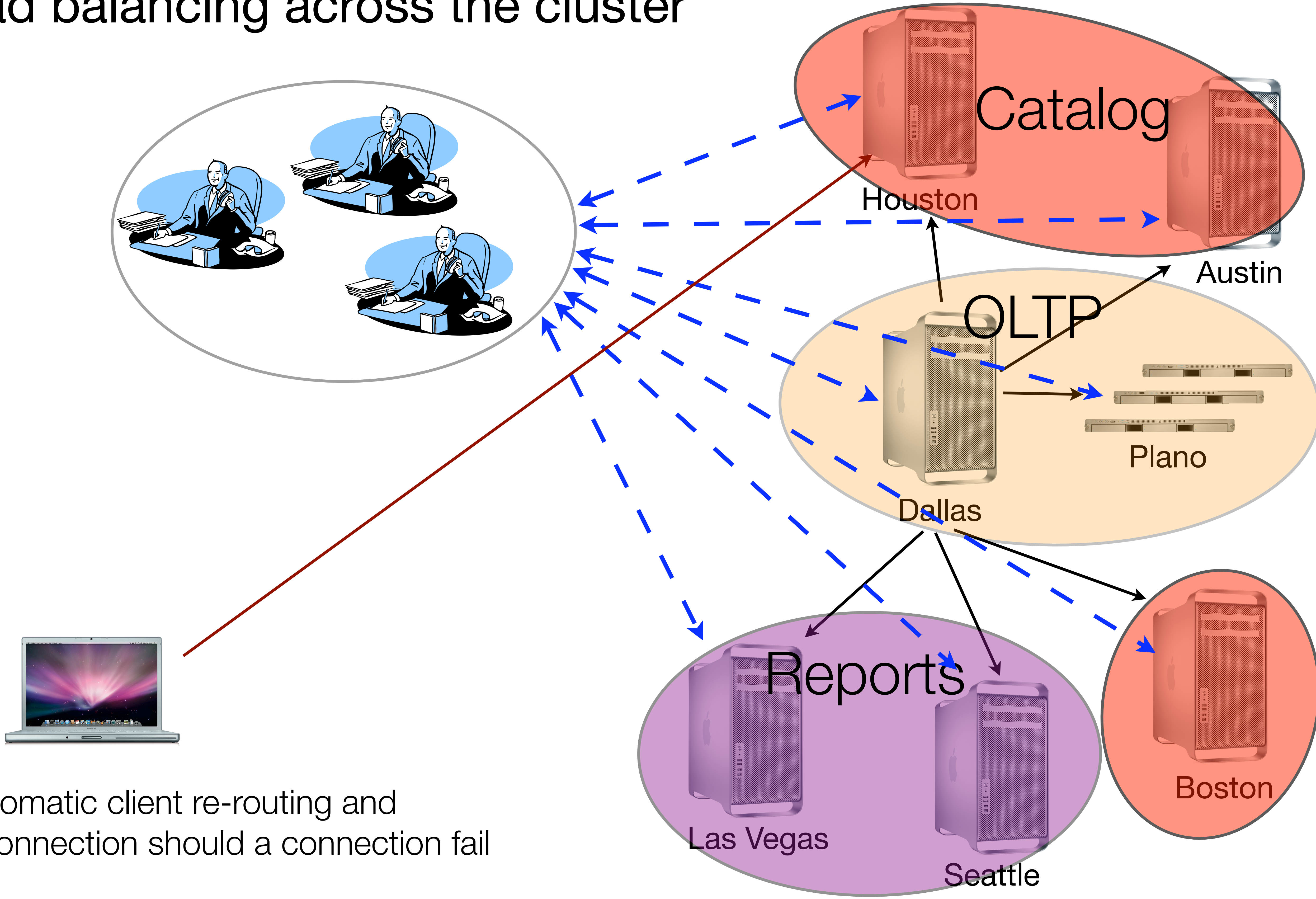
Deriving additional business value

- Load balancing across the cluster



Deriving additional business value

- Load balancing across the cluster



Automatic client re-routing and reconnection should a connection fail

End of part 1, data replication

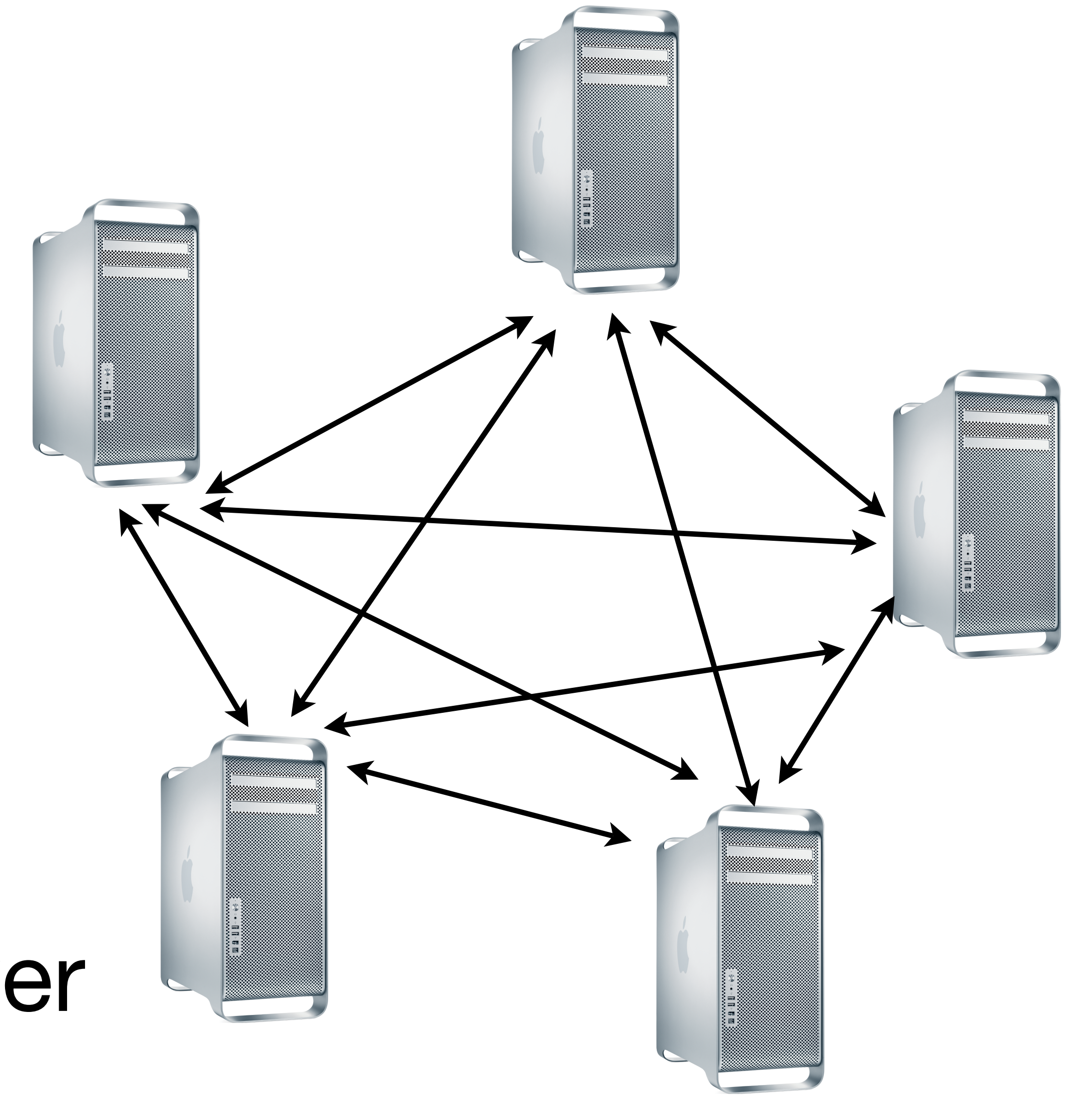
Data Distribution --
Full or selective copies over one or more topologies

Data Distribution

- Guiding principle is data distribution throughout the company or enterprise. Will call it **Enterprise Replication** (ER)
  - Data server driven (no application code)
  - Is data object driven
    - Defined via SQL at the column, row, table or database level
      - Can be selective within
      - Can mix and match as needed

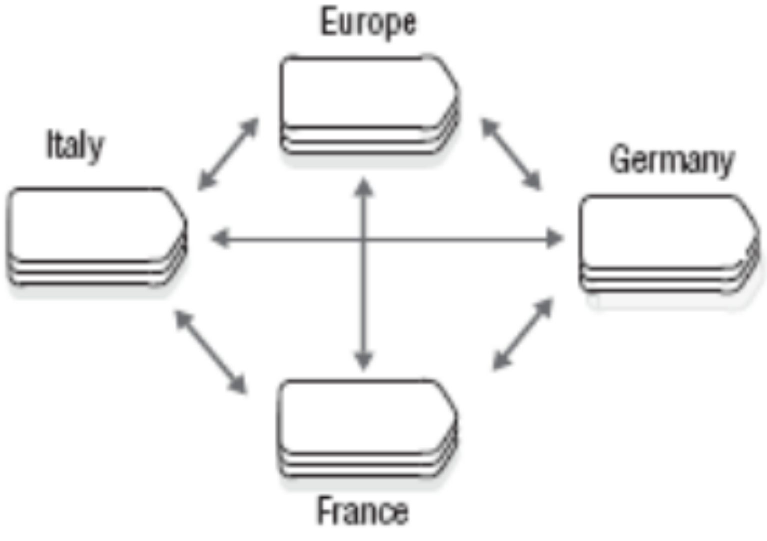
Data Distribution

- Enterprise data distribution
  - Designed for active/active use
  - Very low latency
  - Synchronize regional and enterprise data
- Flexible scalability
  - Subset of data
  - Local data design
- Integrated, not add-on
- Compatible with other data server availability solutions
- Can provide secure data communication between servers

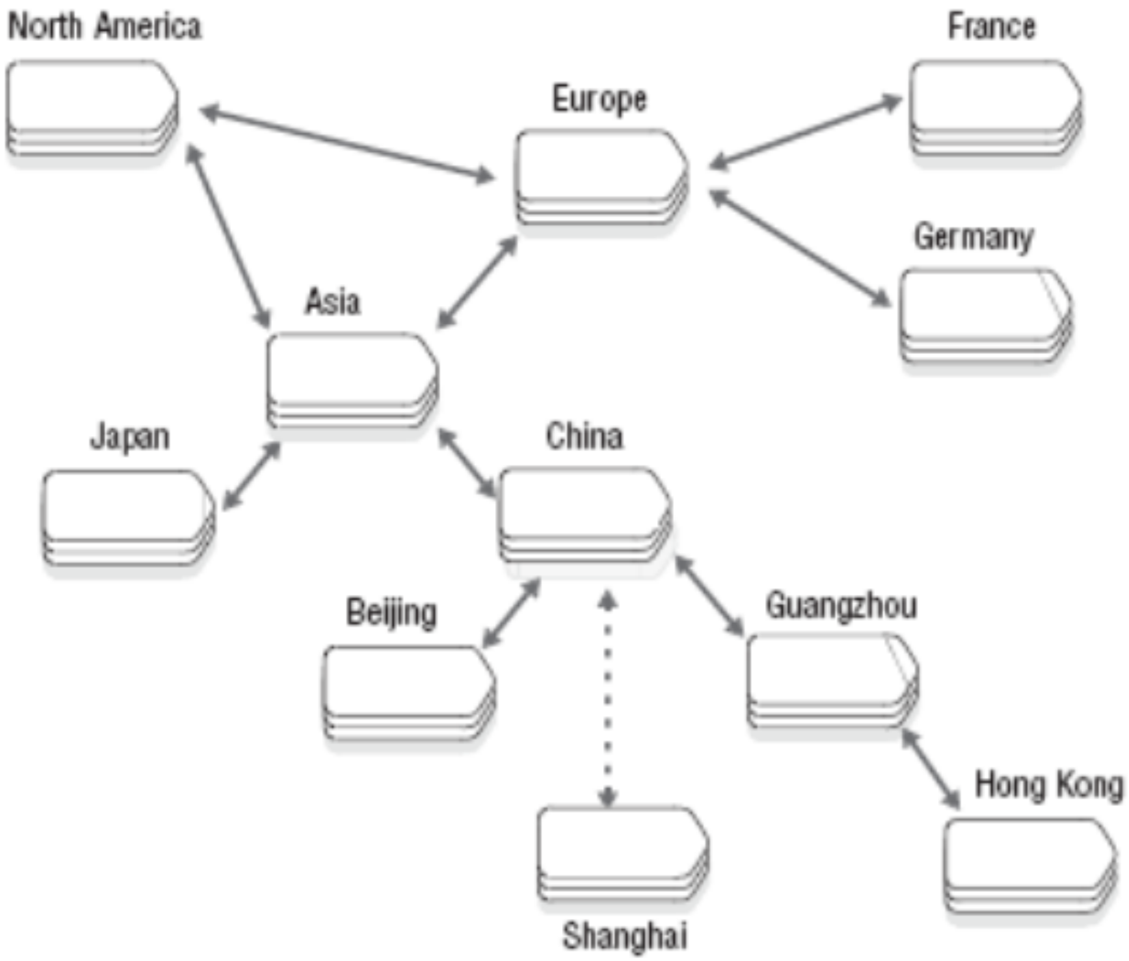


Data Distribution

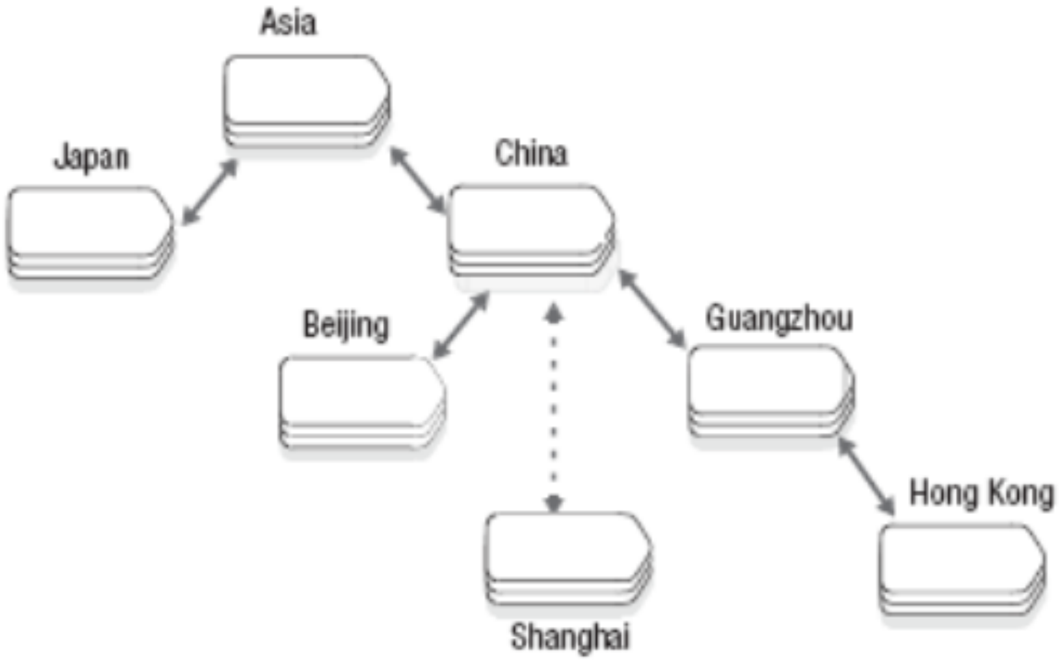
- The three “Ws” - What? Where? When?



Fully meshed – Update anywhere



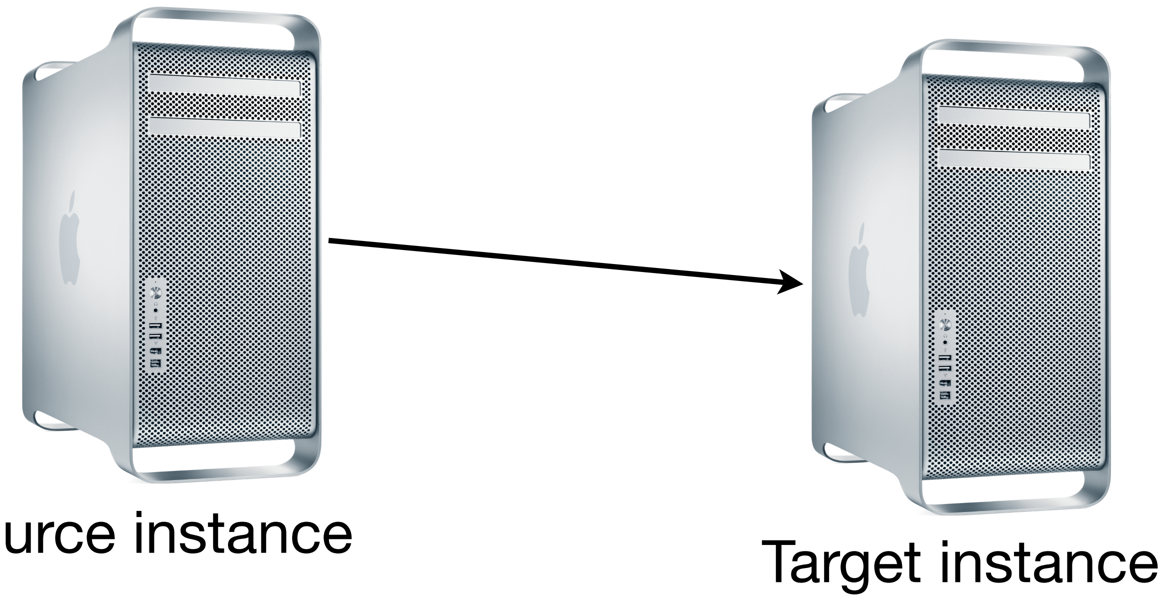
Snowflake / Forest of trees
Nodes share more with closest nodes, less with nodes farther away



Hierarchical – Push down, consolidate up

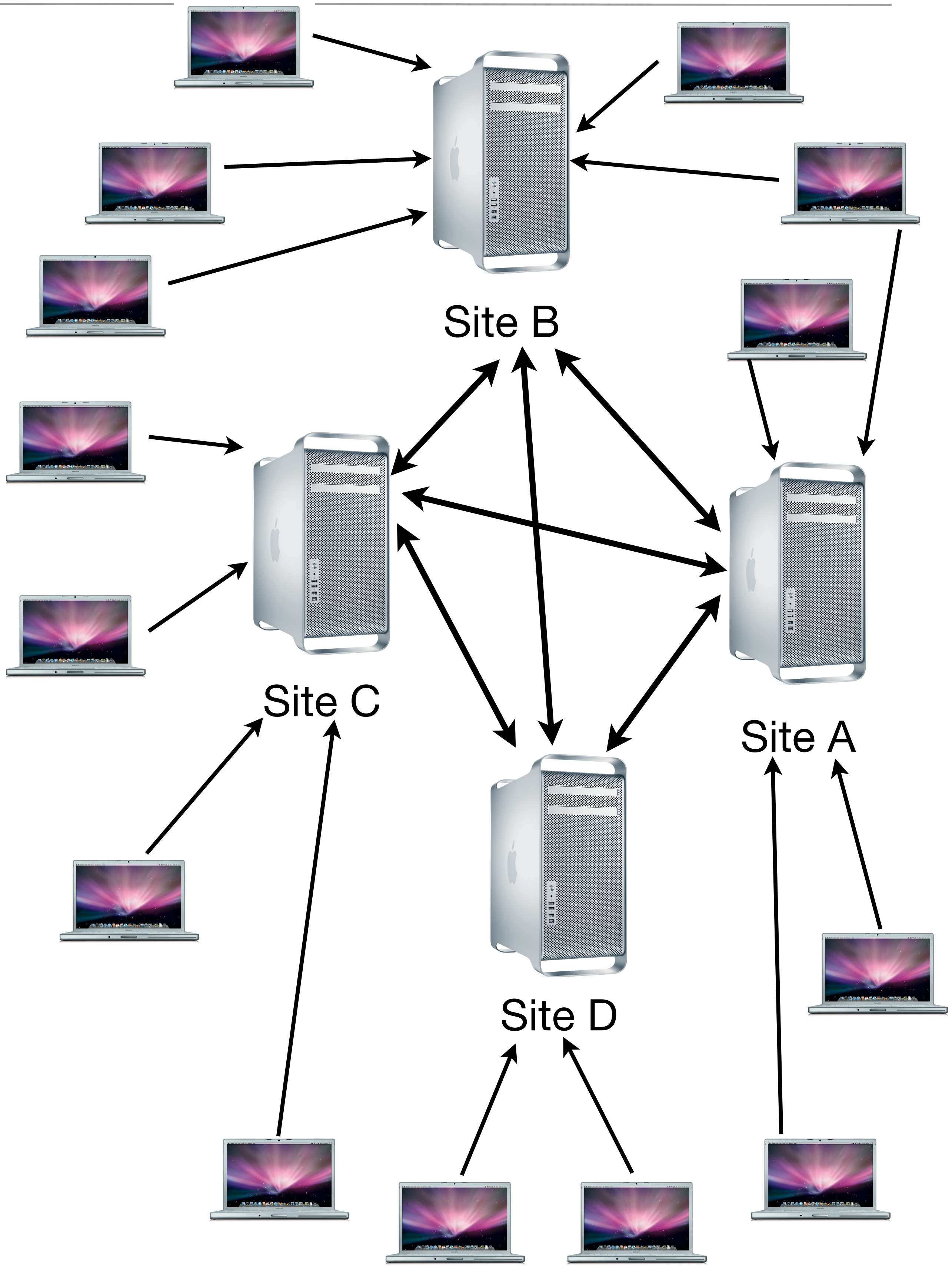
Data Distribution

- Scenario: Point-to-Point
- Is a DR, NOT an HA solution
  - It replicates data AFTER it’s been committed, not as part of the transaction
  - Can be configured to be a “close enough” option though
    - All work replicated
    - Target instance acts as a “very, very warm” stand by
      - Less than 1 minute latency depending on transaction volume, size and network considerations
    - Target available for reporting



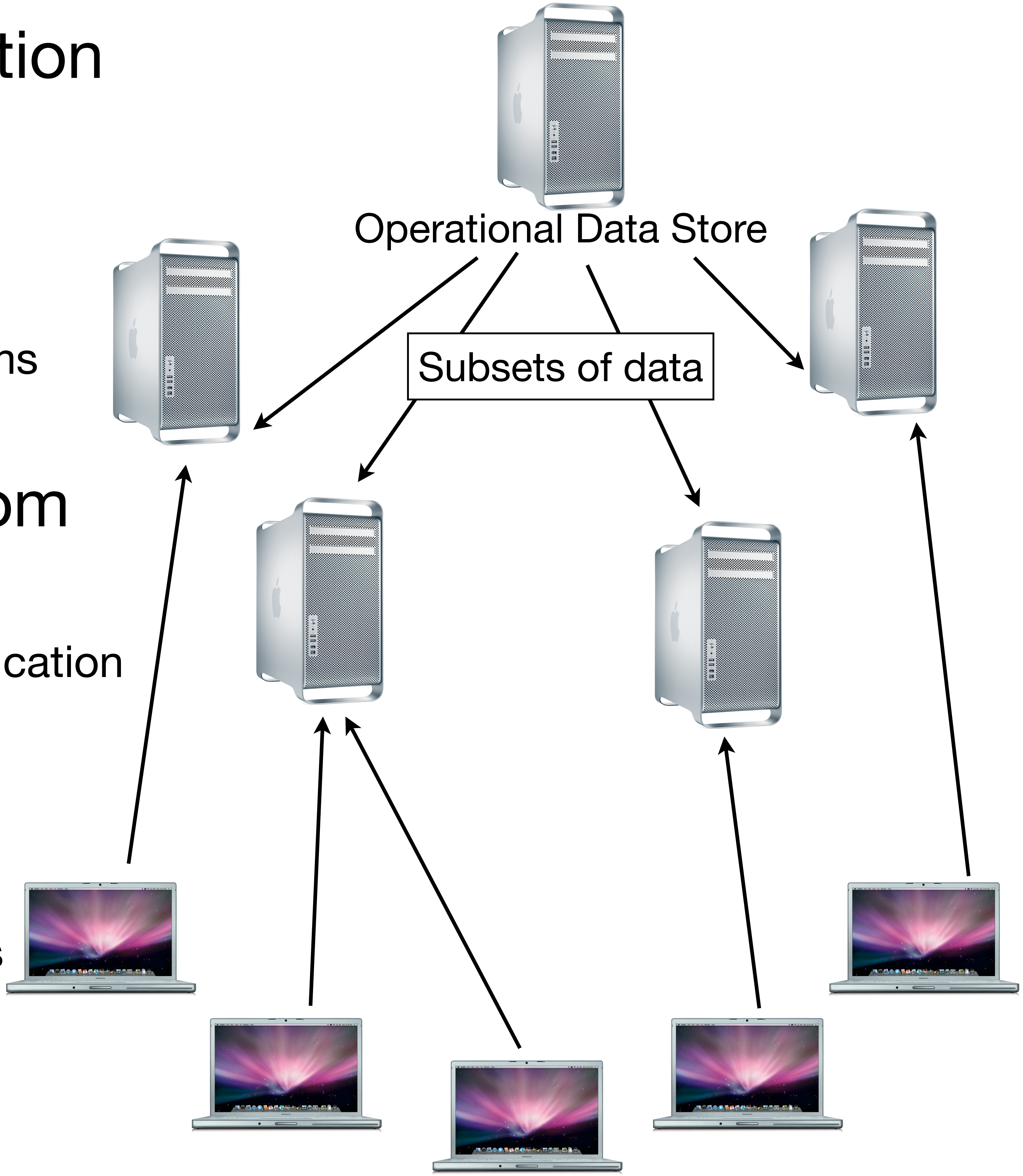
Data Distribution

- Scenario: Share Everything for Application Load-Balancing
- E-commerce site
  - Need to distribute data services load
  - Servers are located throughout the geographic service area
  - Customers can connect anywhere and complete transactions
  - Identical and <u>independent</u> data model and data across all data servers
- Is a “Close enough” solution
  - Data replicated after transaction
    - variability in moment-in-time absolute counts
  - Different sites would have different order number sequences
- Easily set up and administered
  - Check and/or resync data with one command



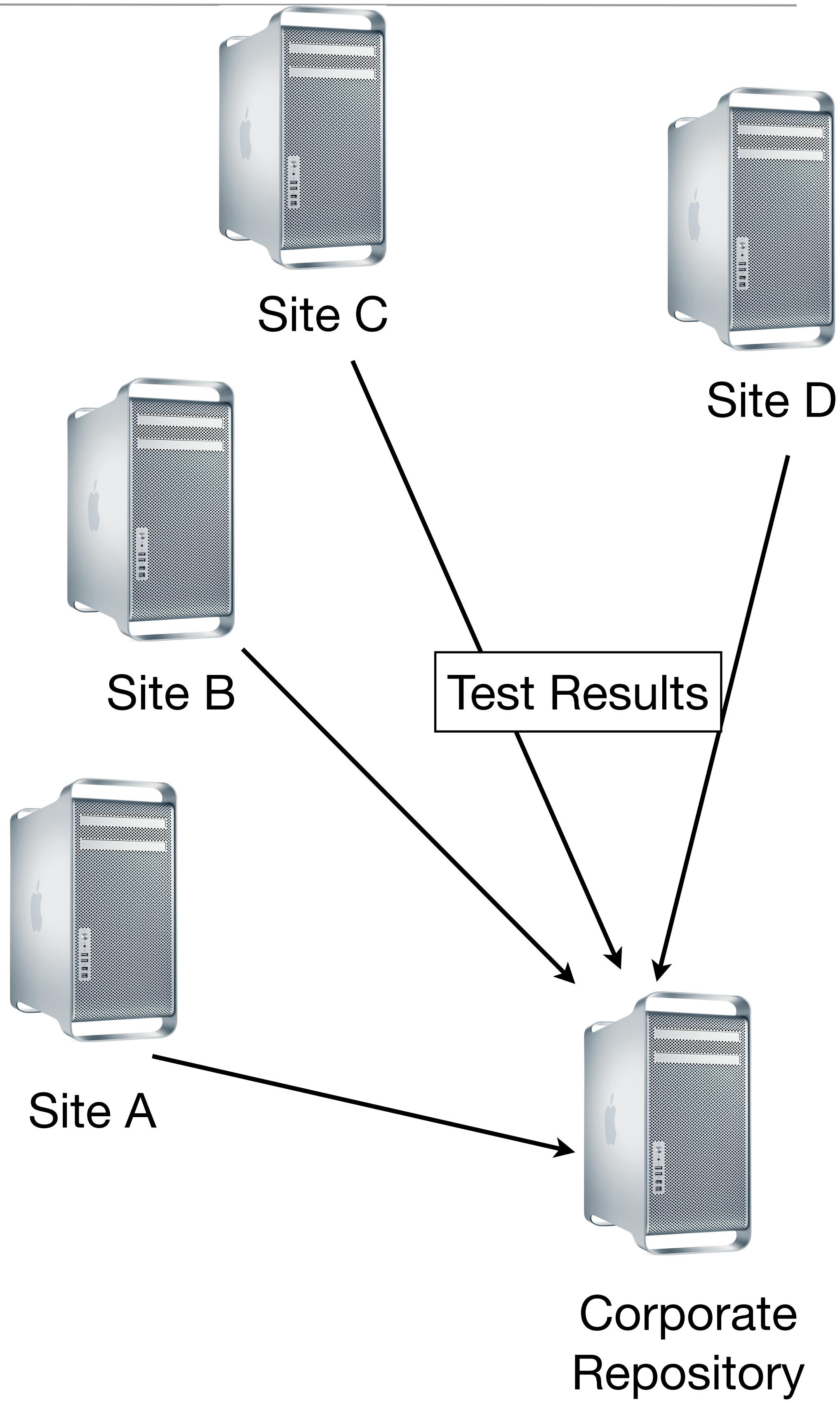
Data Distribution

- Scenario: Fan-out Distribution
- Financial Institution
  - Using large Operational Data Store
  - Collects transactional information
  - Common data from disparate systems
  - High volume of small transactions
- ER used to extract data from ODS
  - SQL based selection of data for replication
- One or many targets
  - Other operational systems
  - Mart sized databases for BI queries
  - Geographical and business divisions defined by use cases
- Easily administered



Data Distribution

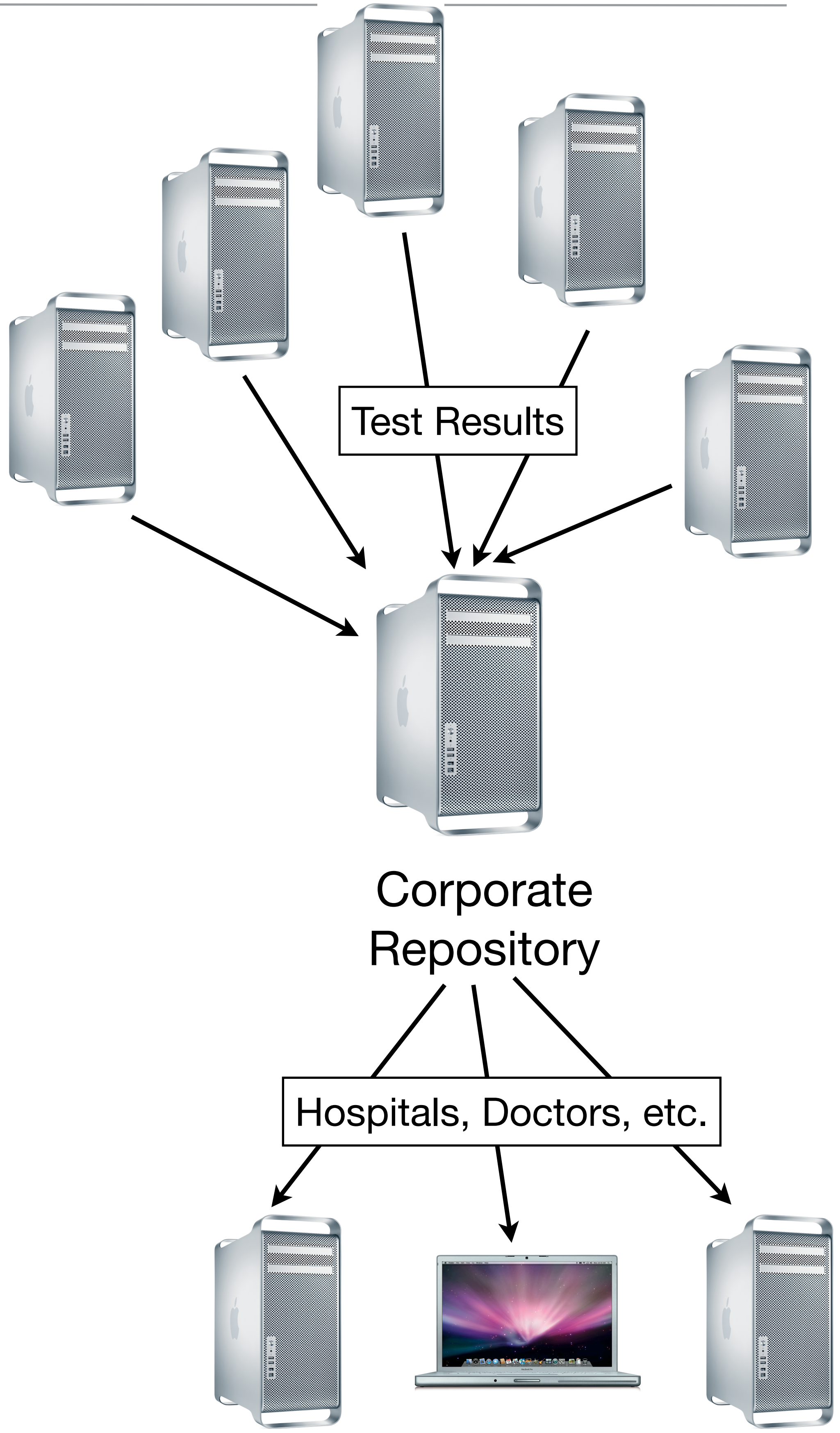
- Scenario: Fan-in for Consolidation
- World-wide drug testing company
  - Running clinical trials for different universities, pharmaceutical companies, researchers
- Conducting tests in remote areas, in areas where diseases occur – result in limited hardware and connectivity
- Data collected and aggregated into a composite view of testing, results
- Information obscured to protect identities of patients / participants to meet legal requirements
- Data-based research is conducted on complete, worldwide view of results



Data Distribution

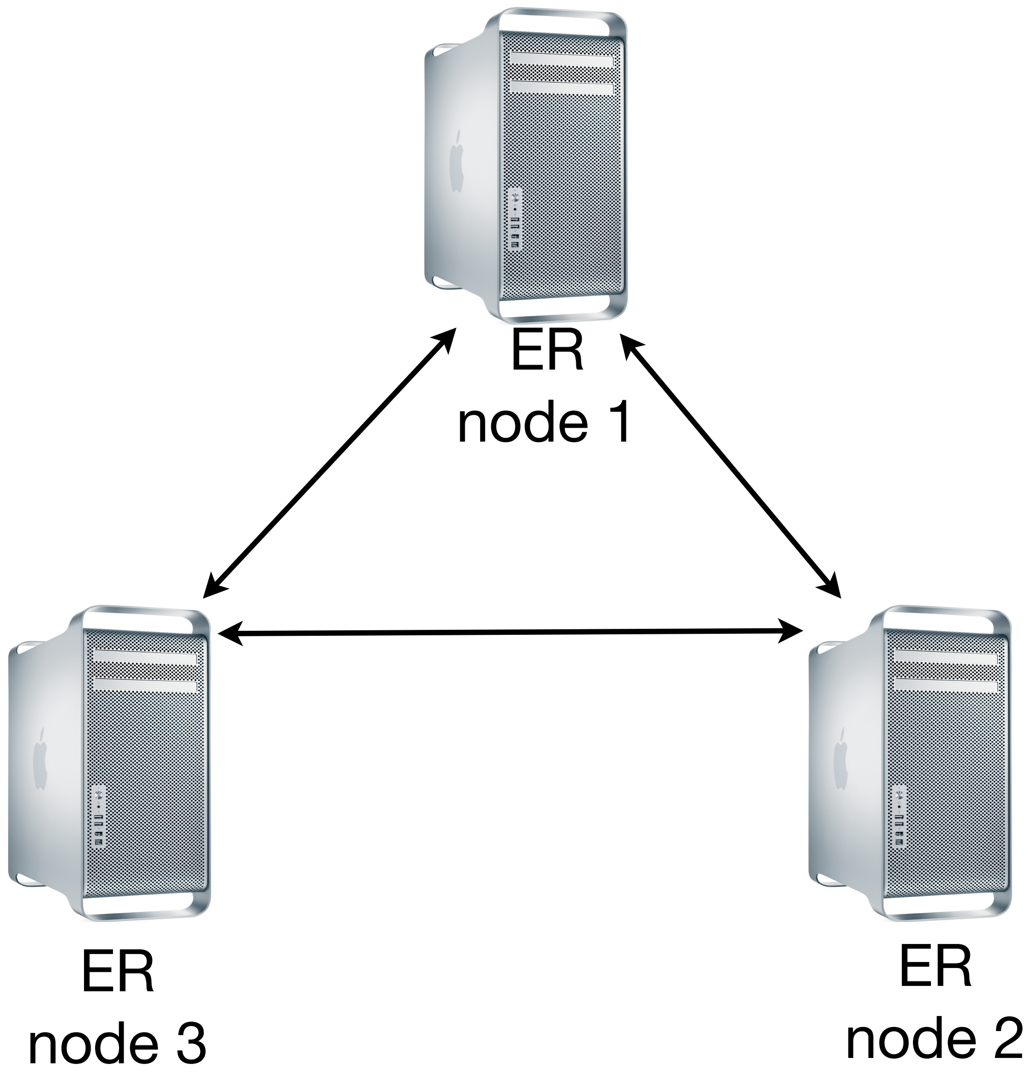
- Scenario: Fan-out and Fan-out (once the trial information is gathered)
- Data is disseminated in controlled fashion to
  - Doctors
  - Hospitals
  - Pharmaceutical Companies
  - Federal Reporting NIH and FDA
- Data is further refined to remove location, participants, and affiliate specific information
- Dissemination tends to be aggregated
- Grouping by affiliation becomes more important than time

Different targets can require different columns be included and removed



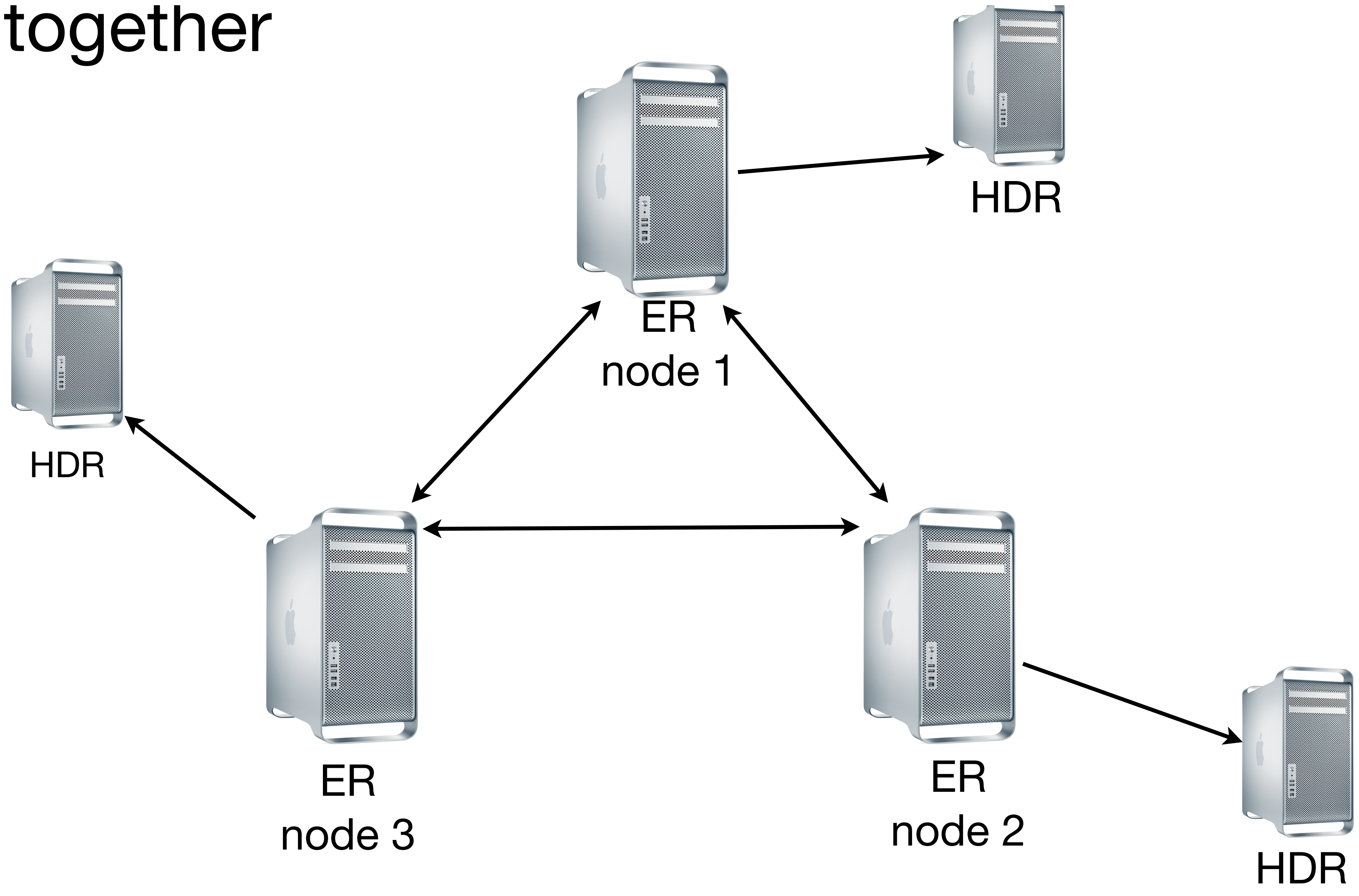
Putting it all together

- You can combine Enterprise Replication and HA/DR technologies together



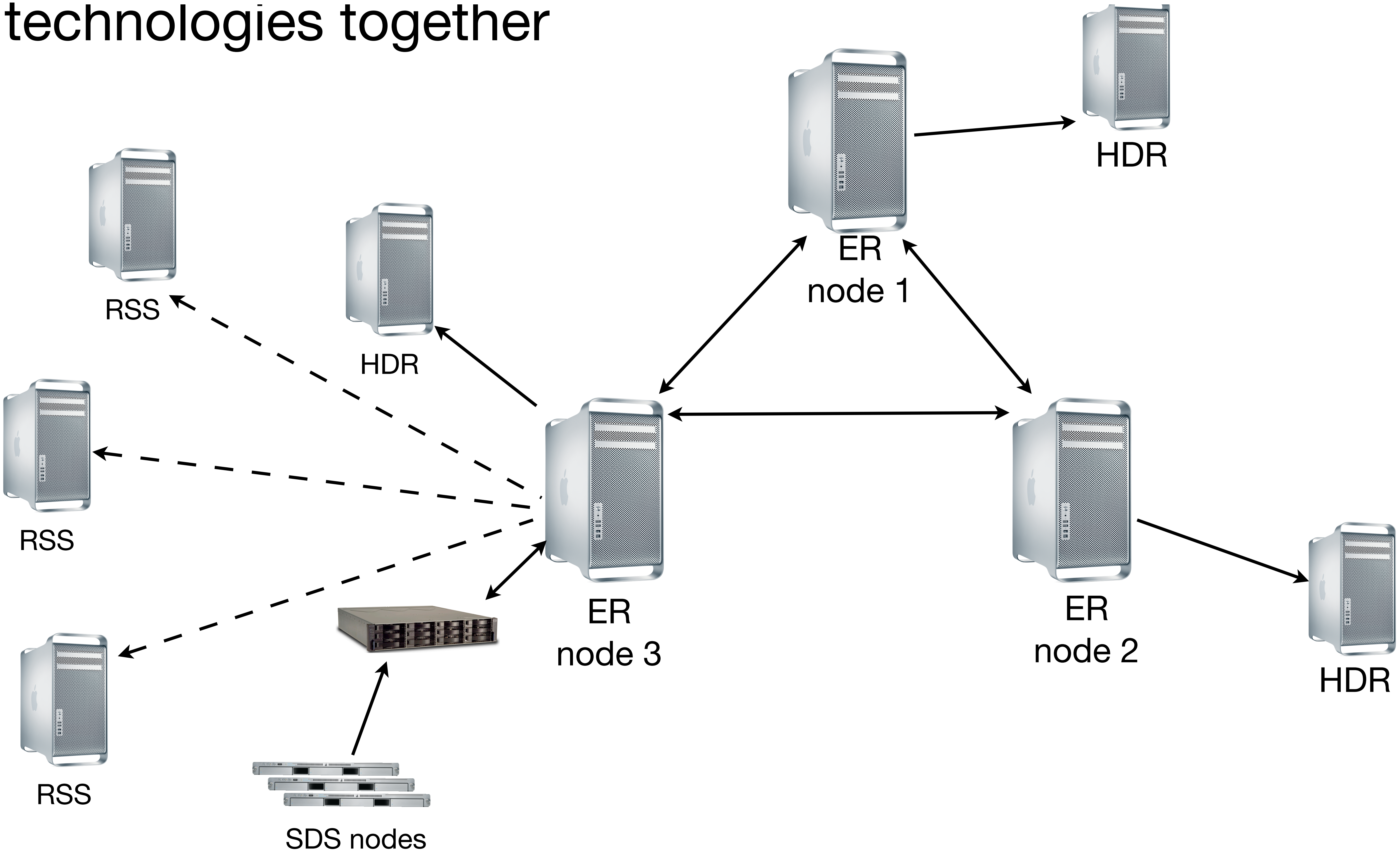
Putting it all together

- You can combine Enterprise Replication and HA/DR technologies together



Putting it all together

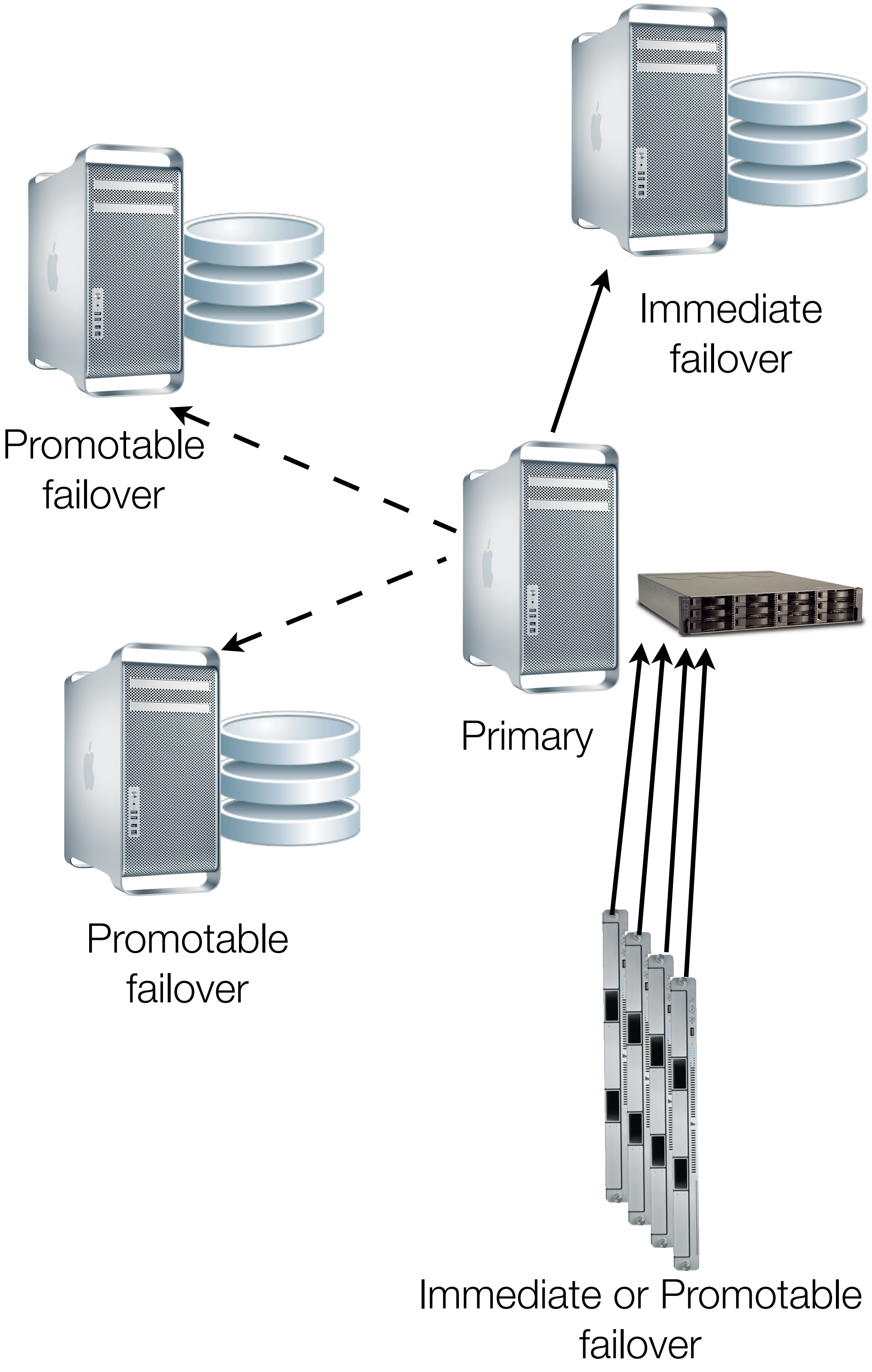
- You can combine Enterprise Replication and HA/DR technologies together



Review and Summary

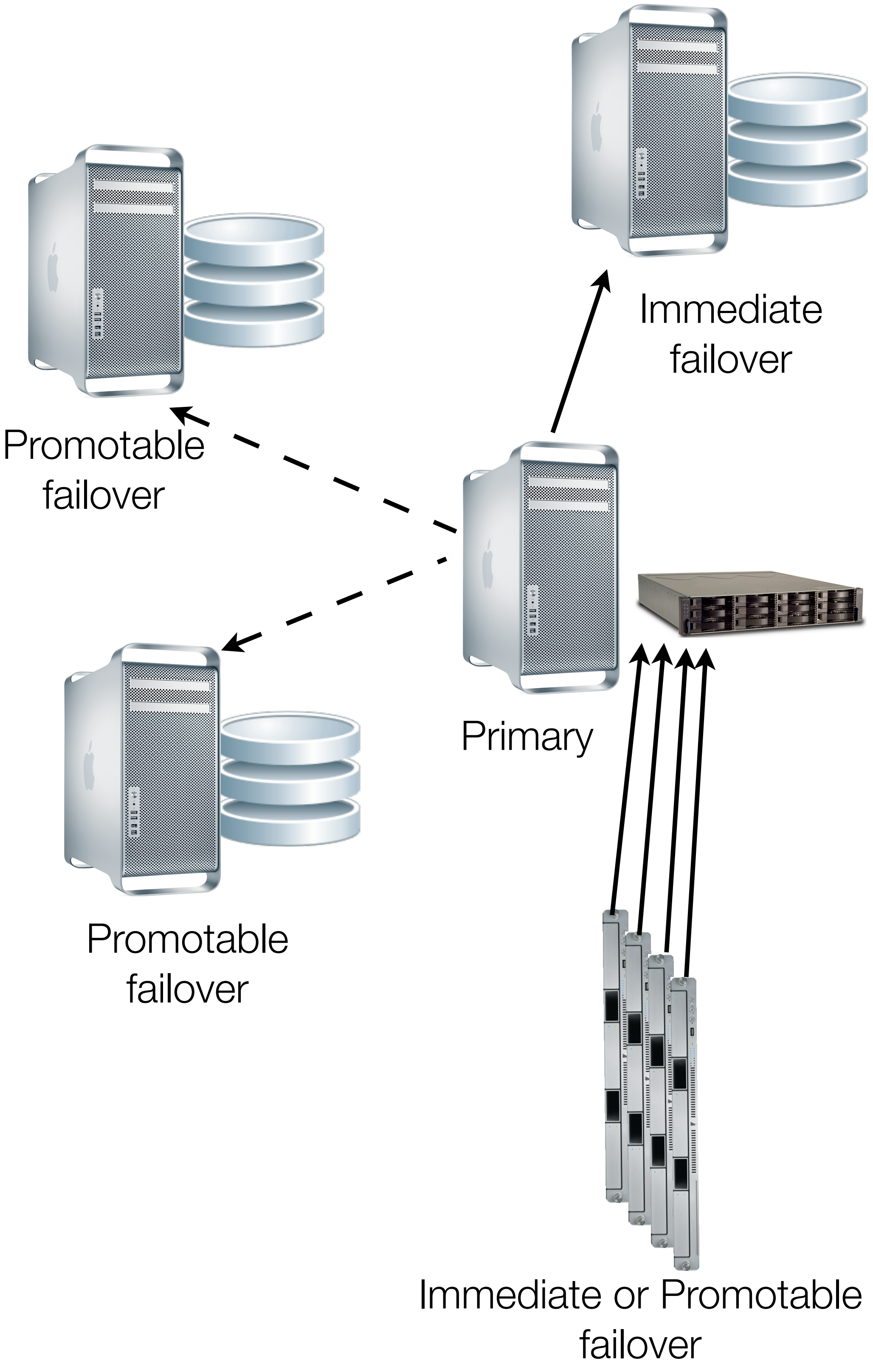
Review - Data Replication Characteristics

- Focuses on data server SPoF conditions to provide "always-on" capability
- One or more full copies of primary data server for transaction processing failover
  - Synchronous communication to immediate fail-over nodes
  - Asynchronous to promotable nodes with guaranteed data synch
- Flexible server types
  - Can provide protection against different SPoFs
    - Disk
    - Server
    - Network
    - Location



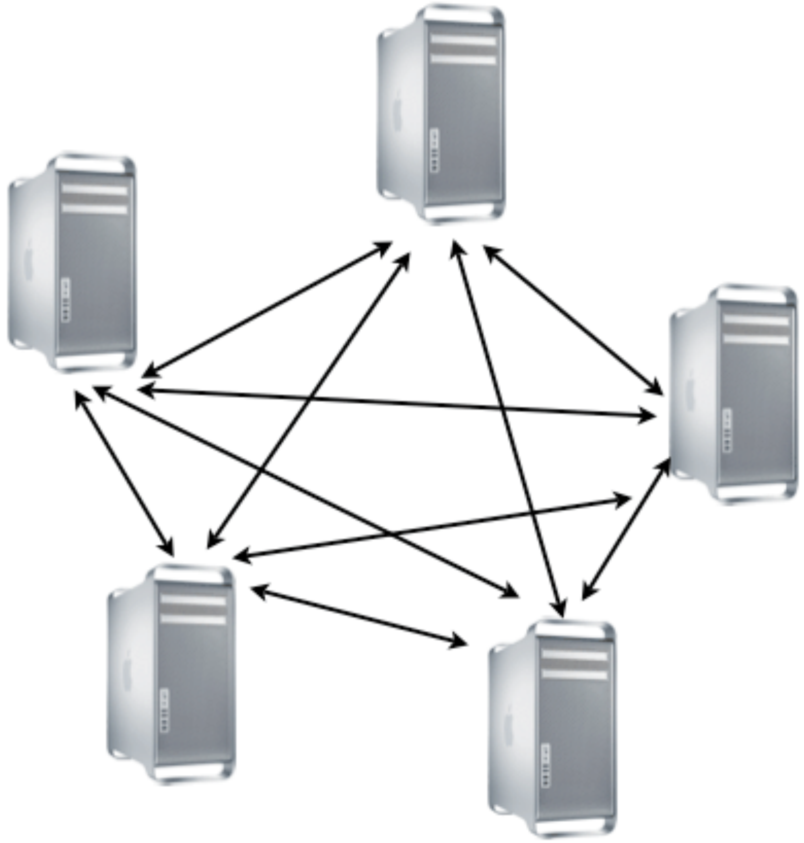
Review - Data Replication Characteristics

- Can support active/active updates
    - full application support anywhere in the cluster
        - no code changes required to deploy
- Integrated, not add-on
- Compatible and interoperable with data distribution technology
- Can provide secure data communication between servers



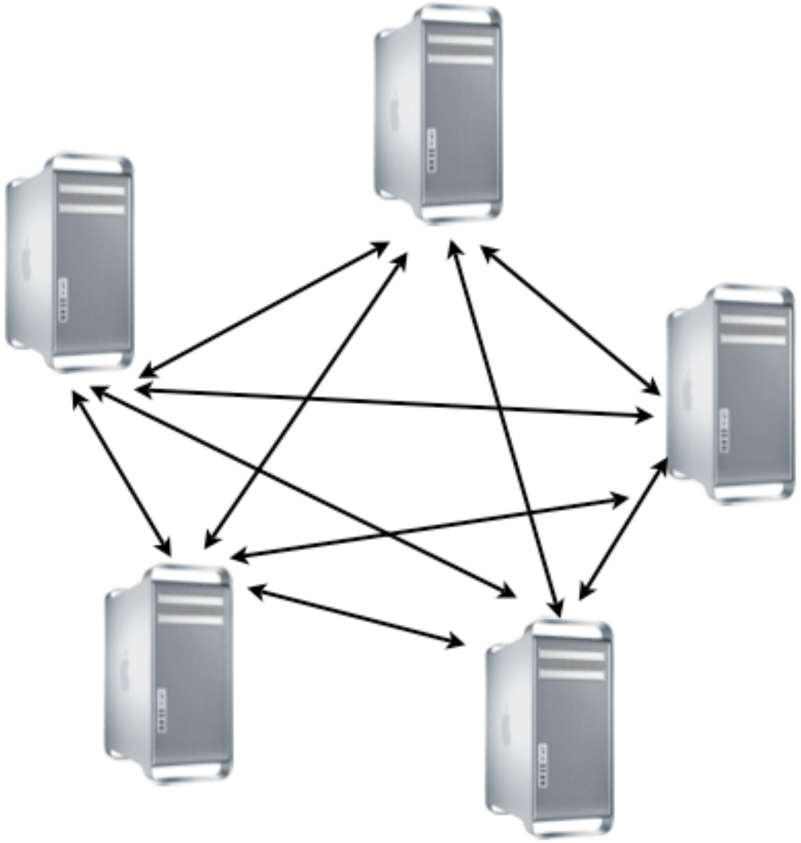
Review - Data Distribution Characteristics

- Asynchronous data replication
  - You define the interval when data is place on the outbound queues
  - Does not require specialized hardware or networking infrastructure
- Provides transaction consistency
  - Data to be replicated is stored in stable queues for transfer to defined targets
- Uses transaction log discovery for data to be transferred
  - Eliminates conflict with transaction processing



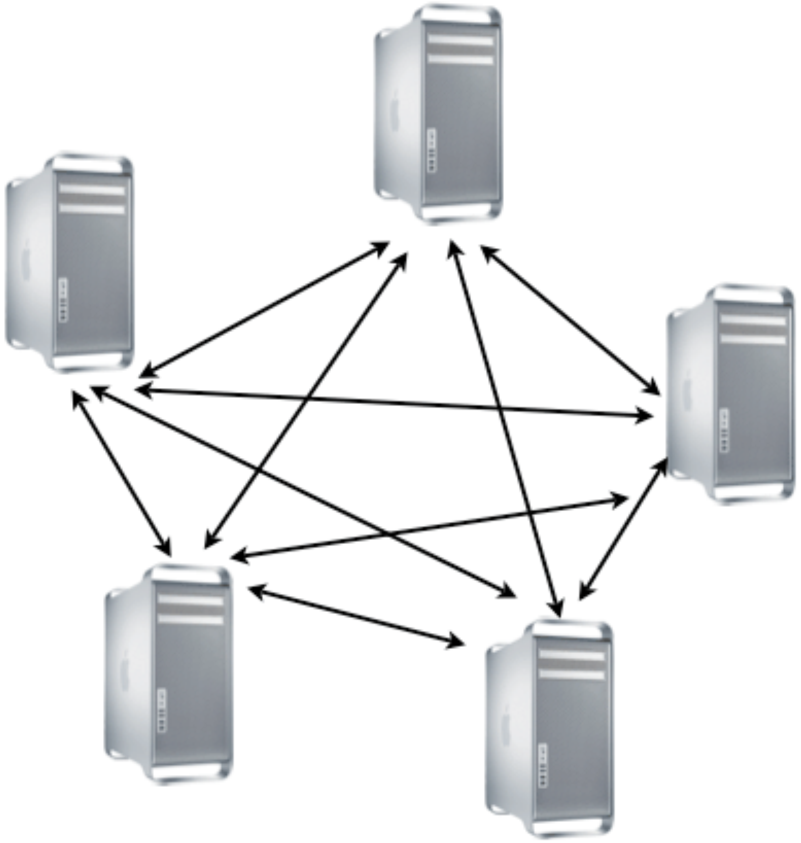
Review - Data Distribution Characteristics

- Flexible
  - You can define the granularity of data to be replicated
  - You can define the sources and targets of replication
  - Is heterogeneous in terms of platform and data server version support within a cluster
    - All supported O/S ports can replicate to each other
    - All current versions of the data server can replicate to / from each other
- Easy to administer
  - Centralized administration through Master Replicates, templates, remote connections and more



Review - Data Distribution Characteristics

- Allows for minimal downtime in critical applications
  - Uptime of Overall system is in 99.999%
- Provides for rolling upgrades
  - Updates to application performed on target
  - Access redirected to target, becomes new source
  - Initial source is updated and synched with new source
  - Source / target either reversed or left as is
- Complies with stringent Service Level Agreements
  - Definition dependent
    - If the local server is unavailable, a replicated copy of the data may be accessed through a remote server.



Summary

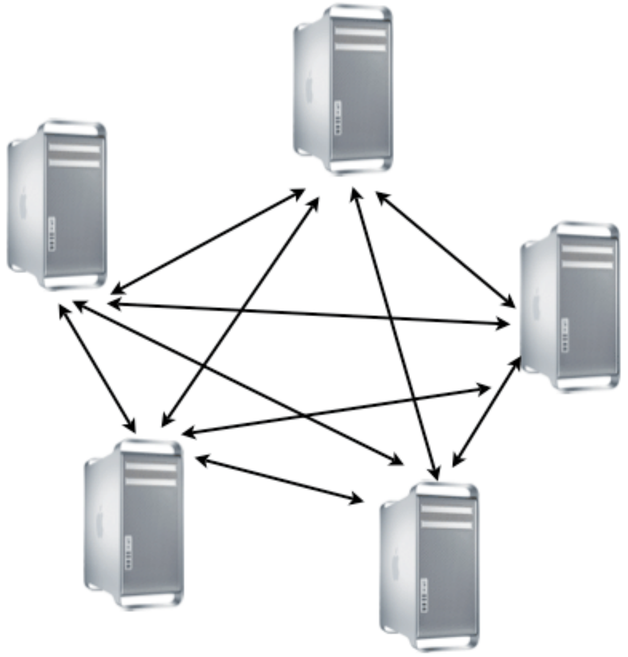
- What’s the right answer for you and your environment?

Summary

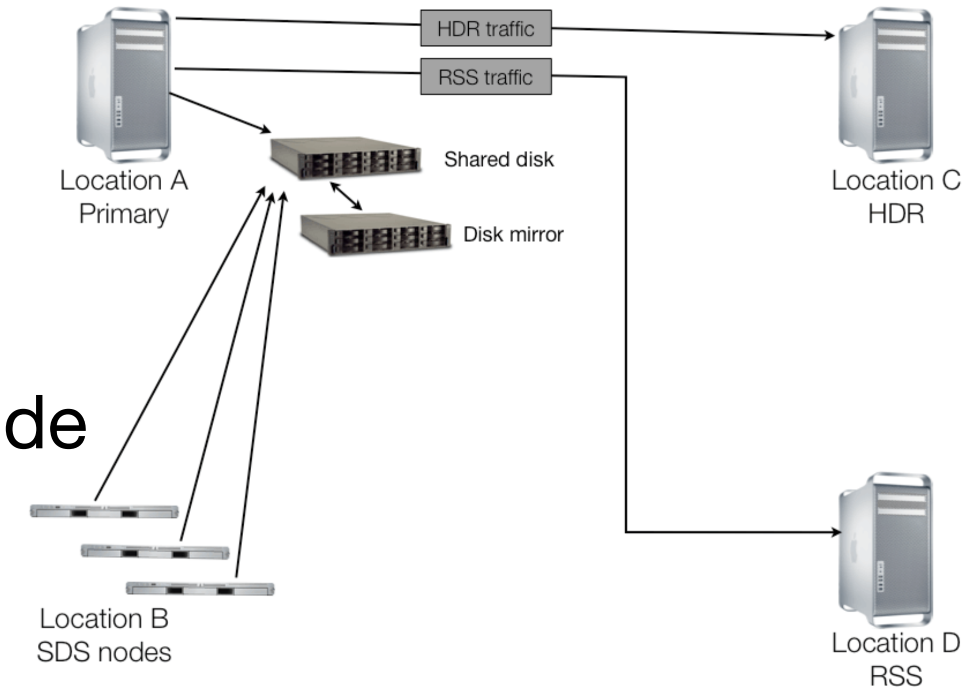
- What's the right answer for you and your environment?
  - It depends!!!
    - What is/are your most important business needs?
    - How much time are you prepared to invest in identifying potential SPoFs?
    - How much time and money are you willing to invest to mitigate against those SPoFs vs the risk of not taking any action?

Summary

- Enterprise Replication gives you
  - Flexibility and options
    - Support for multiple O/S's within the cluster
    - Different data distribution topologies
    - Support for multiple data server versions
  - Downside
    - More complicated to set up and administer



- Data Replication server gives you
  - Simplicity
    - Complete copy in one operation
    - Better transactional security
    - Simpler administration and maintenance
  - Downside
    - No options, it's all or nothing but can provide multiple layers
    - Homogeneous



Where can you get more information?

- The software:
  - http://www-01.ibm.com/software/data/informix/mac/ OR
  - http://www.apple.com/downloads/macosx/development_tools/ibminformixdynamicserverdeveloperedition.html

- Quick start guide:
  - http://www.xmission.com/~dbaresrc/Intro_IDS_on_Mac_v3.pdf

- Application development talk:
  - http://www.ibm.com/developerworks/db2/events/macosx.html

- Excellent book on using the data server:
  - *Administering Informix Dynamic Server, Building the Foundation*

    ISBN-10: 158347076X ISBN-13: 978-1583470763

Questions??

Thank you!

Session: IT864
Carlton Doe
cdoe@us.ibm.com

http://macpres09.shownets.net