Customer Releasable

SAP ASE Always-On Option:
Solution Overview & Update

Jeff Tallman jeff.tallman@sap.com
SAP ASE Product Management

Agenda

The State of HA

- Drivers for Continuous Availability
- Common HA Solutions
  - ✓OS Clusters
  - ✓Shared Disk Clusters
  - ✓Asynchronous Replication
  - ✓HADR Clusters

ASE Always-On

- Architecture & components
- Performance/Sizing
- Failovers
- Upgrades w/ Zero Downtime
- Comparing the solutions

Always-On Future Development

Understanding HADR Clusters

Competition & Trends

Technology Trends

Most DBMS HA solutions are moving to HADR clusters using streaming replication

Rationale

- Hardware, OS & Storage agnostic
  - ✓No need for shared disk
  - ✓No need for OS HA services nor special storage protocols
- Much more supportable in cloud deployments (private or public)
- Supports in-memory processing techniques vs. shared disk clusters (SDC)

Technology du-jour for "NewSQL"

- MemSQL, Postgres, et al.

Only vendors staying with HW/OS implementations are Oracle & IBM

- But then they have a vested interest in HW
- Both also have HADR clusters

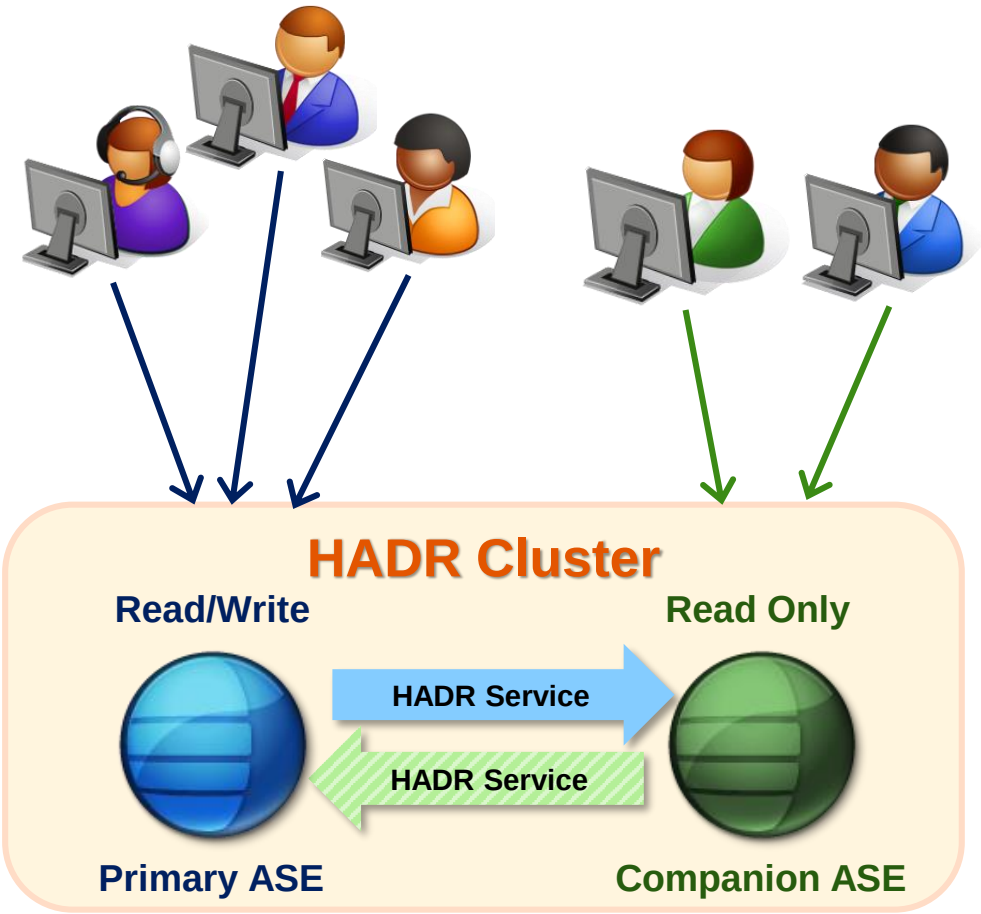
HADR Clusters (e.g. ASE Always On)

Core technology

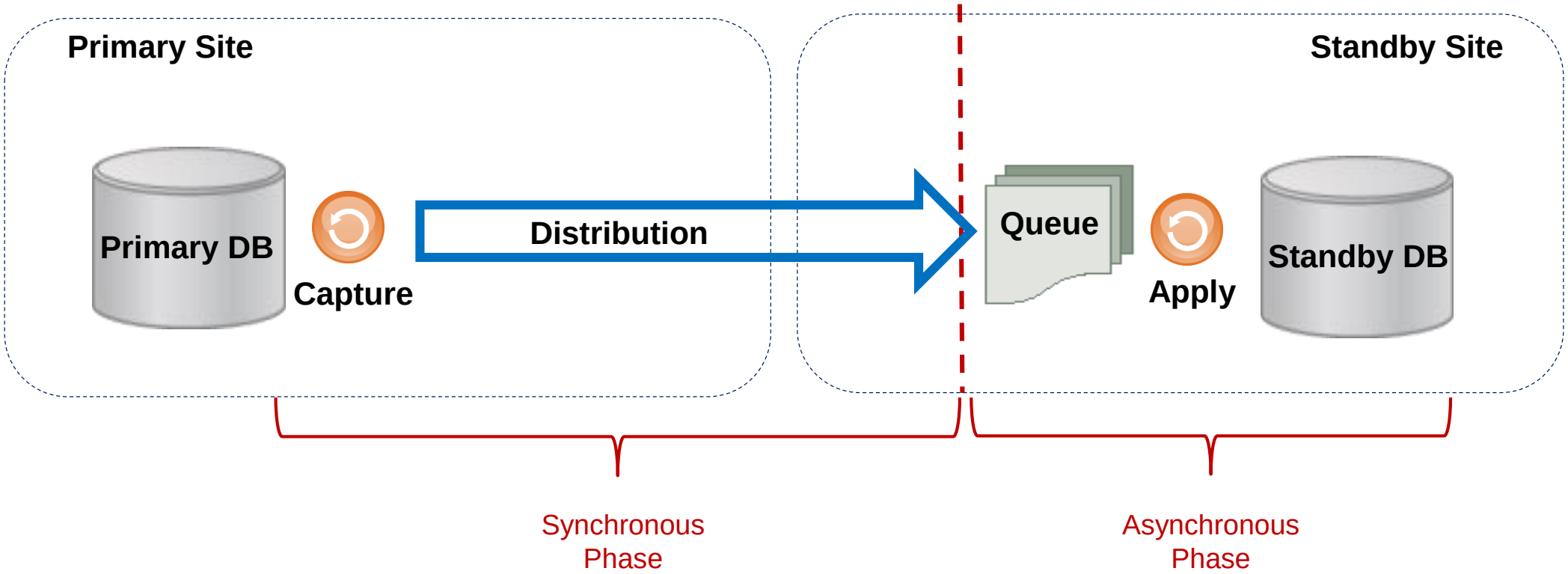
- Log record-based *streaming* data replication
- Usually supports sync, near-sync and async modes
  - ✓Sync → commit on recv persistence/async apply
  - ✓Near-Sync → commit on recv receipt/async apply
  - ✓Async → commit immediately/async send & apply

2 Classes Types

- Physical
  - ✓Log records are applied as log records/log replay mode
- Logical
  - ✓Log records are translated into SQL for apply



HADR Fundamental Data Synchronization



The 2 choices: Physical vs. Logical Apply

Physical Apply

- Copies log records by copying log blocks physically and transmitting binary page/block image to remote copies
- Remote system simply re-applies log image
- Advantages
  - ✓No problems with large transactions, SQL handling
- Disadvantages
  - ✓Cannot handle schema changes (application upgrade availability scenarios)
  - ✓If log block images vs. log records, still could incur log page corruptions
  - ✓Database must be page-for-page mirror image
  - ✓The above blocks DB maintenance on standby

Used by:

- Everyone else (Oracle, IBM, MSSQL)

Logical Apply

- Copies & batches up only necessary log records
  - ✓E.g. can skip index inserts, allocation records, etc.
- Remote system applies via SQL language
- Advantages
  - ✓Lower bandwidth
  - ✓Don't need to replicate reorg actions, etc.
  - ✓Database does not have to be completely page for page mirror image
    - – Allows reorgs, update stats, etc. on standby
  - ✓Can handle schema changes (provided transport supports transformation capabilities)
- Disadvantage
  - ✓Large txns, long running txns, SQL issues
  - ✓Performance can take some tuning

Used by:

- ASE Always-On, Oracle DataGuard (Logical Apply mode)

The Competition: Oracle DataGuard & IBM DB2 HADR

No one is TRULY Synchronous

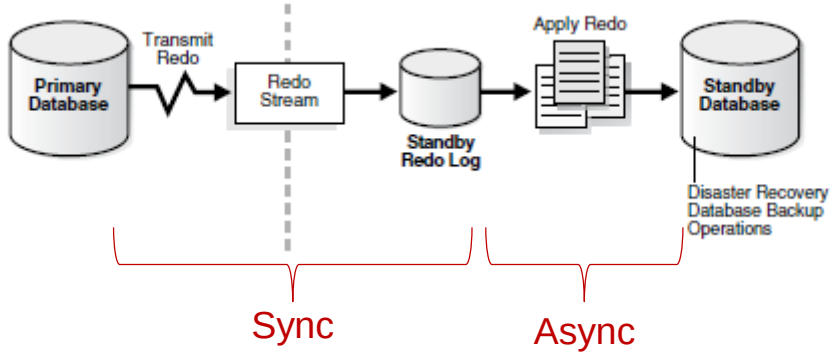
Oracle DataGuard

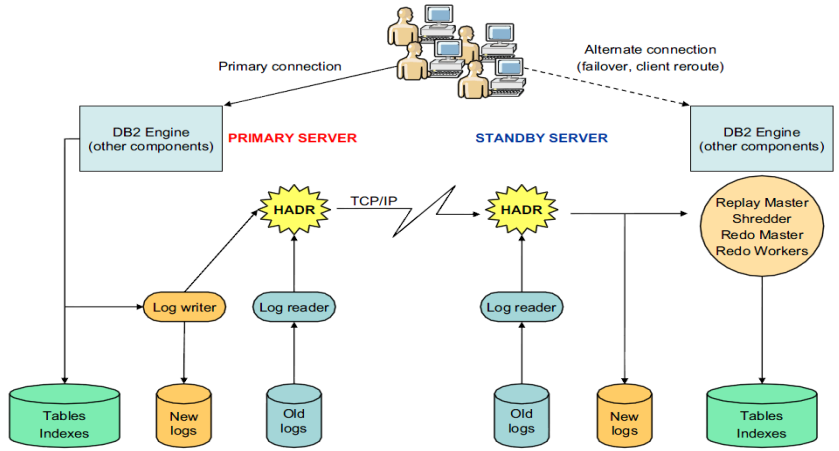
- Synchronous to Remote Redo Log
- Asynchronous Redo Apply
  - ✓Real-Time Apply is supposed to reduce the latency
  - ✓Applies by reading redo log from received buffers vs. rescanning redo log from disk

IBM DB2 HADR

- Synchronous to remote txn in-memory buffer
- Async apply from in-memory buffer
- If buffer fills….primary also suspends
  - ✓You can tune buffer size (DB2_HADR_BUF_SIZE)

Source: Oracle® Data Guard Concepts and Administration 12c Release 1 (12.1), September 2014





(source: High Availability and Disaster Recovery Options for DB2 for Linux, UNIX, and Windows; IBM Redbooks; October 2012)

Common Issues with HADR Clusters

Applying at standby is slower than primary

- One common reason is that at primary there is a lot of effort in logging to speed txn throughput
  - ✓Hence txn rollbacks tend to be slower than txn commits.
  - ✓Group commits, ULC caches/PLC queues, etc.
- Another reason is that user actions (e.g. query) prefetches page to cache
  - ✓At standby, to reapply the insert, often requires a physical read of data & index pages
- Another reason is that users on standby running reports may contend with replay
- Use of in-memory queues tends to limit surge capacity and cause primary outages (e.g. DB2)
- Oracle has attempted to work around this lately by supporting parallel threads in sender/receiver as well as out of order commit sequencing

Biggest issues

- Due to page-for-page mirror image, can't run reorgs, update stats, alternative indexing on standby
  - ✓Would cause pages to change or move wrecking page mirror image
  - ✓Result is primary is inflicted with memory and cpu requirement of DBMS maintenance
- No real integration with enterprise data replication for other topologies
  - ✓They have pictures in the book suggesting how to do it, but largely, it becomes a science project to implement
    - Commonly point to replicating from standby - but what happens in a failover - how do you fail replication over to new standby….if it is available
  - ✓Because of the typical ASE customer has a large SRS topology, this is one of the biggest challenges we needed to address

ASE 16sp02 Always-On Option

ASE High Availability + Disaster Recovery

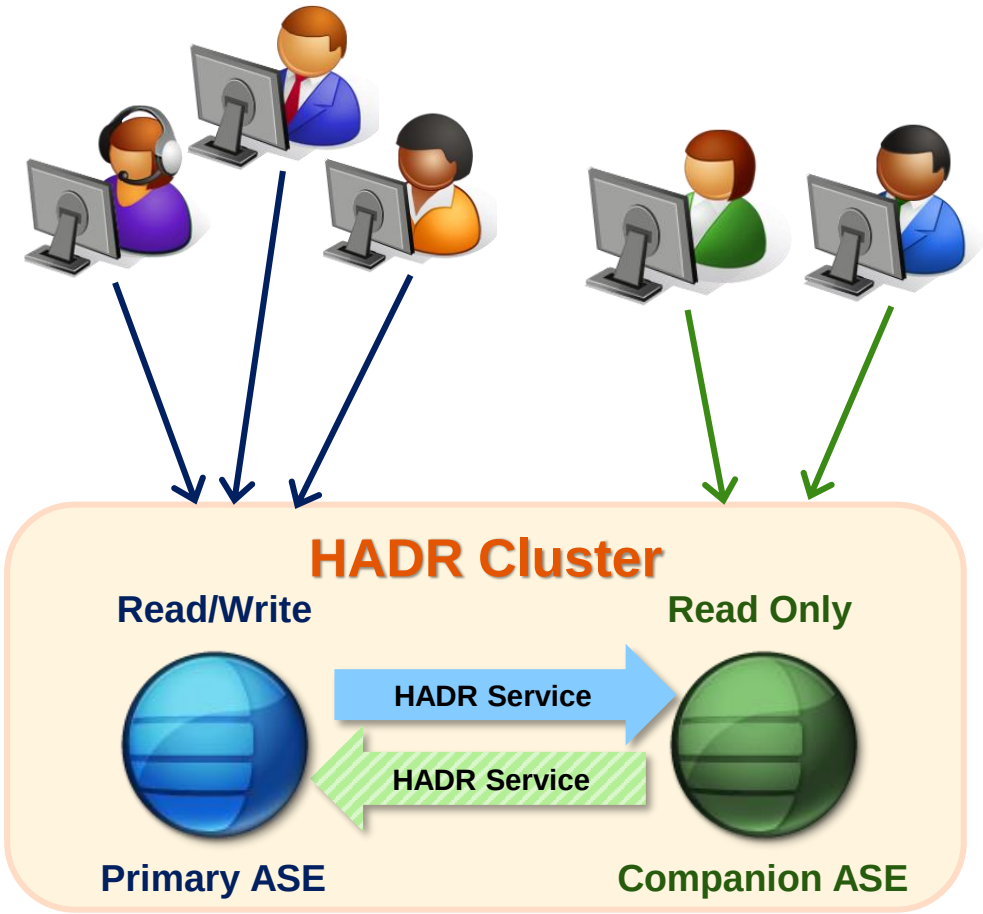
Always-On

HADR Cluster

- Single cluster is limited to 2 nodes
  - ✓Additional standby nodes via external replication
- Log-based Logical Replication Based
  - ✓Synchronous, Near-Synchronous, Asynchronous
  - ✓Zero Data Loss in Synch (RPO=0)
- Fast failover (<2 minutes normally)
  - ✓Planned failovers <1 minute
- GUI (ASE Cockpit – replaces SCC)

Capabilities

- Automated fault detection
- Automated transparent client failover
  - ✓Planned and unplanned failover support
- Companion can be read-only for reporting
- Zero-down time major upgrades
- Cloud friendly deployment
  - ✓vs. OS & Shared Disk Clusters
- Supports In-Memory XOLTP optimizations in ASE



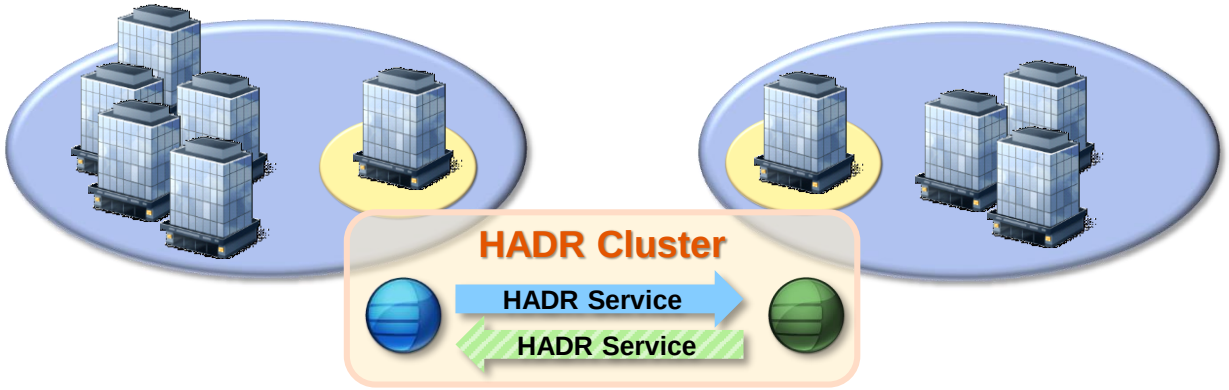
Two Common Installation Architectures

Within same datacenter → HA Focus

- One of the biggest outage reasons within datacenters is individual rack failures or entire row failures due to localized power/switch failure
  - ✓This often takes out clusters as commonly the nodes of the clusters are within the same rack/row to shorten interconnect distance
  - ✓…or shared disk SAN is impacted which takes out entire cluster
- HADR allows two different independent systems on opposite ends of datacenter

Between two datacenters → HA + DR

- Must be short distance due to synchronous replication
  - ✓e.g. similar to disk replication distances
  - ✓The higher speed the link, the longer the distance
  - ✓Needs to be at least 1Gbs or higher
- Bandwidth between sites also needs to support user activity if offloading reporting



Always-on HADR cluster architecture & components

Primary & Companion ASE's (ASE 16sp02+)

- HADR mode enabled (e.g. soft quiesce support, etc.)

Primary & Companion SRS

- HADR capable (SRS 15.7.1sp303+)
- Pre-tuned for high volume/low latency

Primary & Companion RMA

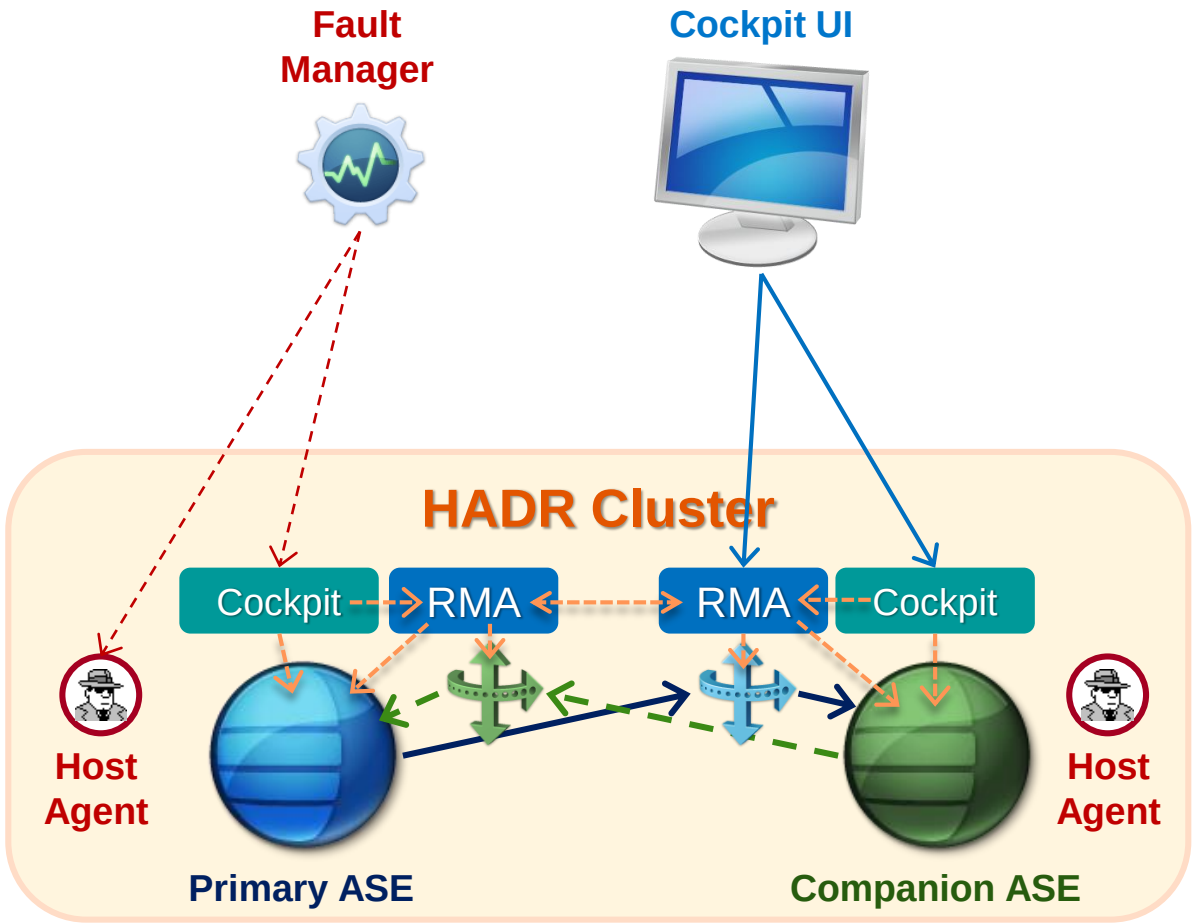
- Provides simplification installation & operations
- E.g. sap_materialize, sap_failover

Primary & Companion Cockpit

- Server side agent has logic
  - ✓ Server-sides supports stop/start/errorlog scan
  - ✓ Issues commands to RMA for HADR operations
- Client UI is web browser

Fault Manager

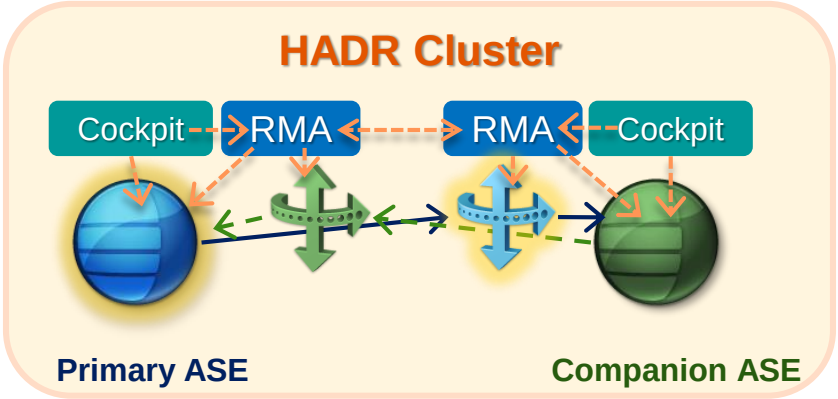
- Installed on 3rd node
- Uses ASE Cockpit and SAP Host Agent to detect and control failover



Old friends….New features

ASE 16sp02 support for HADR

- HADR (virtual) cluster aware w/o HW or quorum devices
  - ✓ Knows which other nodes are in the cluster & state (primary/standby)
- Supports for soft quiesce
  - ✓ Allows zero-downtime planned failovers vs. the typical brief stopping of applications
- Supports client failover & login redirection
  - ✓ None privileged users connecting to the standby are transparently redirected to the primary
- New failover API
  - ✓ Provides state transition messages during planned failovers
- HADR permissions and roles for limiting standby access and HADR admin



SRS 15.7.1sp30x support for HADR

- CI Mode RepAgent with synchronous transfer
  - ✓ New high speed queue for CI mode RepAgent
- Pre-tuned out of the box for high-speed/low latency
  - ✓ All the magic go faster features enabled and memory caches pre-tuned for performance
  - ✓ May need minor tweaking for large txns/batch
  - ✓ Possible future T-shirt sizing pre-tuning to eliminate need to tweak
- Large transaction mode
  - ✓ Allows large txns to start being applied at the standby prior to SRS seeing the commit from primary
  - ✓ Reduces latency caused by waiting for the commit

Recent Developments

ASE 16sp02 pl05+

3 Nodes (HA+DR, HADR w/ 2nd time delayed DR)

3rd Node: HA+DR

- HADR cluster for HA within the datacenter
- DR coverage is purely disaster recovery
  - ✓Not intended for long term usage
  - ✓Apps may have degraded capabilities
    - e.g. no HA, no reporting offload, etc.

3rd Node: Delayed

- HADR primarily for HA or HADR
- 3rd node primarily to protect against errant transactions
  - ✓Assumption is that errant transaction can be spotted and blocked from 3rd node within the delay time frame
  - ✓Data values would be extracted from 3rd node and re-injected into primary vs. failover to 3rd node
    - Attempts to skip the errant transaction could result in further issues due to subsequent transactions which would prolong the failover to the 3rd node if intent was to failover to it once it was back in sync time wise

Dual HADR clusters: Long Distance (HA+DR)^2

1st Goal – survive multiple failovers

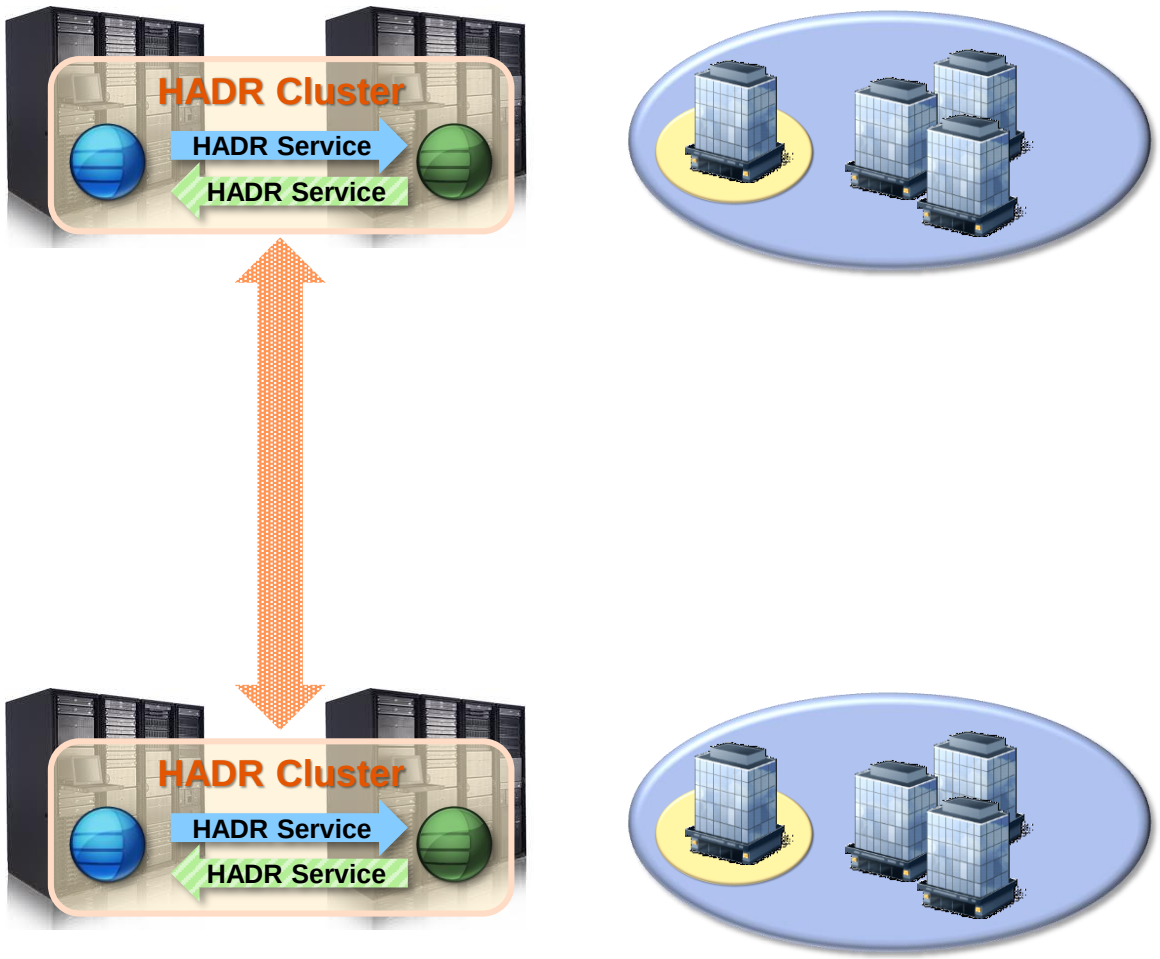
- DC failure
- Subsequent HW or SW failure in DR site
- In other words, full HA capabilities and report offloading is maintained even if DC fails

2nd Goal – R/O scale out

- Users in second business location have reduced latency for reading data/reduces bandwidth requirements to primary site
- Spread out reporting across nodes
  - ✓Current OLTP reports and historical often have competing resource requirements
  - ✓DC 1 Standby → OLTP reports (e.g. order status checks)
  - ✓DC 2 node 1 → EOM/EOY reports
  - ✓DC 2 node 2 → adhoc/historical reports

3rd Goal – Bi-directional Apps

- Site autonomy/better HW utilization



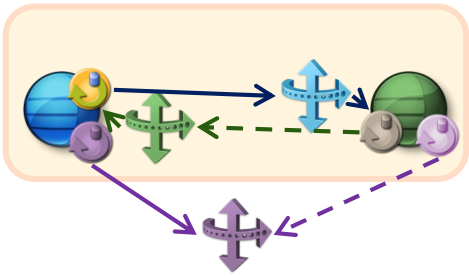
External Replication Support

Split into two problems

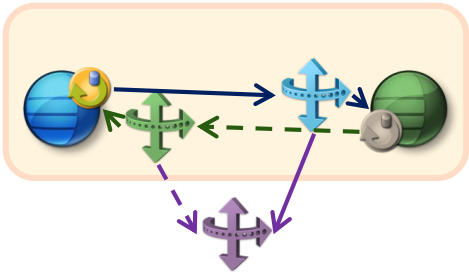
- HADR as a source
- HADR as a target

HADR as Source: 4 Solutions

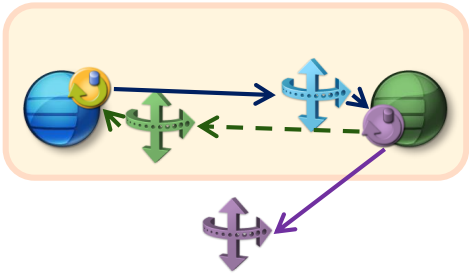
- MPR RepAgent
  - ✓If sync, slows down HADR
  - ✓If async, susceptible to data loss & impacts STP
  - ✓Failover would result in different OQID's
- Route from HADR
  - ✓Failover coordination issues
    - Wait for route queue to drain before failover or …
    - Wait for route queue to drain before starting failover RepAgent
  - ✓Different OQID issues with Route
- Re-replicate from Standby
  - ✓Competition uses this
  - ✓Issue is on failover - either MPR or manually switch RepAgent to old primary….if available
    - Coordination issue - wait until log read before switch (remember we have different OQIDs from other source)???
- Something else (option 4)



MPR RepAgent



HADR Route



Re-Replicate fm Standby

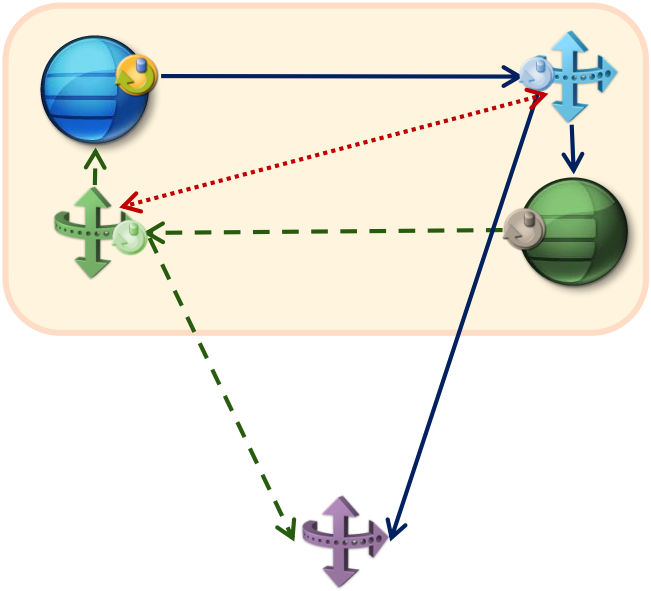
External Replication Support: HADR as a source

CI RepAgent internal to SRS (SRS 15.7.1 sp305+)

- Would scan log records from SPQ
- Connects to external SRS as a database

Advantages

- Zero Data Loss
- Looks to external as ASE RepAgent
  - ✓Preserves existing topologies with no need to drop/recreate repdefs
- Reading inbound queue too late due to RepDef normalization already missed



Challenges

- On failover, HADR needs to switch RepAgents and tell SRS to rs_zeroltm

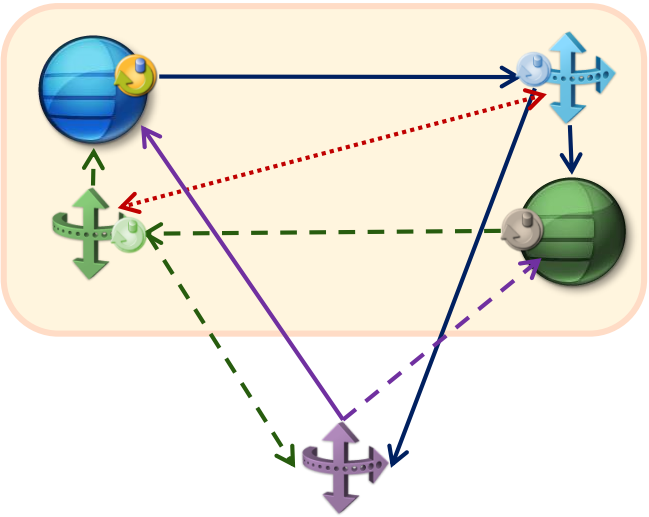
External Replication Support: HADR as a target

On Surface, this appears easy

- DSI has been HA aware since SRS 12.0

Challenges

- We want external SRS DSI's to failover but not the HADR DSI's
- We don't want to confuse rs_lastcommit, etc.
- Need to avoid cyclic replication when HADR is both source & target



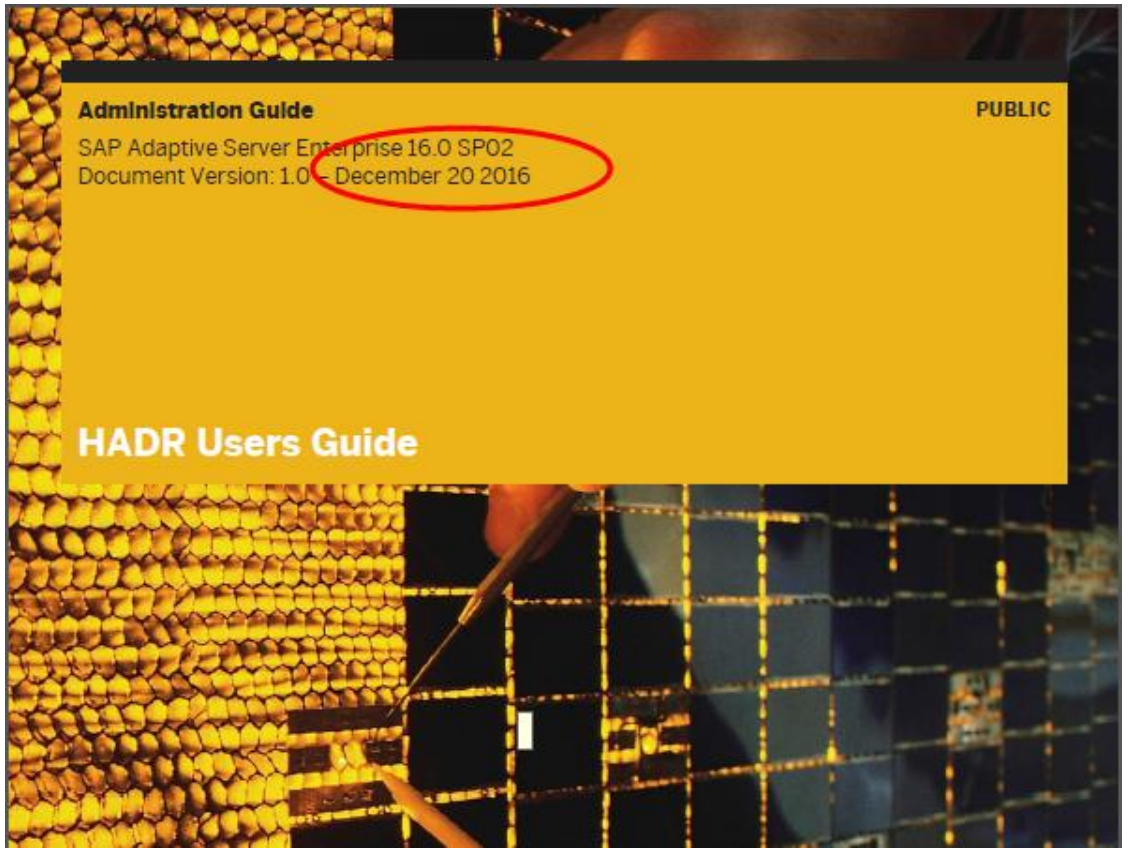
Proposed Solution

- External SRS connects as different maintenance user vs. DR_maint
  - ✓ Separate set of rs_lastcommit tables…e.g. dbmaint.rs_lastcommit
- External SRS connection doesn't have HADR privileges
  - ✓ So it would failover with other connections
  - ✓ Would need replication_role due to column encryption, materialized columns, etc. replication
- SPQ RepAgent can filter out external maint user txns as it does today
- HADR RepAgents run in WS mode to pick up and fwd warm standby txns to other HADR nodes

External Replication Support

Key Assumptions

- If replication already exists, no need to tear down implementation to implement Always-On for any single node
- HADR Users Guide PL05 has details on how to setup/configure in Chapter 5
  - ✓Sections 5.2.1 & 5.2.2 setting up new
  - ✓Section 5.2.4 migrating existing systems



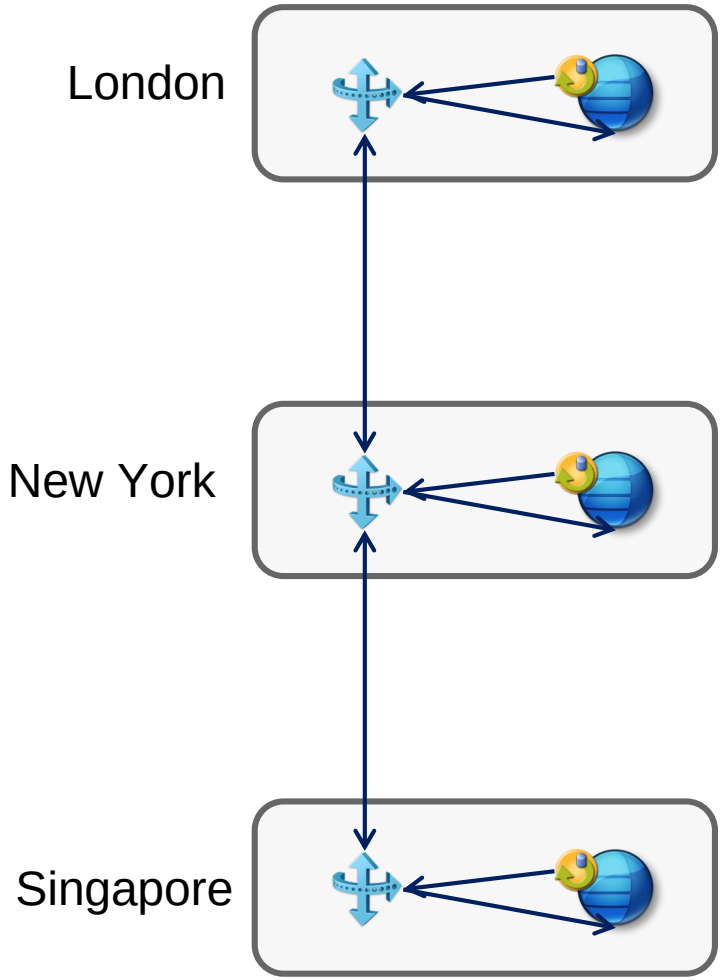
Some notes/restrictions

- Will require some downtime to avoid data loss between HADR cluster and external nodes
- May need to change some corp ‘standards’ (such as aliasing maintuser to dbo)

A Sample Walkthrough

Key Assumptions

- If replication already exists, between 3 sites
- We want to implement Always-On in NYC first (and then may also in London)
- ASE is already ASE 16sp02 PL05+



High-level Steps To Enable External Replication (1)

Upgrade external SRS to 15.7.1 sp305+

- Assumption is ASE is already sp02 pl05

Suspend DSI into target

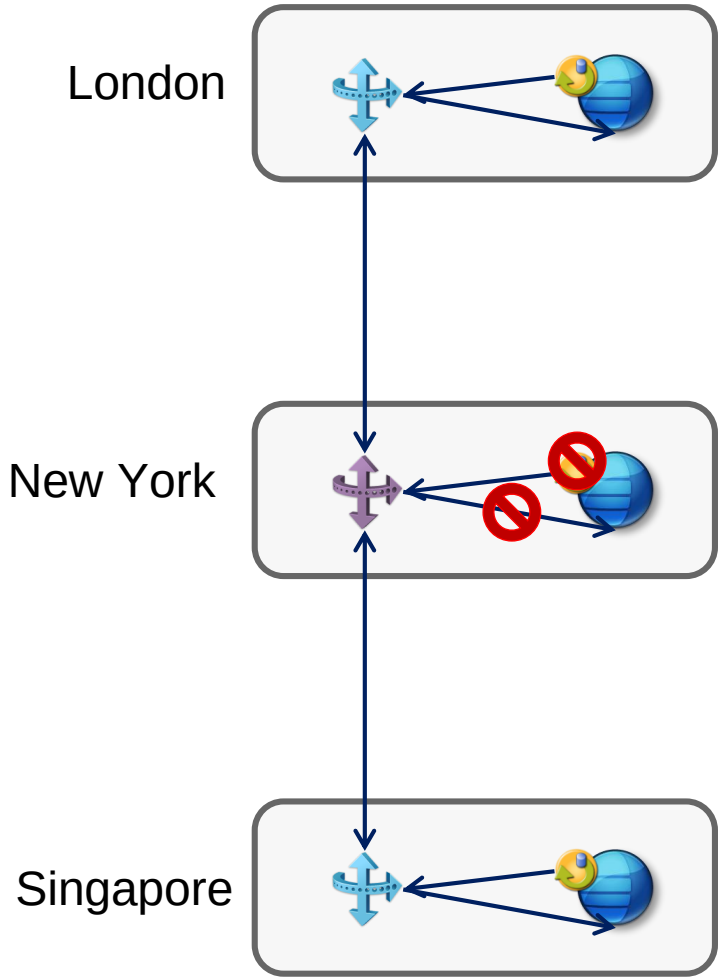
- Ideally you will want to do this during a lull in upstream/downstream activity so that the build up in OBQ is minimized

Teardown the Source RepAgent

- Remove secondary truncation point

Fix the maintuser

- Unalias as dbo
- Revoke sa_role, hadr privileges from maintuser if previously set



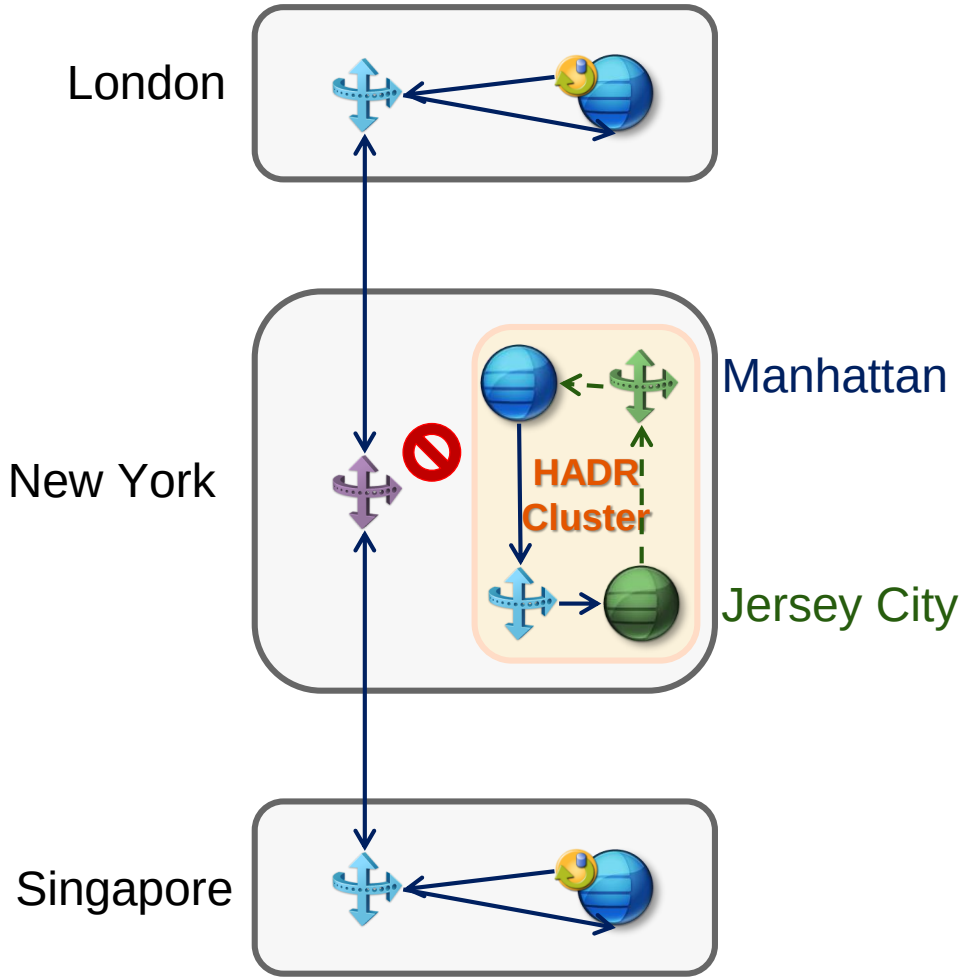
High-level Steps To Enable External Replication (2)

Setup Always-On for current site

- Ideally you will want to do this during a lull in upstream/downstream activity so that the build up in OBQ is minimized

During this time, primary apps should be down

- Otherwise, you will lose data going to external sites as external replication isn't enabled yet



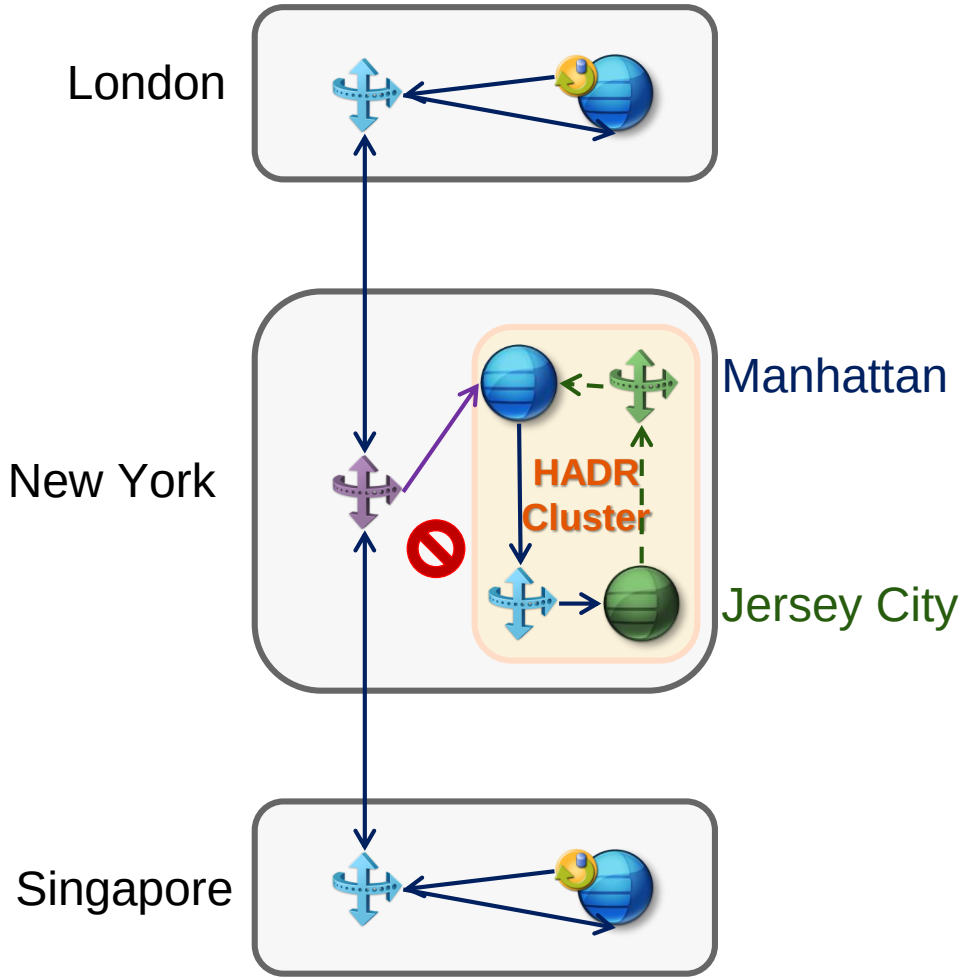
High-level Steps To Enable External Replication (3)

Prepare to re-enable Replication into cluster

- Load rs_install_primary as maintuser
- Grant maint user granular permissions
- If replicating DDL (MSA), grant maint user proxy authorization and set dsi_replication_ddl to true
- Revoke sa_role, hadr privileges from maintuser if previously set

Enable DSI to primary

- Add both primary & standby to external SRS interfaces
- Alter connection to connect to HADR
  - ✓See documentation
- Resume connection



High-level Steps To Enable External Replication (4)

Prepare to re-enable Replication into cluster

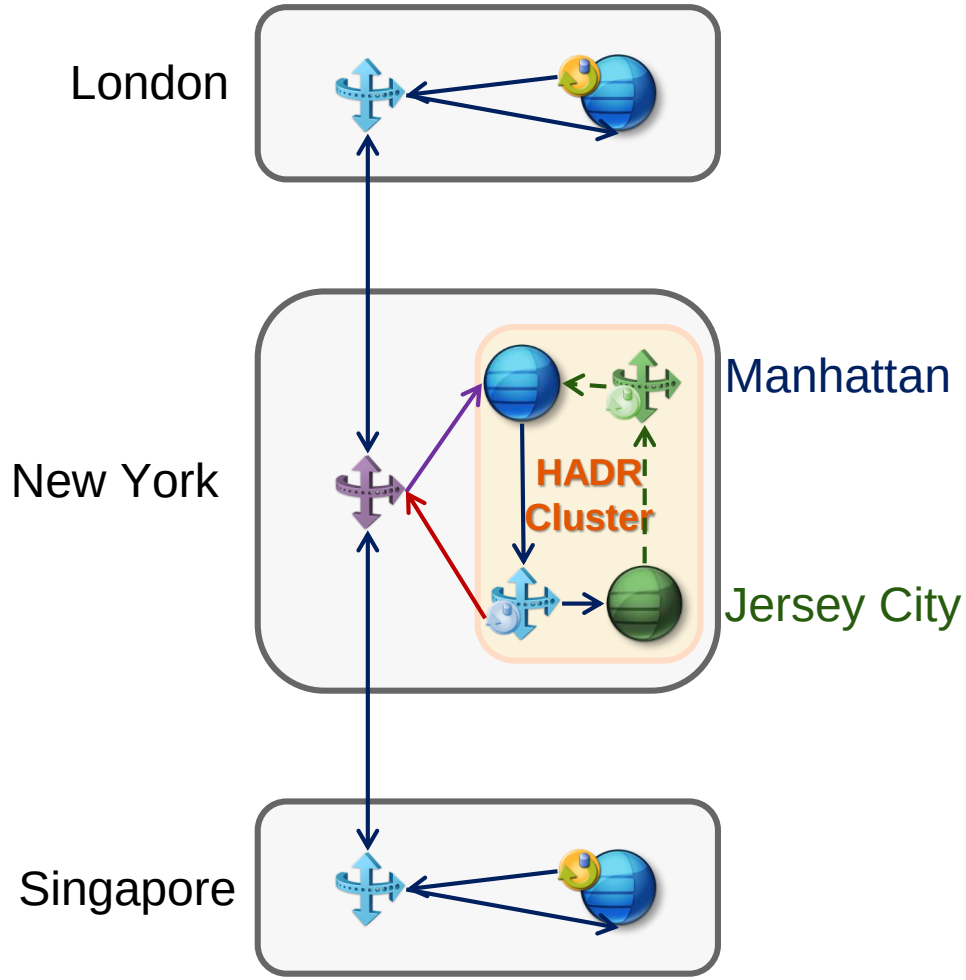
- Load rs_install_primary as maintuser
- Grant maint user granular permissions
- If replicating DDL (MSA), grant maint user proxy authorization and set dsi_replication_ddl to true

Add users to SRS's

- Add db maint user to HADR SRS's and grant manage spq_agent permission
- Add spq_ra_user to external SRS and grant connect source permission

Alter connection to external SRS to add SPQ repagent

Issue sap_enable_external_replication <dbname> in RMA



Roadmap/Future Development

Improvements we are planning or thinking about

Legal Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. This presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this document is not a commitment, promise or legal obligation to deliver any material, code or functionality. This document is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This document is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP´s willful misconduct or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

SAP Adaptive Server Enterprise (ASE)

Product road map overview – key themes and capabilities

Recent innovations

XOLTP Enhancements
- Lockless Cache
- Latch-Free B-Tree
- NVCache
- SNAP (Compiled Queries)

Data Center Operations & Security
- Always-On
  - HADR Clusters
  - External Replication Support
- Workload Analyzer
- DSAM (storage tiering)
- SAP ASE Cockpit

Cloud Services
- AWS, Azure as BYOL
- Docker support
- HCP & MCD DBaaS

SAP HANA Integration
- A4A

Business Suite/SAP Applications
- CDS functionality Phase 1

ASE 16 SP02 PL05 is current release

2017 - Planned innovations

XOLTP Enhancements
- In-Memory Row Store
- Hash based index
- MVCC

Data Center Operations & Security
- Always-On Enhancements
- CCL for SSL
- Idle timeout
- Granular Auditing
- On Demand Network Encryption

Cloud Services
- Cloud services phase 1

SAP HANA Integration
- SAP HANA Schema
- SAP HANA SQL Script

Business Suite/SAP Applications
- CDS functionality Phase 2
- Technical Monitor Cockpit
- Built-in SAP ASE Long term performance Data Repository (BALDR)
- Read-Only Standby

2018 - Product direction

XOLTP Enhancements
- In-Memory Only Tables
- Temporal SQL/Time Series
- >4TB memory & >32K connections
- Proc cache enhancements
- C UDF, JSON, etc.

Data Center Operations & Security
- 64 bit MDA + MDA repository
- Role based resource limits
- Support CI mode for normal SRS
- Always-On Enhancements
  - XA Support, Standby Database
- HSM, LDAP Groups
- Data Masking

Cloud Services
- Cloud services phase 2

SAP HANA/IQ Integration
- Optimized, zero loss data movement to SAP HANA & IQ
- Common Tooling (phase 1)

Business Suite/SAP Applications
- CDS functionality Phase 3

2019 - Product vision

XOLTP Enhancements
- Lazy Persistence
- Non-locking R/O tables/partitions

Data Center Operations & Security
- Workload Analyzer with MDA
- Workload network replay
- Page migration utility
- Undo/redo log utility
- User certificate authentication

Cloud Services
- Cloud services phase 3

SAP HANA/IQ Integration
- Query Enhancements
- Common Tooling (phase 2)

Business Suite/SAP Applications
- CDS functionality Phase 4

FSI Solutions
- Blockchain, Data lineage, Forensic auditing

This is the current state of planning and may be changed by SAP at any time.

For more information on SAP ASE 16 visit:

www.sap.com/ase

http://help.sap.com/ase1602/

https://ideas.sap.com/SAPASE

Jeff Tallman
jeff.tallman@sap.com