Red Hat OpenShift Storage Architecture

Alfred Bach

Principal Learning and Development Instructor

Agenda

- **General Requirements for storage environment**
- Options for providing storage
- Backup & Disaster Recovery
- Combining Solutions
- Feasibility check

OpenShift Virtualization Storage Requirements

Feature	Storage Requirement	Comments
Live Migration	RWX Volumes	Works with both File and Block volumes. RWX on Block volumes is harder
Quick Machine Provisioning	Volume Cloning	If not available, Higher latency to VM provisioning
VM Backups / VM Templates	Volume Snapshots	
Backups at scale	Change Block Tracking (CBT)	Now part of CSI, being implemented by storage and data protection vendors.
Storage live migration	Ability to change the storage class of a volume, without turning off the VM	In TP in 4.17
Support for DR	Volume Based Replication	Different implementation from different vendors
Support for stretched cluster	Stretched storage SANs	Some vendors implement it, CSI support varies.

OpenShift Virtualization Storage

Challenges

RWX Storage Required for Live-Migration:

- anything ?
- anything else ?

Speed of provisioning:

- high number of VMs - migrate or create - parallel operations
- wait time for CSI - operations may time out

Speed of management:

- traditional SAN is not built for it - limited # administrative operations in parallel - slow time of operation
- impact to admins - may affect traditional management operation

Scale of virtual disks:

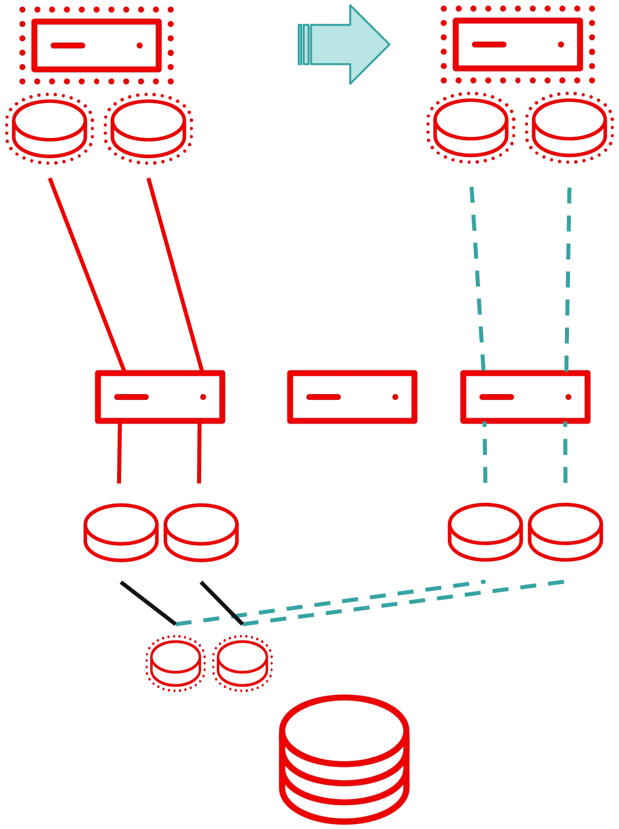
- high number of virtual volumes in enterprise storage - to handle in parallel
- overwhelming the systems admin - understand the actual changes, available resources - reports

Agenda

- General Requirements for storage environment
- **Options for providing storage**
- Backup & Disaster Recovery
- Combining Solutions
- Feasibility check

Direct consumption: SAN /NAS

Reuse existing storage



Architecture:

- direct use of existing storage through CSI driver by vendor
- SAN / dedicated storage network can be leveraged

Configuration:

- 1:1 use of virtual volumes by existing storage
- all nodes must have SAN access + Fibre Channel SAN zoning must include all worker nodes

Performance:

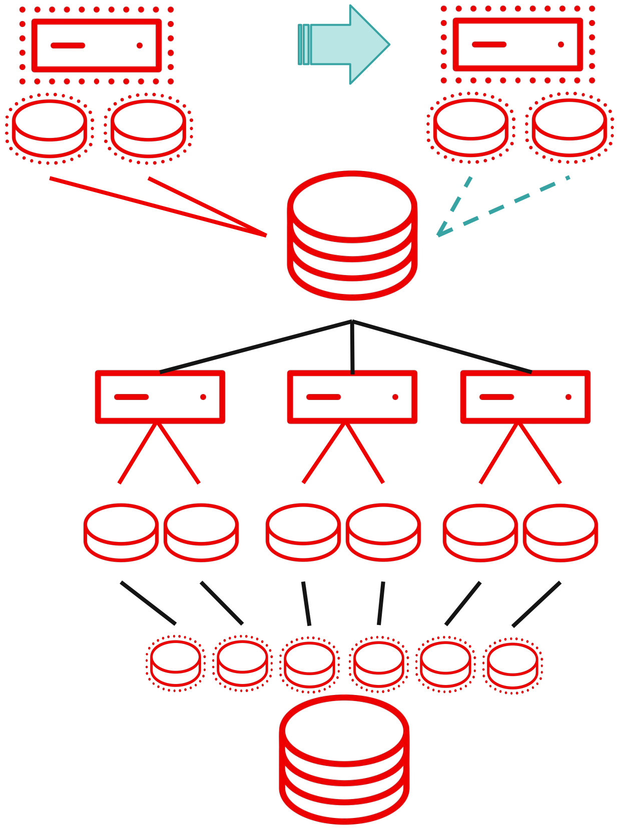
- leverage capacity/performance/latency directly from enterprise storage

Risks:

- potential high number of LUNs per node and multi-pathing challenges
- frequent changes of volume mappings - latency due to rescans and multipathing management
- speed of de-/provisioning & mapping/parallelism of changes => VM migration & provisioning
- Possible limitations of snapshot & cloning
- OpenShift cluster needs permissions on the storage system to de-/provisioning and possibly other operations.

Abstracting existing storage: SDS over SAN

Investment protection



Architecture:

- use of CSI driver by SDS vendor; SDS layer runs internal or external to the OCP cluster

Configuration:

- only SDS nodes must have SAN access + SAN zoning
- PVs by SDS layer; consumes larger virtual volumes from existing storage
- security abstraction for sensitive environments

Performance:

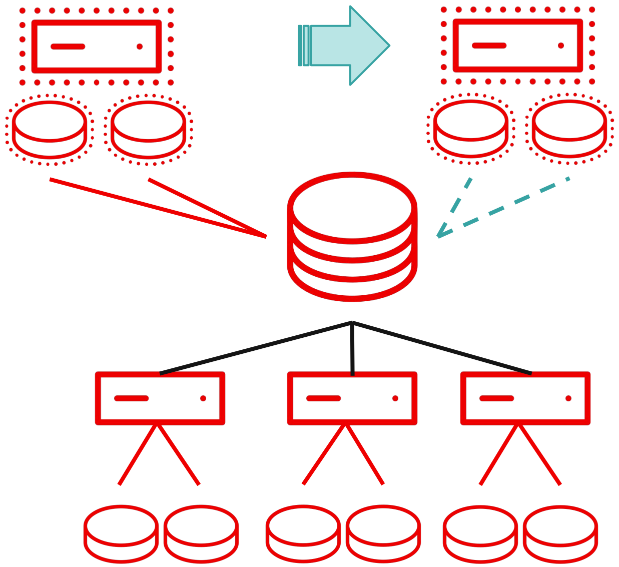
- SDS adds latency - consumes capacity/performance/latency from enterprise storage
- load on array procs - IO structure changed - reduced performance for some workloads to expect

Risks:

- high footprint - in most cases, additional replication may reduce usable capacity
- recommended: separate storage network + use DCB (Data Center Bridging)
- OpenShift cluster might need admin access for de-/provisioning to existing storage
- higher latency than existing storage
- costs

Software-Defined Storage

cloud-native



Architecture:

- SDS layer runs internal or external to the OCP cluster

Configuration:

- PVs by SDS layer; internal to OCP nodes or external on standard servers
- any number of nodes and VMs

Performance:

- capacity/performance/latency – as per design
- recommended: separate storage network + use DCB

Risks:

- layered customer organization in IT
- existing storage not used – phase out or use otherwise
- slightly higher latency than traditional storage (YMWV)
- servers must provide slots for media or external chassis required + CPU/MEM for data layer

Agenda

- General Requirements for storage environment
- Options for providing storage
- **Backup & Disaster Recovery**
- Combining Solutions
- Feasibility check

Backup & Restore

Backups methods

File-level backup

- Requires an agent inside the VM in order to get access to the mounted file system.
- Leverages filesystem metadata and content hashes to recognize changes since the last backup.

Block-level backup

- Works using volume snapshot + copy to a safe location.
- Requires Changed Block Tracking (KEP-3314) for efficiency
- Changed Block Tracking will be possibly delivered as a Dev Preview in ODF 4.19 (RHSTOR-6095)

Backup & Restore

Not a feasible solution at the moment

Red Hat

OADP Operator

provided by Red Hat

OADP (OpenShift API for Data Protection) operator sets up and installs Data Protection...

Velero

- Primarily a block-level backup solution
- CSI snapshot + data mover
- No Changed Block Tracking: need to read the whole CSI snapshot
- Data mover (Kopia) copies changed blocks only and de-duplicates data across volumes in the same namespace.
- CBT must be supported by the CSI storage provider before Velero can start implementing support for CBT
- **Not a feasible solution at the moment**

Backup & Restore

Keep using existing backup solutions

VMware-dependent solutions won't work with OpenShift Virtualization

- For example NetWorker VMware Protection vProxy won't work with OpenShift Virtualization

Backup & Restore

Keep using existing backup solutions

Agent-based solutions will work

- Agent-based backups continue working after migrating to OpenShift Virtualization. No changes needed.
- Need to maintain backup agents for different operating systems

Backup & Restore

Backup ISV Partners

storware
BACKUP AND RECOVERY

IBM Fusion

portworx®
by Pure Storage

TRILIOVAULT

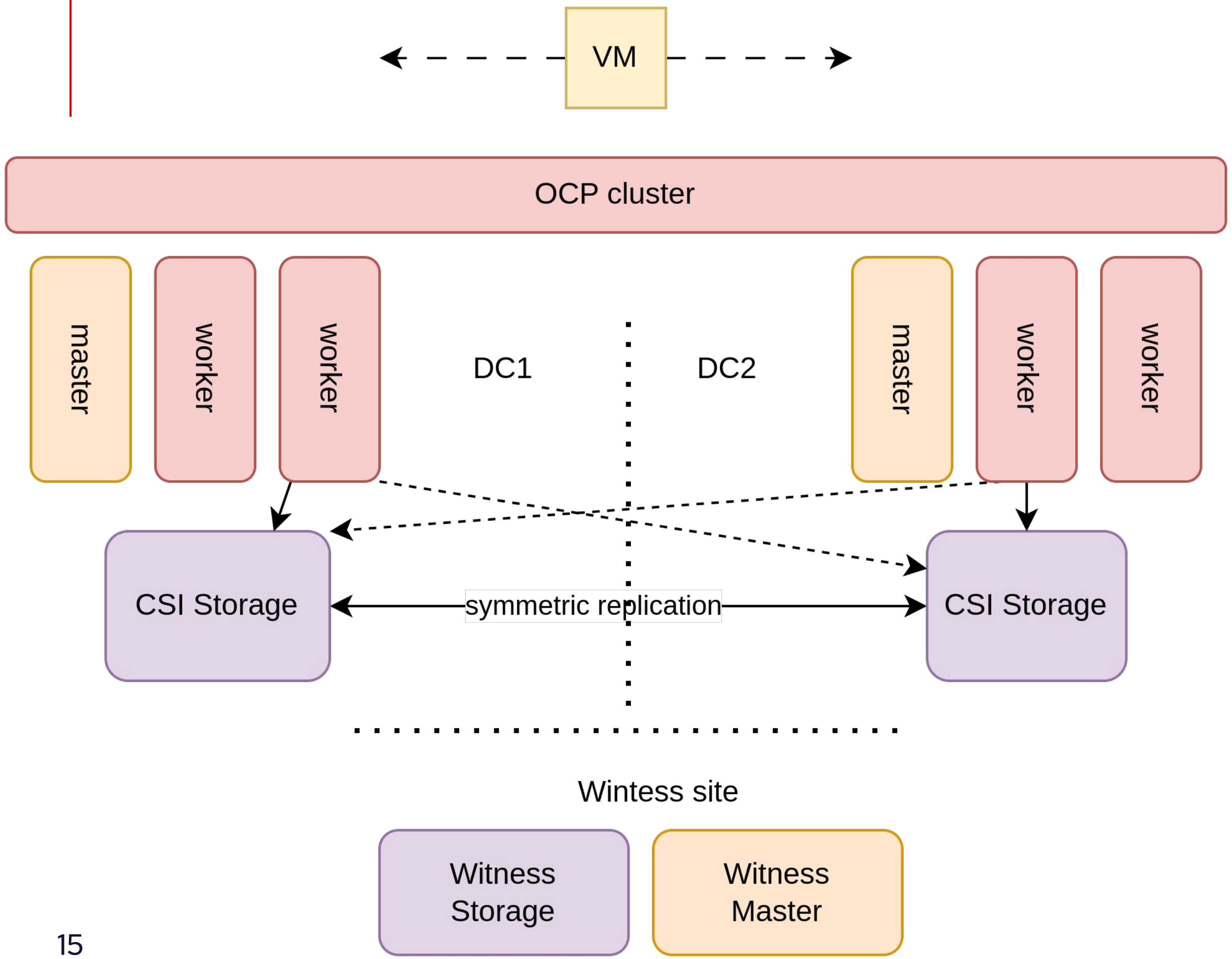
veeam

ot

VERITAS™

Metro DR (Stretched OCP)

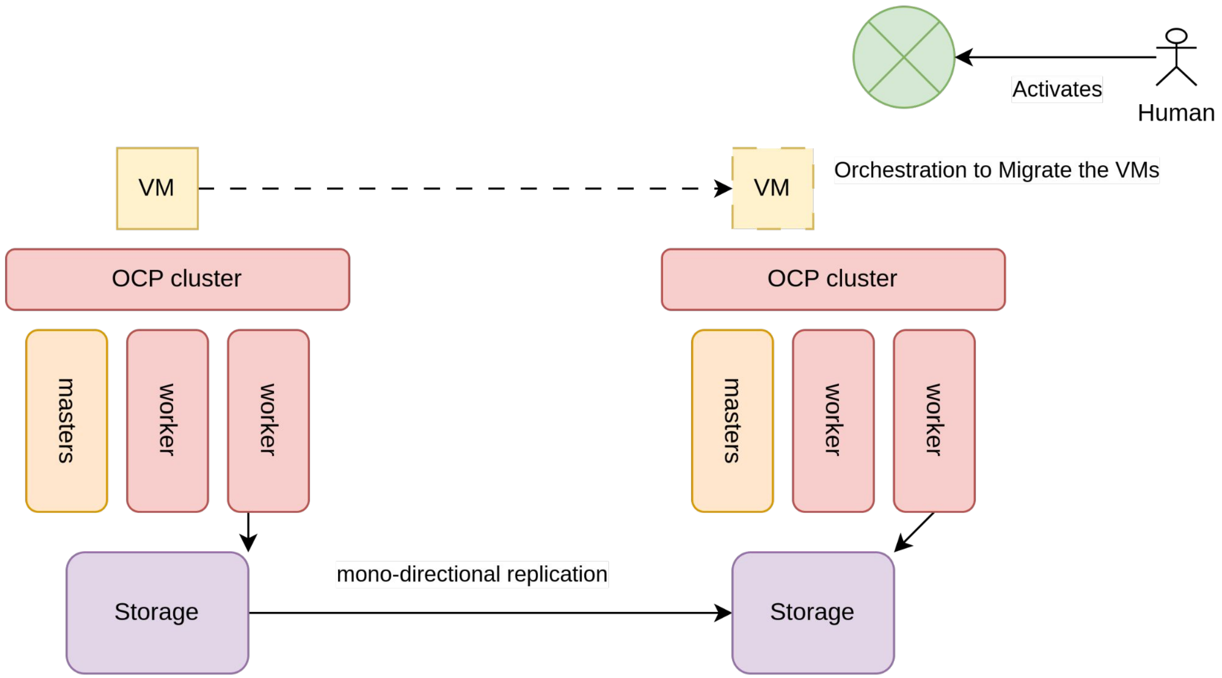
Active/Active -- Latency <5-10ms between DC1 and DC2



- OCP and Storage are stretched across DC1 and DC2
- A witness site is required for both OCP and storage
- Very little latency between DC1 and D2, more latency is usually tolerated between the witness site and the main DCs
- In case of a disaster VMs will be restarted to the healthy DC.
- In case the storage fails, the CSI driver should implement the failover of the multipath connection.
- L2 VLANs for VMs must exist in both DCs.
- Labels and affinity rules might be used to have VMs prefer one DC or the other.

Regional DR

Active/Passive -- Latency > 10ms between DC1 and DC2



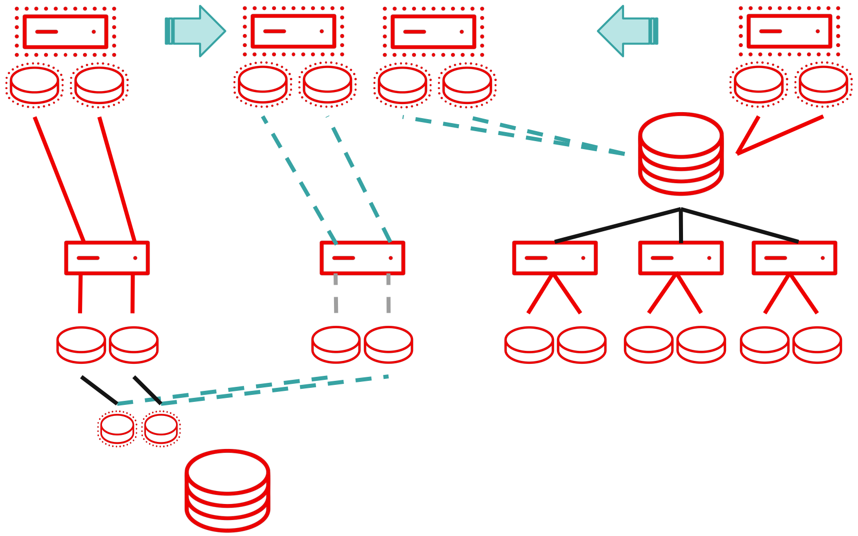
- VM volumes are replicated to the other site (either sync or async replication)
- DR must be triggered externally, typically by a human
- The DR process requires the following components:

Component	Description
Volume Group support	Ability to consistently replicate a group of volumes
Volume Replication Status management	Manage the direction of the replication and whether replication is active or not.
VM Management	Ability to start/stop VMs in the correct DC, attaching them to the same volumes. ACM might help in this space.
Recovery orchestration	Process that restarts all the VMs in the healthy site in case of a disaster, throttling the VM restarts and managing dependencies

Agenda

- General Requirements for storage environment
- Options for providing storage
- Backup & Disaster Recovery
- **Combining Solutions**
- Feasibility check

Combining the solutions



	Existing storage	SDS internal	SDS external
Large VMs	Easy to manage	Ok, size constraints	Good match
Lowest latency (VMs / containers)	Direct low latency	Higher latency	Higher latency
Highly dynamic deployments	Challenging	Best match	Good match
Small VMs (many)	Speed of change	Best match	Good match
Containers	Speed of change	Best match	Good match
High capacity data	Expensive over time	Limits in capacity	Best match
Other storage kind	Limited	Best match	Good match

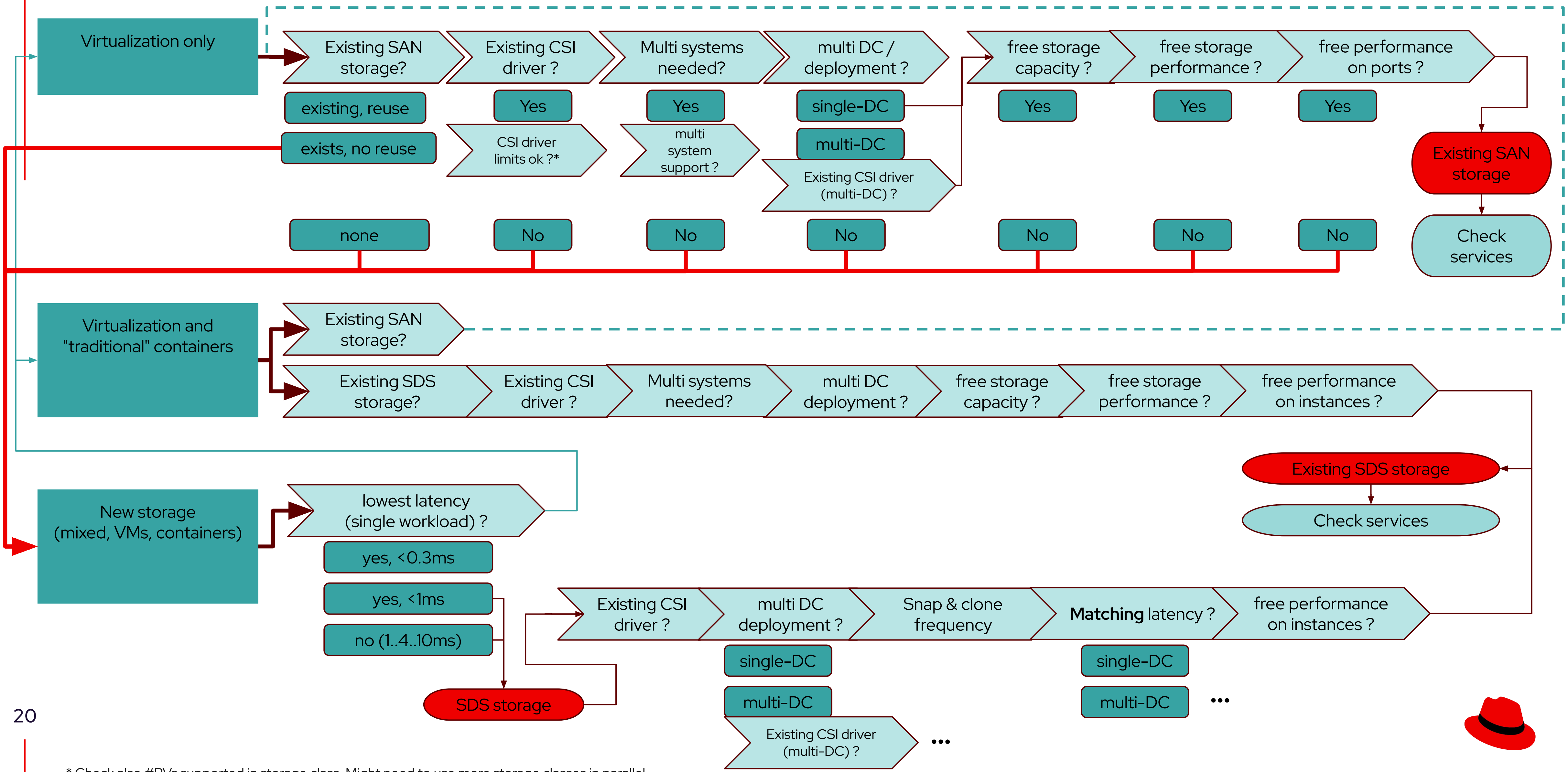
Agenda

- General Requirements for storage environment
- Options for providing storage
- Backup & Disaster Recovery
- Combining Solutions
- **Feasibility check**

* Check also #PVs supported in storage class.

CONFIDENTIAL Red Hat Associate and
NDA partner use only, no further distribution

Feasibility check: Base System evaluation => walk-through (1)



Virtualization only

Existing SAN storage?

existing, reuse

exists, no reuse

none

Existing CSI driver ?

Yes

CSI driver limits ok ?*

No

Multi systems needed?

Yes

multi system support ?

No

multi DC / deployment ?

single-DC

multi-DC

Existing CSI driver (multi-DC) ?

No

free storage capacity ?

Yes

No

free storage performance ?

Yes

No

free performance on ports ?

Yes

No

Existing SAN storage

Check services

Virtualization and "traditional" containers

Existing SAN storage?

Existing SDS storage?

Existing CSI driver ?

Multi systems needed?

multi DC deployment ?

free storage capacity ?

free storage performance ?

free performance on instances ?

Existing SDS storage

Check services

New storage (mixed, VMs, containers)

lowest latency (single workload) ?

yes, <0.3ms

yes, <1ms

no (1..4..10ms)

SDS storage

Existing CSI driver ?

multi DC deployment ?

single-DC

multi-DC

Existing CSI driver (multi-DC) ?

...

Snap & clone frequency

Matching latency ?

single-DC

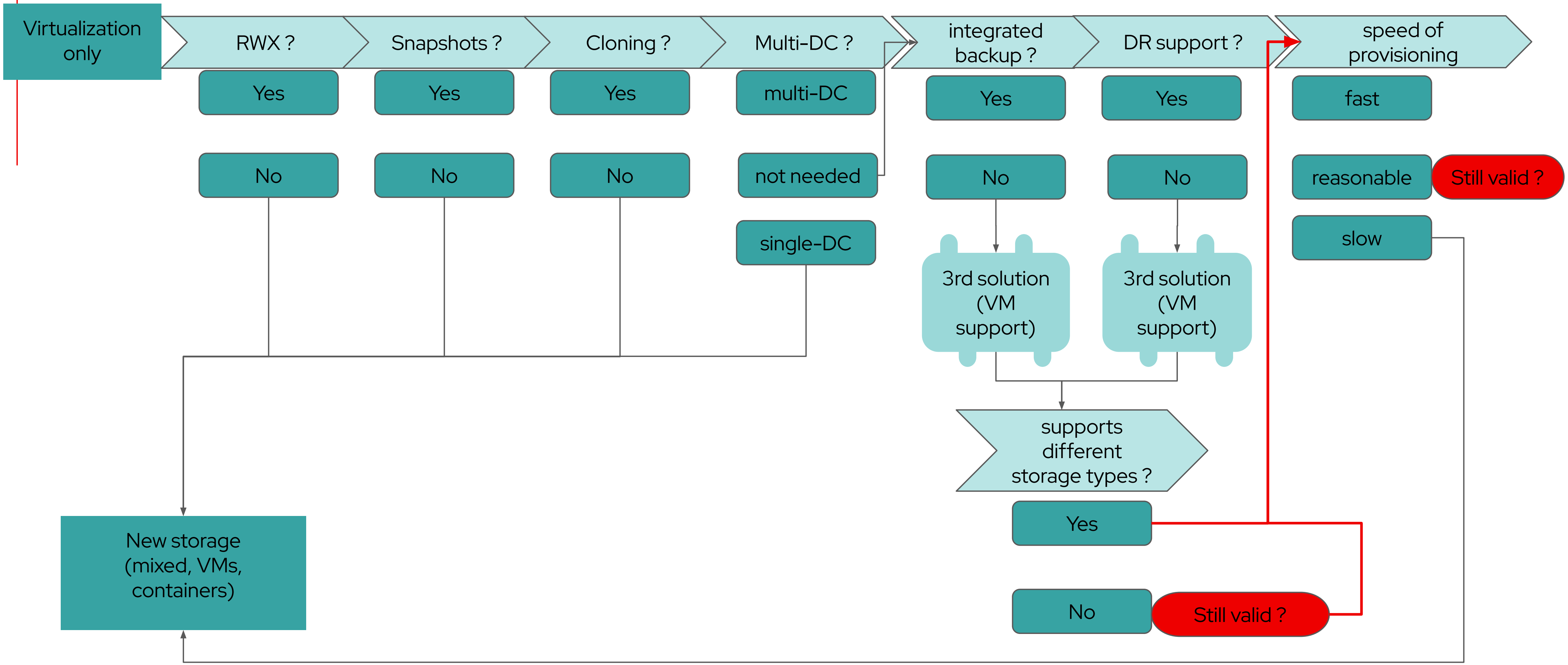
multi-DC

...

free performance on instances ?

* Check also #PVs supported in storage class. Might need to use more storage classes in parallel.

Feasibility check: Check Services for Virtualization (2)


Virtualization only
RWX ?
Snapshots ?
Cloning ?
Multi-DC ?
integrated backup ?
DR support ?
speed of provisioning
Yes
Yes
Yes
multi-DC
Yes
Yes
fast
No
No
No
not needed
No
No
reasonable
Still valid ?
single-DC
slow
3rd solution (VM support)
3rd solution (VM support)
supports different storage types ?
Yes
No
Still valid ?
New storage (mixed, VMs, containers)