Cinder Thin Provisioning

A comprehensive guide

Erlon R. Cruz

NetApp™

Gorka Eguileor

redhat.

Tiago Pasqualini da Silva

NetApp™

Cinder Overprovisioning

What you'll be learning

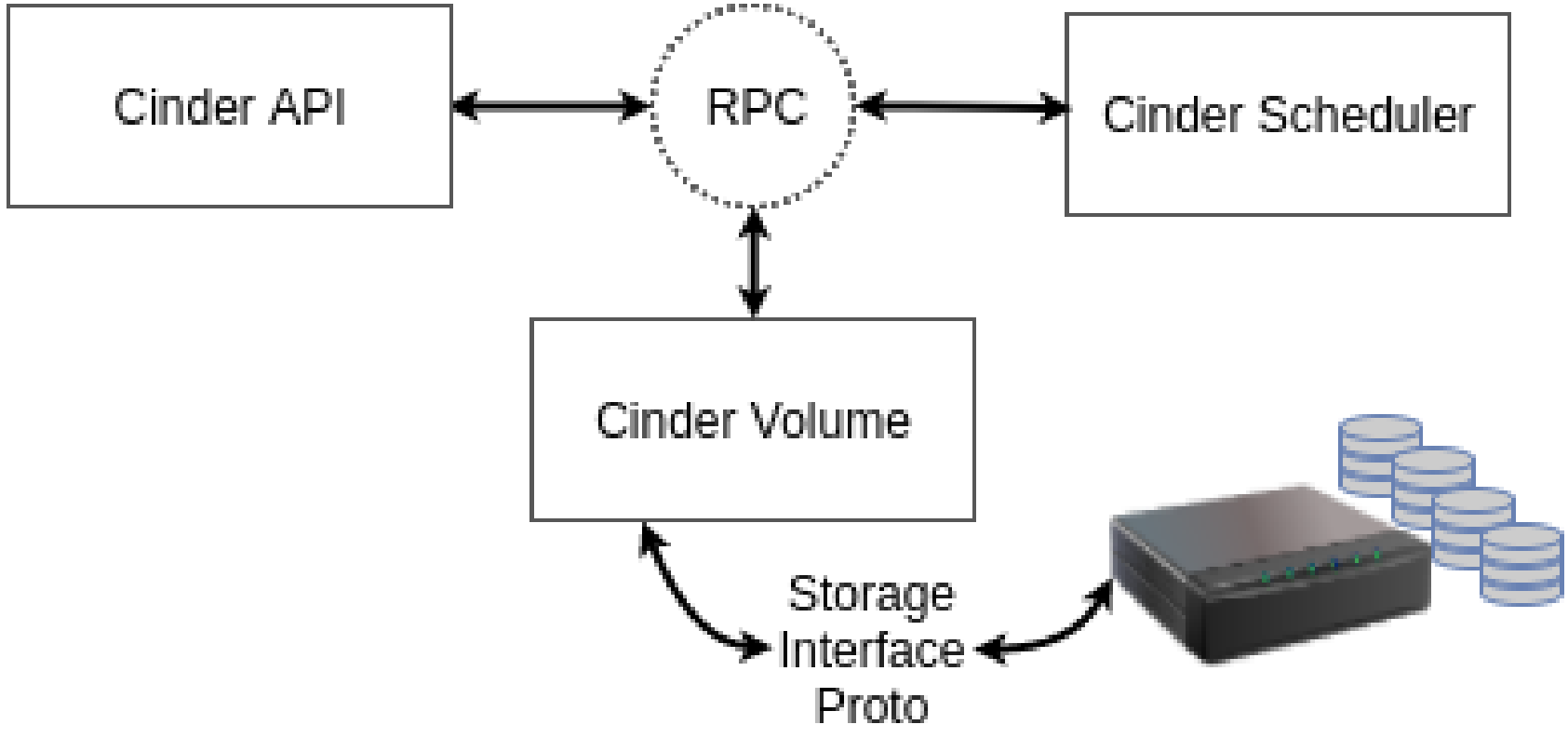
- How scheduling decisions are made
- Filters and how they affect scheduling
- Weighers
- Thin provisioning on Cinder
- How to use thin provisioning
- How to trobleshoot problems
- The future of thin provisioning and Cinder scheduler

redhat.

NetApp™

Cinder architecture

How scheduling decisions are made



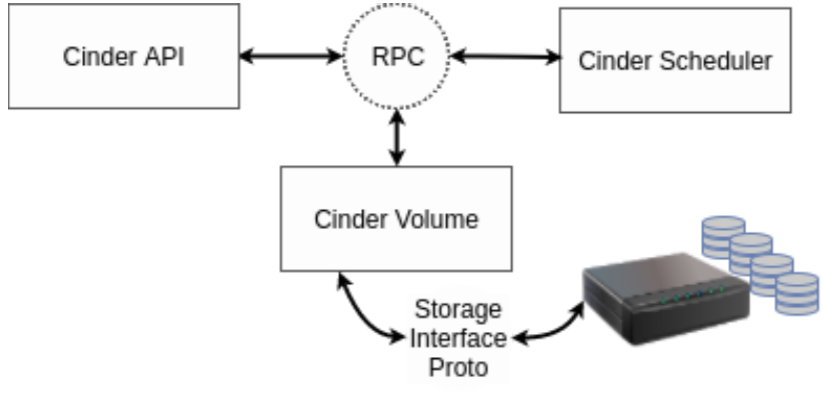
redhat.

NetApp™

Cinder architecture

How scheduling decisions are made

- The API is always the entry points for user requests
- Some requests are handled in the API (list, show, reset-state)
- Some requests go straight to the volume service (delete, delete_snapshot, upload_to_image)
- Most requests go through the scheduler (create, extend, manage, migrate, create_group, migrate and retype)

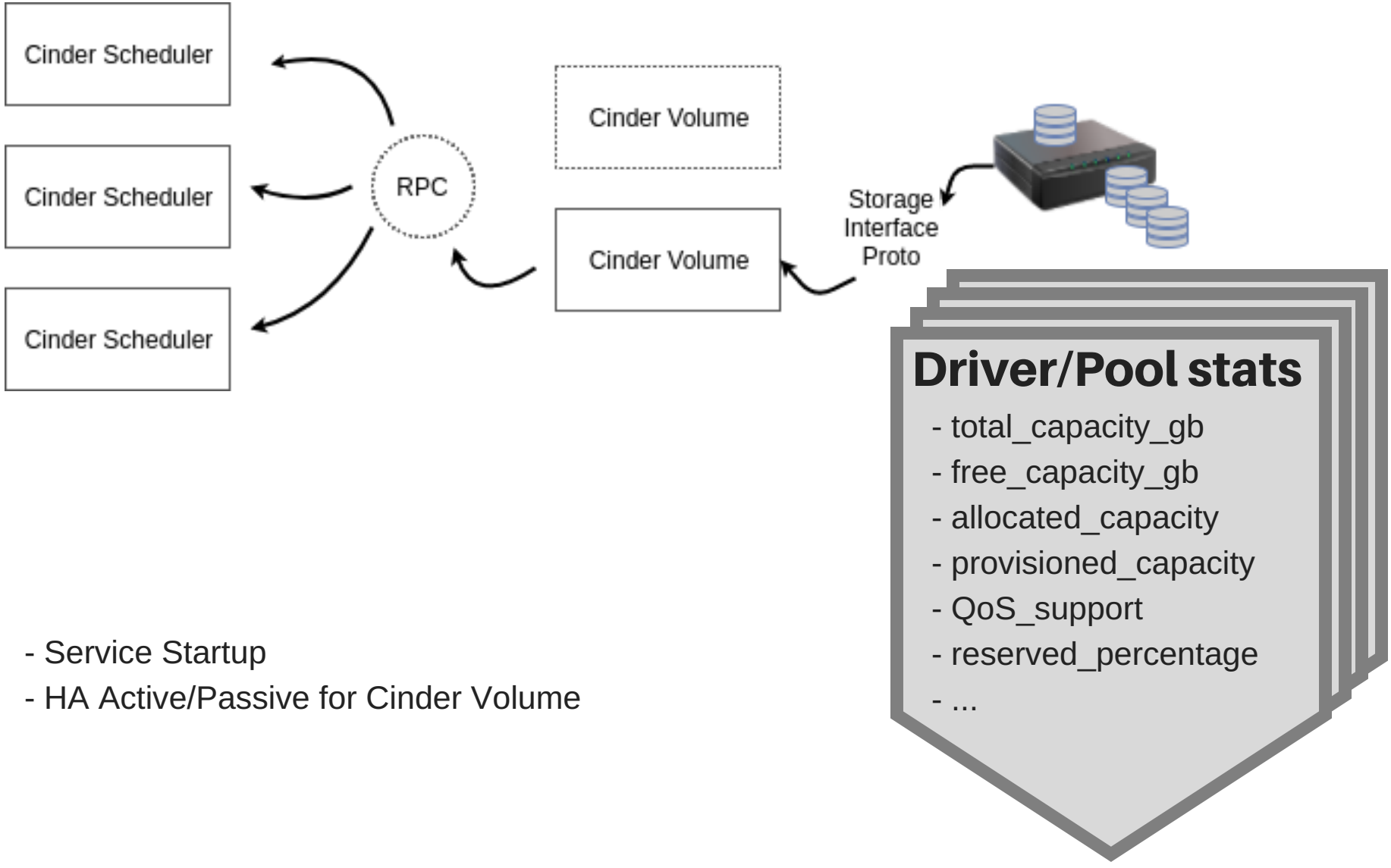


redhat.

NetApp

Cinder architecture

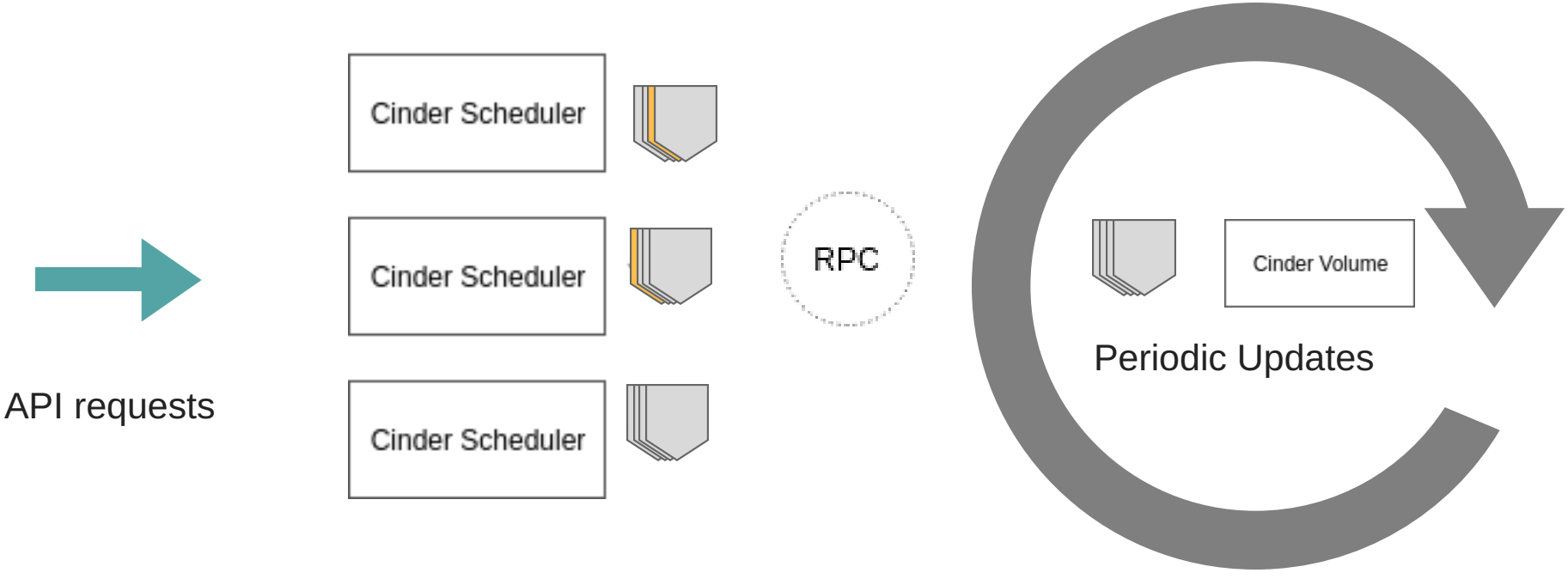
How scheduling decisions are made



- Service Startup
- HA Active/Passive for Cinder Volume

Cinder architecture

How scheduling decisions are made



Stats are not shared/synchronized among services

Filters and filter functions

Given a set of pools, filter out based on defined criteria which services are capable of attending the request.

- 100GB
- QoS
- multi-attach
- az1

- multiattach: True
- tt_cp_gb: 2054
- QoS_support: True
- free_cp_gb: 1580
- reserved: 15
- az1

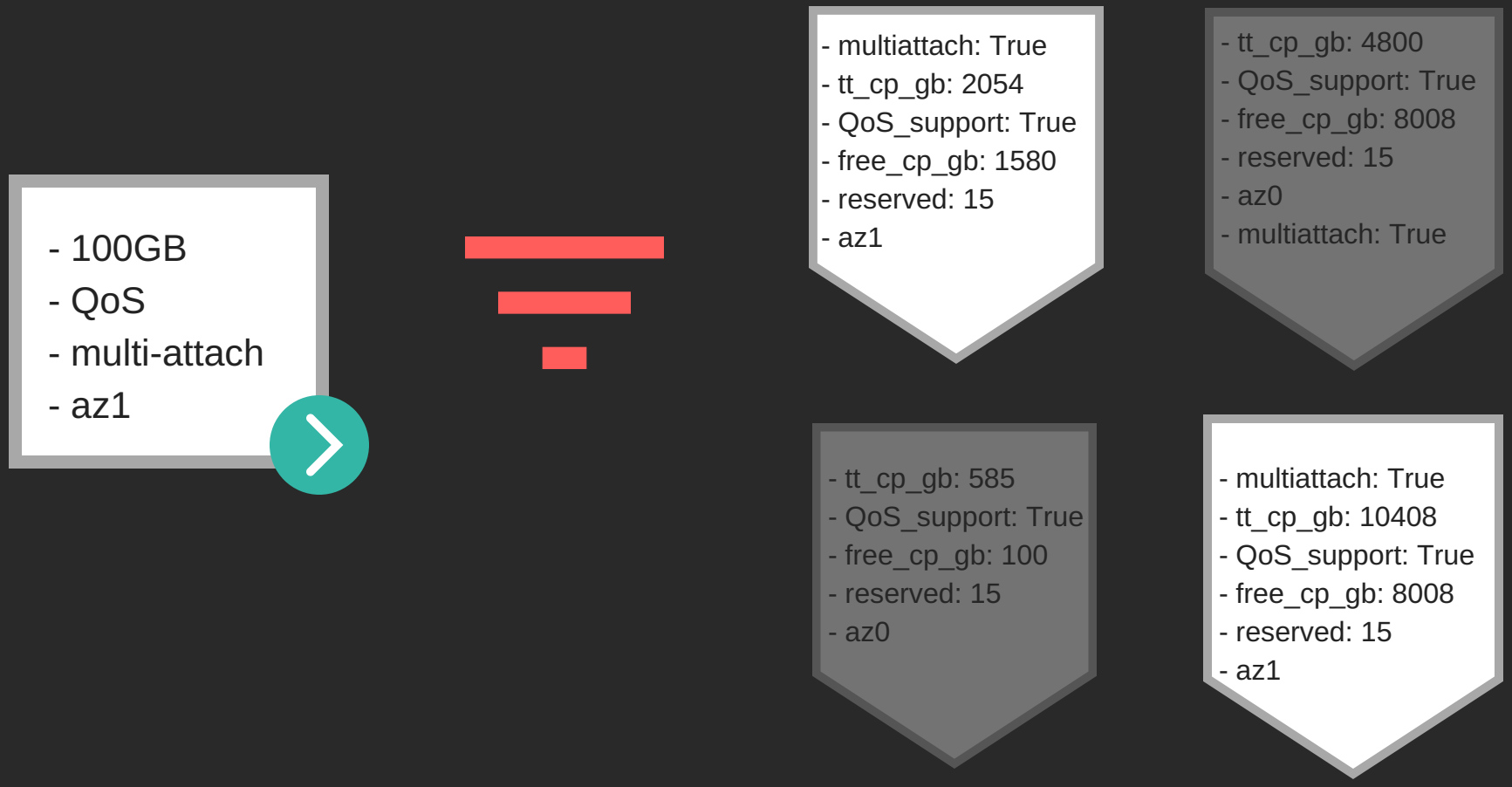
- multiattach: True
- tt_cp_gb: 4800
- QoS_support: True
- free_cp_gb: 8008
- reserved: 15
- az0

- tt_cp_gb: 585
- QoS_support: True
- free_cp_gb: 50
- reserved: 15
- az0

- multiattach: True
- tt_cp_gb: 10408
- QoS_support: True
- free_cp_gb: 8008
- reserved: 15
- az1

Filters and filter functions

Given a set of pools, filter out based on defined criteria which services are capable of attending the request.



Filters and filter funcions

Affinity Filter

Capacity Filter

Capabilities Filter

Driver Filter

Bypass Attempted

Json Filter

AZ Filter

Instance Locality

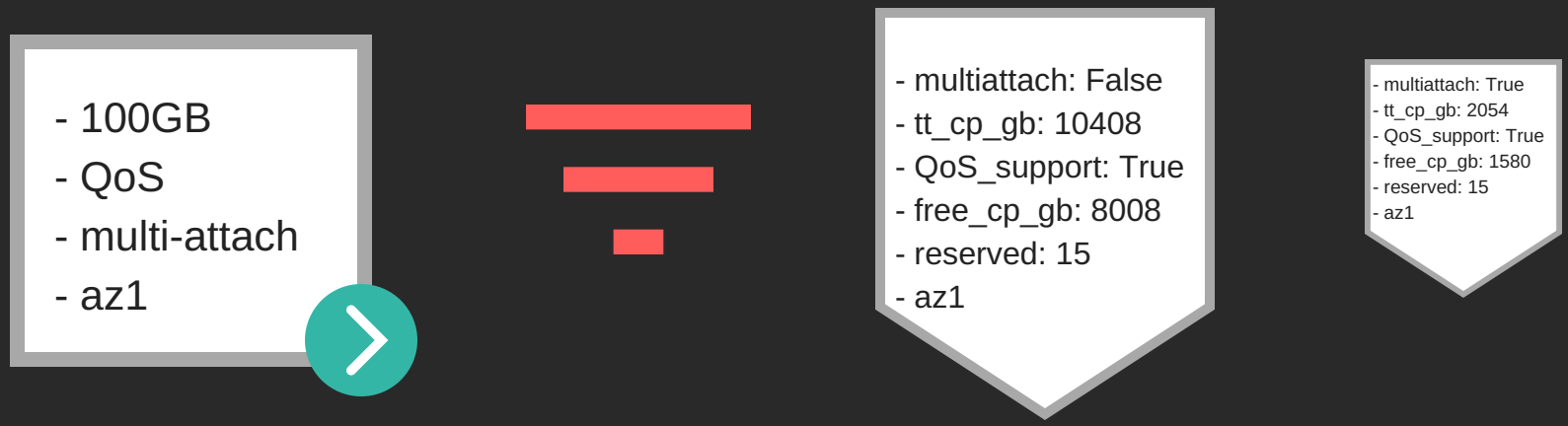
scheduler_default_filters = AvailabilityZoneFilter,CapacityFilter,CapabilitiesFilter

redhat.

NetApp™

Weighers

Given a set of pools, sort based on a given criteria which is the best pool to serve the request.



redhat

NetApp™

Weighers

Allocated Capacity Weigher

Goodness weigher

Capacity Weigher

Volume Number Weigher

Stocastic weigher

scheduler_default_weighers = CapacityWeigher

redhat.

NetApp™

Thin-provisioning support

How everything started

- No way to support storages that supported the feature
- Drivers reported 'infinite' or 'unknown'
- No overprovisioning control
- Initially added in Kilo
- Driver adoption in Liberty (NetApp, NFS Generic, Dell, ScaleIO, etc)

Thin-provisioning support

How it was supposed to work: use cases

- Multiple tiers (platinum, gold, silver) with defined max_oversubscription ratios
- Pools reporting support to thick or thin (each pool being only thick or thin)
- Pools reporting thick and thin at the same time

Thin-provisioning support

Definitions

- **Total capacity:** It is the total physical capacity that would be available in the storage array's pool being used by Cinder if no volumes were present.
- **Free capacity:** It is the current physical capacity available.
- **Allocated capacity:** The amount of capacity that would be used in the storage array's pool being used by Cinder if all the volumes present in there were completely full. Calculated by Cinder.
- **Provisioned capacity:** The amount of capacity that would be used in the storage array's pool being used by Cinder if all the volumes present in there were completely full. Calculated by the driver.
- **Over-subscription ratio:** ratio between **provisioned** and **total capacity.**
- **Reserved percentage:** reserved from total capacity.

Thin-provisioning support

How it was supposed to work: driver side

Drivers service would report

- provisioned_capacity_gb
- max_oversubscription_ratio (from config options)
- reserved_percentage were to be measured against the total_capacity (not free capacity)
- thin_provisioning_support/thick_provisioning_support

Volume service would calculate allocated_capacity for drivers not capable of reporting

Scheduler would filter out pools once they reached their maximum provisioned capacity

Thin-provisioning support

How it was supposed to work: admin actions

Extra-specs should have

- 'capabilities:thin_provisioning_support': '<is> True' or '<is> False'
- 'capabilities:thick_provisioning_support': '<is> True' or '<is> False'

Or:

- 'thin_provisioning_support': '<is> True' or '<is> False'
- 'thick_provisioning_support': '<is> True' or '<is> False'

Configuration should have

- max_oversubscription_ratio

redhat.

NetApp™

Thin-provisioning support

It didn't go so well

Volumes being allowed to be created when they should not be allowed to.

Volumes not being allowed to be created when they should be allowed to.

Thin-provisioning support

What didn't go so well

- Driver maintainers confused with terminology and incorrect capacity calculations (reported values didn't mean the same across all driver implementations)
- Some drivers still had their own way to control over provisioning (LVM, NFS, etc)
- Drivers reporting values that should not be reported
- Development bugs
- max_oversubscription_ratio needed to be continuously calibrated, requiring the service to be restarted
- Lack of synchronization between schedulers
- Race conditions on scheduler/volume services

redhat.

NetApp™

Thin-provisioning problems

Improvements done so far

- **Terminology and documentation:** discussed, defined in spec and documented for developers and users[1]
- **Driver bugs:** Patches to fix non-compliant drivers[2]
- Deprecation of driver's provisioning control options[3][4]
- **Re-calibration problem:** Support for max_oversubscription_ratio='auto' [5][6]
- Scheduler race conditions: WIP

Thin-provisioning

Usage guide

- Check if your storage supports it
- Check if your vendor provides Cinder support (grepping from Cinder code: BlockBridge, EMCExtremeIO, EMCVNX, EQLX, GlusterFS, HPE3par, HPELeftHand, Huawei, Infortrend, LVM, NetApp Ontap, NetApp 7mode, NetApp Eseries, NFS, Pure)*
- Configure storage options for thin provisioning
- Set storage specific configuration options
- Set Cinder configuration options
- Create volume types and extra-specs
- Test setup and configuration

* supports Cinder thin provisioning control

Thin-provisioning

Configuration options

max_over_subscription_ratio:

- >=1 or 'auto'
- for most use cases 'auto'

reserved_percentage:

- 0 - 100
- how quickly can you provide more disks?
- always monitor your storage

backend_specific_configs: e.g. nfs_sparsed_volumes, nas_volume_prov_type, netapp_lun_space_reservation, san_thin_provision, etc

Thin-provisioning

Additional configuration options

scheduler_default_weighers:
- CapacityWeigher or AllocatedCapacityWeigher

capacity_weight_multiplier:
- <>0, usually -1 or 1
- stack vs spreading

allocated_capacity_weight_multiplier:
- <>0, usually -1 or 1
- stack vs spreading

Thin-provisioning

Troubleshooting

- What OS release am I? (*for RH users most of upstream fixes were backported)
- When possible get a fresh pool and reproduce the problem
- Release notes are friends
- Check scheduler logs, pay attention on requests' timing
- Get your fists ready: cinder/cinder/scheduler/filters/capacity_filter.py
- Check the related bugs on newer releases

Appendix

Troubleshooting

Liberty

Fix capacity filter to allow oversubscription https://review.openstack.org/185764
Allow provisioning to reach max oversubscription https://review.openstack.org/188031
LVM Thin Provisioning auto-detect https://review.openstack.org/104653
Configure space reservation on NetApp Data ONTAP https://review.openstack.org/211659
Rename free_virtual in capacity filter https://review.openstack.org/214276
Implement thin provisioning support for E-Series https://review.openstack.org/215833
Fix use of wrong storage pools for NetApp Drivers https://review.openstack.org/222413
NetApp: Fix volume extend with E-Series https://review.openstack.org/224285
NetApp E-Series over-subscription support https://review.openstack.org/215801
ZFSSA driver to return project 'available' space https://review.openstack.org/211299
NetApp DOT block driver over-subscription support https://review.openstack.org/215865

Appendix

Troubleshooting

Mitaka

Fix ScaleIO driver provisioning key Fix ScaleIO driver provisioning key
NetApp eseries: report max_over_subscription_ratio correctly
https://review.openstack.org/267726
Set LVM driver default overprovisioning ratio to 1.0 https://review.openstack.org/266986
fix NFS driver max_over_subscription_ratio typo https://review.openstack.org/269830
Fix thin provisioning flags in NetApp drivers https://review.openstack.org/267513
Correcting thin provisioning behavior https://review.openstack.org/275408

Appendix

Troubleshooting

Newton

Fix HNAS stats reporting https://review.openstack.org/344477

Differentiate thick and thin provisioning https://review.openstack.org/315352

Ocata

RBD Thin Provisioning stats https://review.openstack.org/178262

Pike

Don't check thin provisioning when manage volumes https://review.openstack.org/457119

Kamiario: Fix over subscription reporting https://review.openstack.org/492206

SMBFS: enable thin provisioning support flag https://review.openstack.org/484424

Appendix

Troubleshooting

Queens

RBD: Fix stats reporting https://review.openstack.org/486734
Stop overriding LVM overprovisioning ratio and deprecate
https://review.openstack.org/507985
Netapp Ontap: Adds support for auto-max-over-subscription
https://review.openstack.org/534855
Dell EMC PS: Fix over-subscription ratio stats https://review.openstack.org/514338
Check available capacity before creating resources https://review.openstack.org/509011
Dell EMC PS: Fix over-subscription ratio stats https://review.openstack.org/512740
NetApp E-series: Fix provisioned_capacity_gb https://review.openstack.org/518406
Fix allocated_capacity_gb race on create volume https://review.openstack.org/#/c/546983/
NetApp ONTAP: Fix reporting of provisioned_capacity_gb
https://review.openstack.org/#/c/509780/
Fix reporting old stats https://review.openstack.org/546717

References and links

[1] https://docs.openstack.org/cinder/latest/contributor/thin_provisioning.html
[2] https://review.openstack.org/#/q/status:merged+project:openstack/cinder+(message:thin+OR+message:provisioning+OR+message:overprovisioning+OR+message:ratio)
[3] https://review.openstack.org/#/c/269841/
[4] https://review.openstack.org/#/c/564265/
[5] https://review.openstack.org/#/c/534854/
[6] https://docs.google.com/spreadsheets/d/1wpNg-80YkHyrQqSWk120znkKRMOg1xJB8va12-L_vso/edit?usp=sharing

Thank you!

Please don't hesitate to contact us if you have any questions

redhat.

NetApp™