

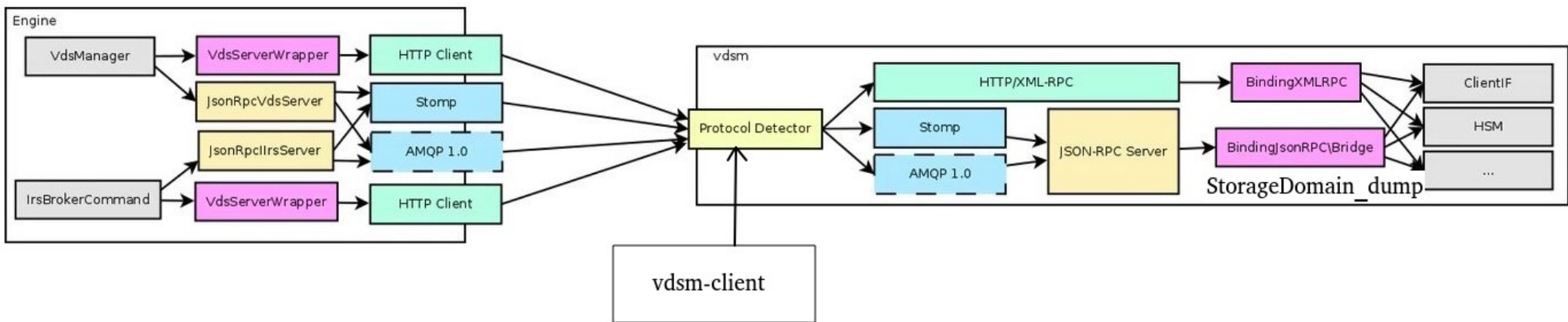
Storage Domain Dump API



Amit Bawer
abawer@redhat.com
July 2020



You are (about) here



Basic requirements:

- 1) vsm 4.40.18 or later.
- 2) An online Storage Domain (Storage has to be accessible from host).



Usage

CLI:

```
vdsm-client StorageDomain dump sd_id=<SD-UUID> [full=1]
```

Python:

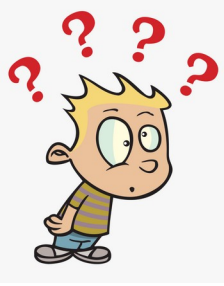
```
from vds import client
...
cli = client.connect(host, port=54321, use_tls=True)
cli.StorageDomain.dump(sd_id=sd_uuid, full=1)
```

Engine (theoretical):

```
public class JsonRpcVdsServer implements IVdsServer
{...
  @Override
  public StorageDomainDumpReturn dumpStorageDomain(String sdUUID, boolean full=false) {
    JsonRequest request =
      new RequestBuilder("StorageDomain.dump")
        .withParameter("sd_id", sdUUID)
        .withParameter("full", full)
        .build();
    Map<String, Object> response = new FutureMap(this.client, request);
    return new StorageDomainDumpReturn(response);
  }...
}
```

Storage Domain Information

- SD Metadata
- Volumes
- Lockspace (Host IDs)
- Leases (SDM & Volumes)
- Xleases (HA VMs leases)



Block SD - MD Layout

- # vgs -o tags 9c9f6d68-1ead-4bf3-a188-6a7145986f50

```
MDT_ALIGNMENT=1048576,MDT_BLOCK_SIZE=512,MDT_CLASS
=Data,MDT_DESCRIPTION=dom2,MDT_IOOPTIMEOUTSEC=10,M
DT_LEASERETRIES=3,MDT_LEASETIMESEC=60,MDT_LOCKPOLIC
Y=,MDT_LOCKRENEWALINTERVALSEC=5,MDT_POOL_UUID=992
af8dc-b623-11ea-a8a2-
525400200635,MDT_PV0=pv:360014058c0cd7f6f3f54ef8a45e8
571f&44&uuid:BF2agT-dNmH-0CqY-lwPB-UAUZ-apex-
IScdt4&44&pestart:0&44&pecount:797&44&mapoffset:0,
MDT_ROLE=Regular,MDT_SDUUID=9c9f6d68-1ead-4bf3-a188-
6a7145986f50,
MDT_TYPE=ISCSI,MDT_VERSION=5,MDT_VGUUID=1xe9D1-
LQH6-5Kn0-o8OM-ZAC2-7bE8-rTtFRI,
MDT__SHA_CKSUM=14b0e69ba44fbc74296035c6503b821d1b3
0eff6,RHAT_storage_domain
```

Block SD Metadata

```
"metadata": {  
  "alignment": 1048576,  
  "block_size": 512,  
  "class": "Data",  
  "metadataDevice": "360014058c0cd7f6f3f54ef8a45e8571f",  
  "name": "dom2",  
  "pool": [  
    "992af8dc-b623-11ea-a8a2-525400200635"  
  ],  
  "role": "Regular",  
  "state": "OK",  
  "type": "ISCSI",  
  "uuid": "9c9f6d68-1ead-4bf3-a188-6a7145986f50",  
  "version": "5",  
  "vgMetadataDevice": "360014058c0cd7f6f3f54ef8a45e8571f",  
  "vguuid": "1xe9D1-LQH6-5Kn0-o8OM-ZAC2-7bE8-rTtFRI"  
}
```

File SD - MD Layout

```
# cat /rhev/data-center/<pool_uuid>/<sd_uuid>/dom_md/metadata
```

```
ALIGNMENT=1048576
```

```
BLOCK_SIZE=512
```

```
CLASS=Data
```

```
DESCRIPTION=nfs1
```

```
IOOPTIMEOUTSEC=10
```

```
LEASERETRIES=3
```

```
LEASETIMESEC=60
```

```
LOCKPOLICY=
```

```
LOCKRENEWALINTERVALSEC=5
```

```
POOL_UUID=992af8dc-b623-11ea-a8a2-525400200635
```

```
REMOTE_PATH=10.35.28.239:/exports/data
```

```
ROLE=Regular
```

```
SDUUID=b0a72804-5a70-4df8-9135-41ed0f755d70
```

```
TYPE=NFS
```

```
VERSION=5
```

```
_SHA_CKSUM=ee2ce6be89d7d2d3e5d70d3b859c198347bec3bf
```

File SD Metadata

```
"metadata": {  
  "alignment": 1048576,  
  "block_size": 512,  
  "class": "Data",  
  "name": "nfs1",  
  "pool": [  
    "992af8dc-b623-11ea-a8a2-525400200635"  
  ],  
  "remotePath": "10.35.28.239:/exports/data",  
  "role": "Regular",  
  "type": "NFS",  
  "uuid": "b0a72804-5a70-4df8-9135-41ed0f755d70",  
  "version": "5"  
}
```


Block Volume - MD layout

LV Tags

```
# lvs -o tags c3b0a94e-ab0e-4221-b5a5-aedc151aa42/b2992531-d40f-448e-8259-39c3c04c9777
  IU_0eb460d9-c5e1-43a9-928a-3b2e786eb039, MD_2, PU_00000000-0000-0000-0000-000000000000
```

Metadata LV

```
/dev/9c9f6d68-1ead-4bf3-a188-6a7145986f50/metadata
```

Addressing

```
MD Base = 1MB (Compatibility with v4)
MD Slot Size = 8KB
MD Offset = (MD Base) + (MD Slot) * (MD Slot Size)
```

Slot Entry

```
CTIME=1593760271
DESCRIPTION={"DiskAlias":"vm1_Disk1","DiskDescription":"super duper vd disk"}
DISKTYPE=DATA
DOMAIN=9c9f6d68-1ead-4bf3-a188-6a7145986f50
FORMAT=RAW
GEN=0
IMAGE=9e7973a8-ecb8-43f4-98a7-d556df72e995
LEGALITY=LEGAL
PUUID=00000000-0000-0000-0000-000000000000
TYPE=PREALLOCATED
VOLTYPE=LEAF
EOF
```

Block Volume

```
"volumes": {  
  "9e498bc3-e995-414d-81aa-242b7749a3b5": {  
    "apparentsize": 1073741824,  
    "capacity": 1073741824,  
    "ctime": 1593760271,  
    "description": "{\"DiskAlias\":\"vm1_Disk1\", \"DiskDescription\":\"vm disk\"}",  
    "disktype": "DATA",  
    "format": "RAW",  
    "generation": 0,  
    "image": "9e7973a8-ecb8-43f4-98a7-d556df72e995",  
    "legality": "LEGAL",  
    "mdslot": 2,  
    "parent": "00000000-0000-0000-0000-000000000000",  
    "status": "OK",  
    "truesize": 1073741824,  
    "type": "PREALLOCATED",  
    "voltype": "LEAF"  
  }, ...  
}
```

File Volume - MD Layout

```
# cat /rhev/data-center/<pool_uuid>/<sd_uuid>/images/<image_uuid>/<vol_uuid>.meta
```

```
CAP=1073741824
```

```
CTIME=1593760654
```

```
DESCRIPTION={"DiskAlias":"vm2_Disk1","DiskDescription":"super duper vm2 disk"}
```

```
DISKTYPE=DATA
```

```
DOMAIN=b0a72804-5a70-4df8-9135-41ed0f755d70
```

```
FORMAT=RAW
```

```
GEN=0
```

```
IMAGE=4dd6bd83-99dc-433f-ba4f-fdb84206f4c0
```

```
LEGALITY=LEGAL
```

```
PUUID=00000000-0000-0000-0000-000000000000
```

```
TYPE=SPARSE
```

```
VOLTYPE=LEAF
```

```
EOF
```

File Volume

```
"volumes": {
  "e7d2866c-6ed5-4e31-a326-6e0784dc6b79": {
    "apparentsize": 1073741824,
    "capacity": 1073741824,
    "ctime": 1593760654,
    "description":
"{\"DiskAlias\": \"vm2_Disk1\", \"DiskDescription\": \"super duper vm2 disk\"}",
    "disktype": "DATA",
    "format": "RAW",
    "generation": 0,
    "image": "4dd6bd83-99dc-433f-ba4f-fdb84206f4c0",
    "legality": "LEGAL",
    "parent": "00000000-0000-0000-0000-0000-000000000000",
    "status": "OK",
    "truesize": 4096,
    "type": "SPARSE",
    "voltype": "LEAF"
  },
}
```

Sanlock ids, leases (full=1)

```
# sanlock direct dump /dev/9c9f6d68-1ead-4bf3-a188-6a7145986f50/ids:0 -A 1M -Z 512
```

offset	lockspace	resource	timestamp	own	gen
0	9c9f6d68-1ead-4bf3-a188-6a7145986f50	caf9ffa0-36f2-4b58-9807-63f3ec18cf82.vm-18-187.	243650	1	6
512	9c9f6d68-1ead-4bf3-a188-6a7145986f50	8ac70da7-59f8-414c-87d9-d68eb0115f30.localhost.	115308	2	3

```
# sanlock direct dump /dev/9c9f6d68-1ead-4bf3-a188-6a7145986f50/leases:0:$((103 * 1024 * 1024))  
-A 1M -Z 512
```

offset	lockspace	resource	timestamp	own	gen	lver
01048576	9c9f6d68-1ead-4bf3-a188-6a7145986f50	SDM	0	1	1	1
105906176	9c9f6d68-1ead-4bf3-a188-6a7145986f50	e17215ed-956b-4f99-8604-2f5b7b6f45f2	0	0	0	0
106954752	9c9f6d68-1ead-4bf3-a188-6a7145986f50	7864a713-481d-4f42-b8e5-38a44864d3cf	0	0	0	0

leases volume layout:

- offset=0 SPM (Backward compatibility with V0 and V2)
- offset=1MB SDM (V3 and later, using sanlock)
- offsets=2MB, ..., 100MB For future use / Unassigned.
- Block volumes leases are starting from offset 101MB
- File volumes leases are kept on a dedicated <vol-uuid>.lease file per image path (excluded from dump)

NamesandNumbers.com	
River Valley	Hailey 788-1206
BATES Paul 118 Willow Rd.	Hailey 788-6222
BATES Steve 105 Audubon Pl.	Sun Valley 788-5050
BATES VICKY - INTERIOR MOTIVES PO Box 1820	Ketchum 726-0722
BATHUM Roy 235 Spur Ln	See West Adam
BATHMAN	726-7494
BATT Jeffrey & Camille	Ketchum 726-8896
BATTERSBY Patricia 116 Ritchie Dr.	Hailey 788-4279
BAUER Charlotte 621 Northstar Dr.	
BAUER CHARLOTTE LINDBERG	
Radiance Skin Care Studio	Hailey 578-2214
BAUER Matt 3340 Woodside Blvd.	Hailey 578-0703
BAUER Rich	720-0165

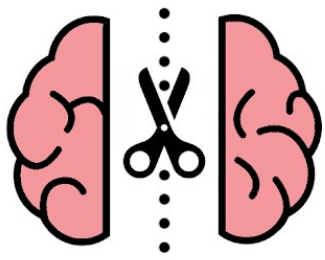
Lockspace (full=1)

```
"lockspace": [
  {
    "gen": 6,
    "lockspace": "9c9f6d68-1ead-4bf3-a188-6a7145986f50",
    "offset": 0,
    "own": 1,
    "resource": "caf9ffa0-36f2-4b58-9807-63f3ec18cf82.vm-18-187.",
    "timestamp": 241039
  },
  {
    "gen": 3,
    "lockspace": "9c9f6d68-1ead-4bf3-a188-6a7145986f50",
    "offset": 512,
    "own": 2,
    "resource": "8ac70da7-59f8-414c-87d9-d68eb0115f30.localhost.",
    "timestamp": 112695
  }
],
```



Leases (full=1)

```
"leases": [  
  {  
    "gen": 1,  
    "lockspace": "9c9f6d68-1ead-4bf3-a188-6a7145986f50",  
    "lver": 1,  
    "offset": 1048576,  
    "own": 1,  
    "resource": "SDM",  
    "timestamp": 0  
  },  
  {  
    "gen": 3,  
    "lockspace": "9c9f6d68-1ead-4bf3-a188-6a7145986f50",  
    "lver": 1,  
    "offset": 105906176,  
    "own": 2,  
    "resource": "5fe7b349-dbf5-4941-97fd-216f5b35e643",  
    "timestamp": 0  
  }, ...  
]
```



Xleases (full=1)

Dumps the xleases index for HA VM leases instead of the /dev/<sd-uuid>/xleases volume

```
"xleases": {  
  "c21aa521-4fdb-43b3-ac54-9c97d66b8dad":  
  {  
    "offset": 3145728,  
    "updating": false  
  },  
}
```




SD Errors

```
# vdsm-client StorageDomain dump sd_id=b0a72804-5a70-4df8-9135-41ed0f755d70
```

- Domain is offline:

```
(code=358, message=Storage domain does not exist: ('b0a72804-5a70-4df8-9135-41ed0f755d70',))
```

- File Domain metadata is inaccessible:

```
(code=100, message=[Errno 13] Permission denied)
```

- File Domain metadata is corrupted:

```
(code=752, message=Meta Data seal is broken (checksum mismatch): 'cksum = ee2ce6be89d7d2d3e5d70d3b859c198347bec3bf, computed_cksum = f6b449957db72b9f99561de9cdc0bceaa337af3f')
```



Block Volume Errors

- Bad LVM md-slot tag

```
"volumes":{  
  "a217c6bf-ed68-40bd-88b7-b022a01297ba":{  
    "status":"INVALID",  
    "image":"b3ab8950-1ea0-48af-9c33-d3659028c527",  
    "parent":"00000000-0000-0000-0000-000000000000",  
    "truesize":1073741824,  
    "apparentsize":1073741824  
  }  
}
```

WARN Invalid tag 'MD_bad-slot-number' for lv
6465a611-f8ce-4234-a7a5-4eaa6f5e1f47/a217c6bf-ed68-40bd-88b7-b022a01297ba
WARN Failed to get metadata from lv tags for lv 6465a611-f8ce-4234-a7a5-
4eaa6f5e1f47/a217c6bf-ed68-40bd-88b7-b022a01297ba: None

- Bad MD slot

```
"volumes":{  
  "a217c6bf-ed68-40bd-88b7-b022a01297ba":{  
    "status":"INVALID",  
    "mdslot":4,  
    ...  
  }  
}
```

WARN Failed to parse metadata slot 4 offset=0: Meta Data key not found error:
("key='SIZE' lines=[],")



File Volume Errors

- Corrupted metadata file

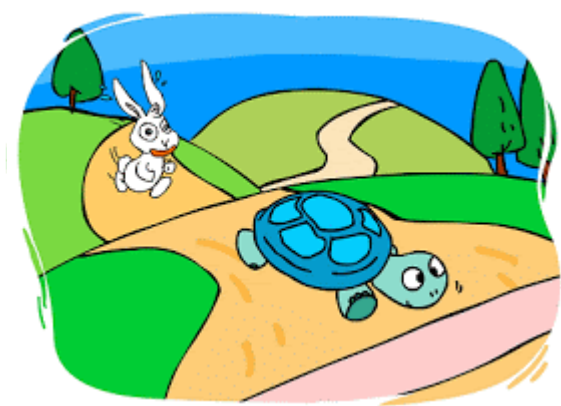
```
"volumes":{  
  "55367923-13fc-448d-aecc-3bbbc3c316ba":{  
    "image":"c7e04d99-c392-441c-b5cc-5d4e8a5a096a",  
    "status":"INVALID",  
    "truesize":4096,  
    "apparentsize":2147483648  
  }  
}
```

WARN Failed to parse meta file for volume
d793f0e6-ac2f-4970-921e-247e50671a6e/55367923-13fc-448d-aecc-3bbbc3c316ba:
Meta Data key not found error: ("key='SIZE' lines=[b'bad-key=bad-value']",)

- No permissions on metadata file

```
"volumes":{  
  "e041420b-8185-490d-a0c8-9d3cd9d240f8":{  
    "image":"53624d2f-7a31-4a31-bbda-1d0ce5bb477f",  
    "status":"INVALID",  
    "truesize":4096,  
    "apparentsize":2147483648  
  }  
}
```

WARN Failed to parse meta file for volume baa5266d-eb82-40b4-b018-a8e0b1ebd91f/1bfefe6f-
09af-445c-9892-aac0ce9a959b:
[Errno 13] Permission denied



Performance

Default Block SD dump with 1030 VMS (Only SD MD & Volumes)

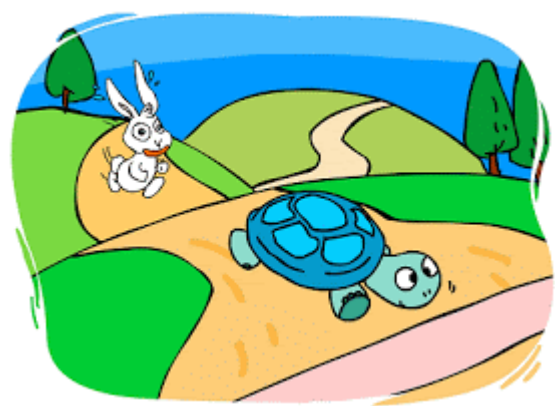
```
# time vdsm-client StorageDomain dump sd_id=38ca2952-9ee4-4443-8f59-3884e294deb9
```

```
real 0m1.293s  
user 0m0.174s  
sys 0m0.026s
```

Full block SD dump with 1030 VMS

```
# time vdsm-client StorageDomain dump sd_id=38ca2952-9ee4-4443-8f59-3884e294deb9 full=1
```

```
real 0m8.430s  
user 0m0.196s  
sys 0m0.020s
```



Performance

Almost 300 times faster than existing dump-volume-chains tool
(API based version is WIP)

```
# time vdsm-tool dump-volume-chains f32458fb-7cb5-431f-8767-5b3cf63fb85b
```

```
real 18m44.384s
user 0m4.729s
sys 0m1.094s
```

```
# time vdsm-client StorageDomain dump sd_id=f32458fb-7cb5-431f-8767-5b3cf63fb85b
```

```
real 0m3.818s
user 0m0.475s
sys 0m0.093s
```

* SD with 1356 disks

References

- RFE for SD Dump API
<https://bugzilla.redhat.com/1557147>
- API spec
<https://github.com/oVirt/vdsm/blob/0218ed7e6ff9fd2a0d8dce5da469d1975a133/lib/vdsm/api/vdsm-api.yml#L10109>
- Xleases Deep Dive by Nir
https://www.youtube.com/watch?v=WLnU_YsHWtU
- Sanlock DevConf by Vojtech
<https://www.youtube.com/watch?v=we0zW7ekfAI>
- Volume legality in live merge
<https://www.ovirt.org/develop/release-management/features/storage/live-merge.html>