Storage Performance Diagnosis and Optimization

2020/9/16

Security Level:

HUAWEI

Foreword

- Storage performance diagnosis and optimization is a systematic topic, need to understand the principle of IO process and grasp many diagnosis and tuning skills
- This cause will introduce how to troubleshooting some basic performance issue on Huawei storage, and typical ways to improve the performance

HUAWEI

Contents

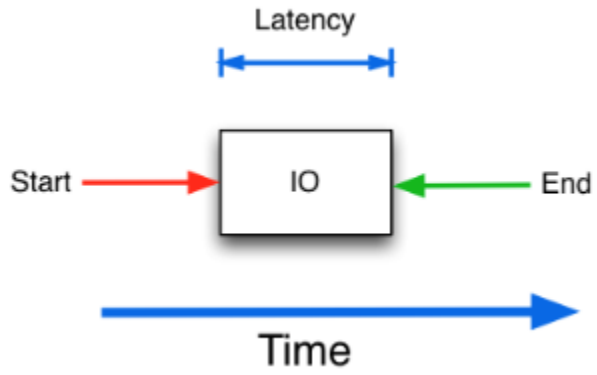
Understand IO

- IO means single read/write request
- From micro perspective, an IO go through application -> host OS(filesystem, volume) -> host HBA -> storage network -> storage array(cache, RAID, disk). It looks like a car/truck on high way, transport person or goods from one place to another.

HUAWEI

Understand Performance Indicators

- Usually, we use IOPS, bandwidth(throughput) and latency(response time) to measure the performance of storage.
- IOPS means how many read/write request can be handled in one second.
- Latency = (1 second / IOPS) * Queue depth
- Bandwidth = IOPS * Average IO size



IO Model

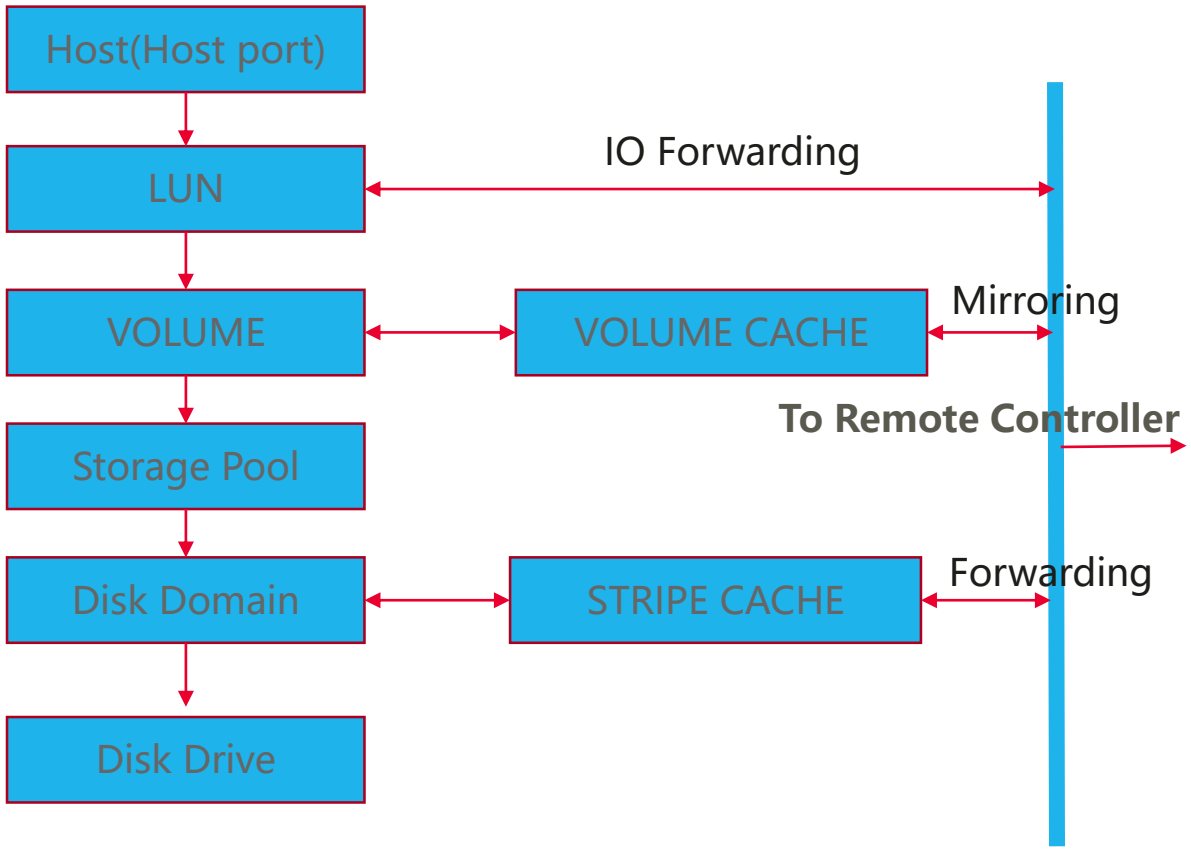
- Read/Write ratio: Read IO and Write IO consumes different system resource, also different handling process. For example, RAID6 storage pool, one write IO means **at least** 2 disk read IO and 2 disk write IO, while one read IO may just need one disk read IO
- Queue depth：Queue exists everywhere on IO handling path. Queue depth means how many IO can be handled at the same time, like the lanes of high way
- Average IO size: Usually, use IOPS to measure the performance for small IO application, and bandwidth to measure the performance for larger IO application
- Randomness: Storage performance is highly related to pre-fetch on read cache and merge on write cache. Random read/write makes storage cache loss effectiveness

Flow Control

- Flow control is a way to ensure the service quality when overloaded
- System flow control triggered when the CPU usage or write cache water level over threshold
- The same as SmartQos, the flow control is based on Token/Credit , it will result in higher latency
- Start from Dorado V300R002, the storage system dynamically flow control “background” IO first, like Formatting, VMware Full Copy, Window ODX, etc. , then limit normal read/write IO

IO Handling Process In Storage(V3 Serial)

- The resource/queue depth at each stage may affect the performance
- Also, system resource, like CPU, memory may affect the performance

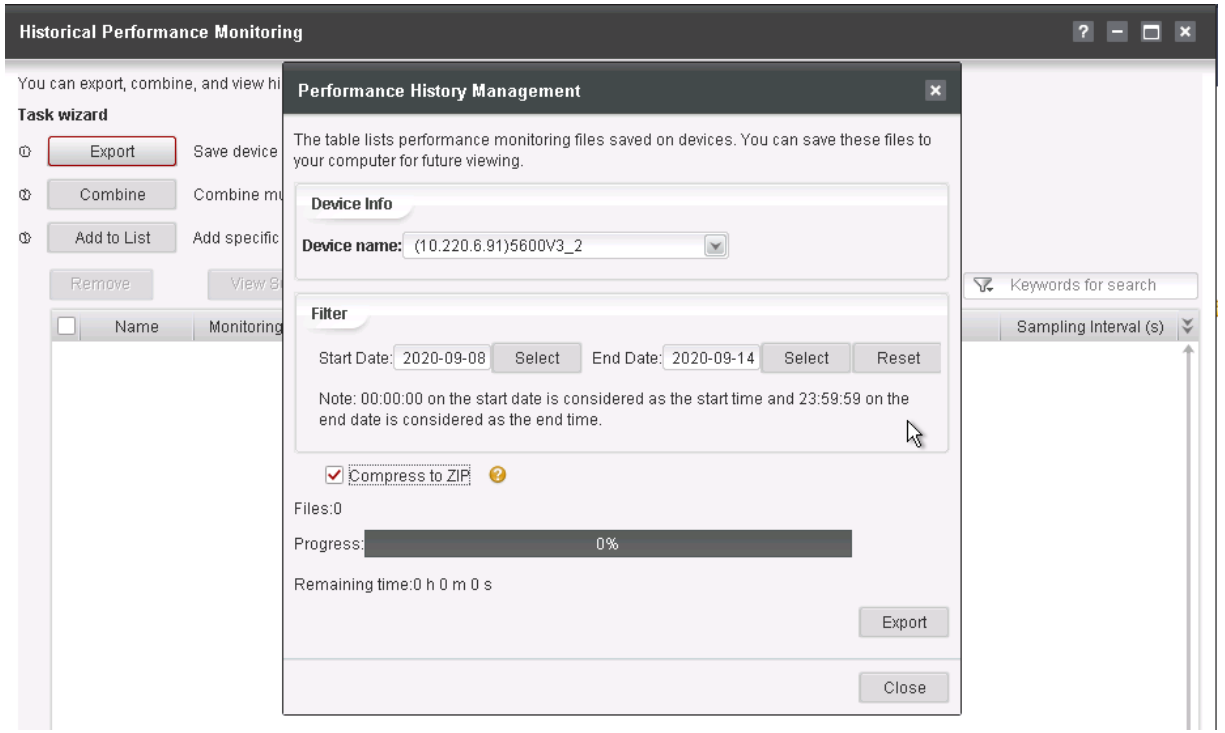


HUAWEI

Contents

Collect Storage Performance Data

- Download performance log files by SFTP tool(Applicable for dual-controller enclosure storage)
- Export performance log by SmartKit(Applicable for all storage)



HUAWEI

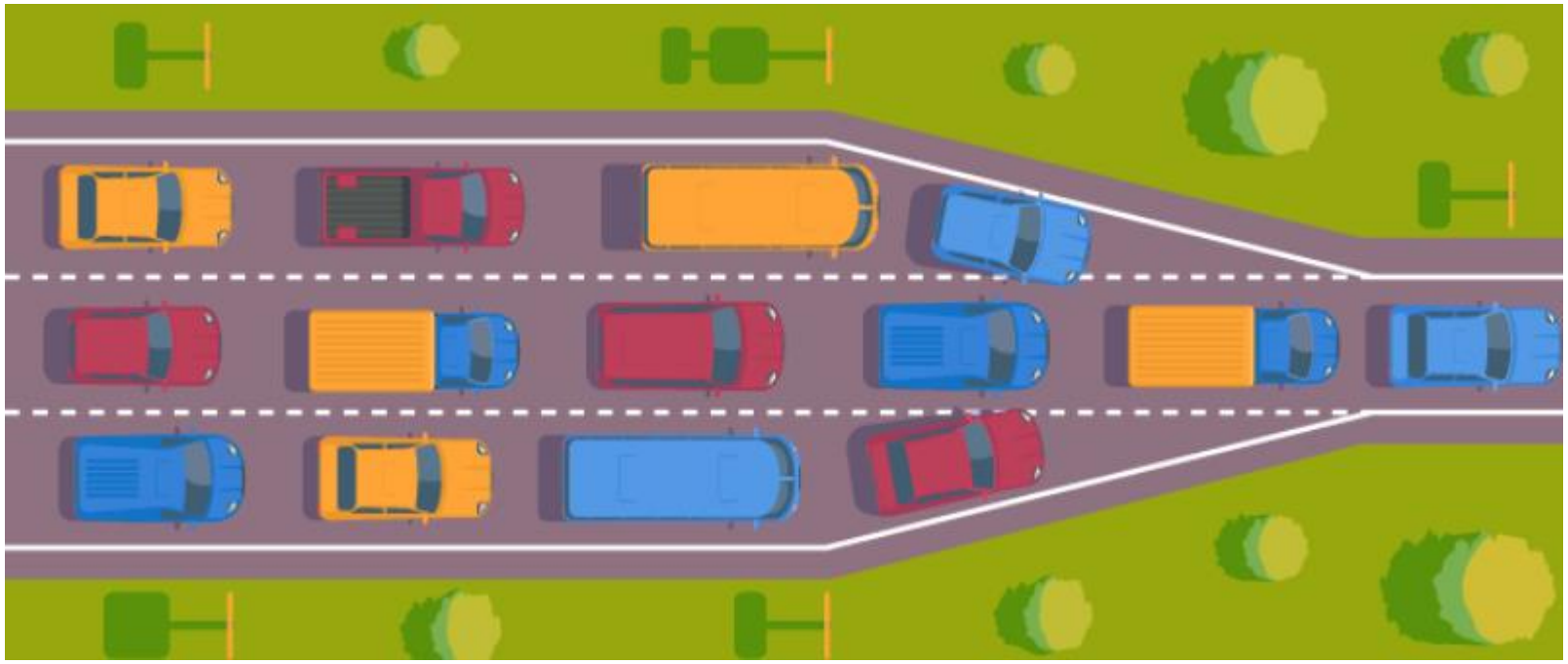
Understand Phenomenon

- Trigger condition, for example, VMware storage vmotion, disk expansion, etc.
- Issue time
  - First occurrence time
  - Periodically(start time, end time) or continuously
- Issue scope, for example, specific host, specific LUNs, etc.
- “Translate” the issue symptom into standard performance indicators. For example:
  - “Write bandwidth can only reach 100MB/s when copy file to storage” VS “copy file slow”
  - “Read latency over 5ms” VS “Database slow”

HUAWEI

Identify Performance Bottleneck

- Performance issue means bottleneck somewhere.
- Performance issue troubleshooting => Find the bottleneck



Contrast and Exclusion Analysis

- Quickly identify the issue component/device based comparison test or performance monitor.
- Exclude the normal component/device in order to narrow the investigation scope
- Example 1:

  VMware VM slow -> Write latency high -> Only HyperMetro LUN write slow -> Write performance become normal after HyperMetro pair split -> Replication link issue

- Example 2:

  VMware VM slow -> Only slow when on one of ESXi -> One of the path slow -> Port bit error

Layer by Layer Analysis

- Analyze from top to bottom
- Analyze latency in priority, also check whether pressure over expectation
- Example 1:

  Host read IO latency over 30ms -> Storage host port IO latency over around 30ms -> storage pool read latency over 30ms -> NL_SAS disk usage reach 100%

- Example 2:

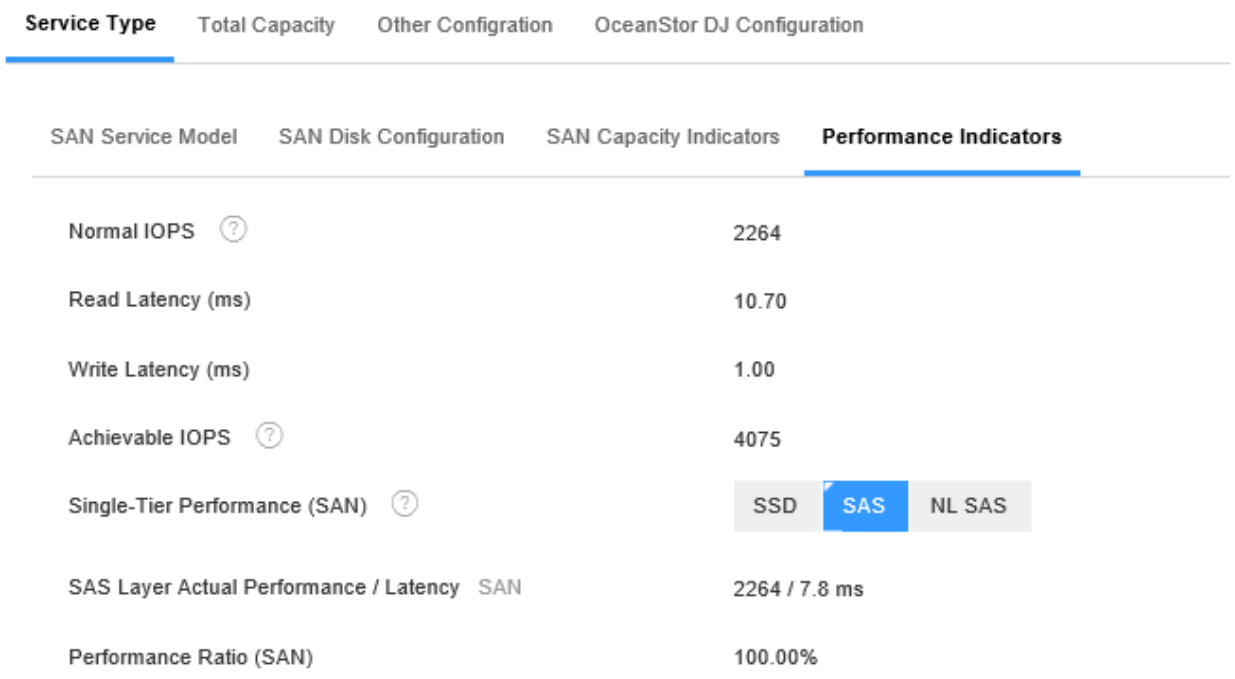
  Host read IO latency over 30ms -> Storage host port IO latency around 2ms -> FC SAN network buffer credit used out

Typical Performance Issue

- Cache write-through because of controller fault, BBU faulty, capacity used out etc.
- High CPU/Cache usage because of Deduplication/Compression，VMware vMotion，Disk formatting, etc.
- Flow control after service pressure over storage/LUN capability

Evaluate Storage Performance by eDesigner

- Evaluate the storage capability based on specific IO model, check whether flow control triggered. Link: http://app.huawei.com/unistar/edesigner/configuration!showParas.action?projectId=16211832&siteId=3214754&projectCategory=0



HUAWEI

Contents

FTDS Introduction

- FTDS(Fault Tracing and Diagnosing System) which used to trace IO process, IO statistics, IO handling time at each stage.
- Usually, this platform is used to analyze complex performance issue.

HUAWEI

Request Time Trace

- Trace the average handling time, maximum handling time, minimum handling time, received request, ended request, total request at each function on IO path

```
developer:/>change rtt_switch general trace_id=all status=open
Command executed successfully.
developer:/>debug
admin:/diagnose>rtt show -t all

  Name                                tid        open       Av         Max        Min        Beg        End        Total
  ----------                          ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
  SSDC_WRITE                               6          1          0          0          0          0          0          0
  SSDC_READ                                7          1          0          0          0          0          0          0
  EXTENT_VOLUME_READ                       8          1          0          0          0          0          0          0
  EXTENT_SUB_READ                          9          1          0          0          0          0          0          0
  EXTENT_VOLUME_WRITE                     10          1          0          0          0          0          0          0
  EXTENT_SUB_WRITE                        11          1          0          0          0          0          0          0
  EXTENT_MIGRATE                          12          1          0          0          0          0          0          0
  EXTENT_INIT                             13          1          0          0          0          0          0          0
  EXTENT_GET_LOCK                         14          1          0          0          0          0          0          0
  EXTENT_GET_META_INFO                    15          1          0          0          0          0          0          0
  EXTENT_GET_BIT_MAP                      16          1          0          0          0          0          0          0
  EXTENT_SET_BIT_MAP                      17          1          0          0          0          0          0          0
  EXTENT_GET_PAGE                         18          1          0          0          0          0          0          0
  EXTENT_STOP_INIT                        19          1          0          0          0          0          0          0
  TGT_WRITE_ALLOC                         20          1          0          0          0          0          0          0
  TGT_WRITE_DATA                          21          1          0          0          0          0          0          0
```

HUAWEI

NAS Performance Statistics

- show nfs nfsv3_stat
- show nfs nfsv4_stat
- show cifs smb1_stat
- show cifs smb2_stat

```
developer:/>show nfs nfsv3_stat
Procedure      Counts      Success     Failure     Drop        Maximal Delay   Average Delay   File System Maximal Delay   File System Average Delay
NULL           0           0           0           0           0               0               0                           0
GETATTR        0           0           0           0           0               0               0                           0
SETATTR        0           0           0           0           0               0               0                           0
LOOKUP         0           0           0           0           0               0               0                           0
ACCESS         0           0           0           0           0               0               0                           0
READLINK       0           0           0           0           0               0               0                           0
READ           0           0           0           0           0               0               0                           0
WRITE          0           0           0           0           0               0               0                           0
CREATE         0           0           0           0           0               0               0                           0
MKDIR          0           0           0           0           0               0               0                           0
SYMLINK        0           0           0           0           0               0               0                           0
MKNOD          0           0           0           0           0               0               0                           0
REMOVE         0           0           0           0           0               0               0                           0
RMDIR          0           0           0           0           0               0               0                           0
RENAME         0           0           0           0           0               0               0                           0
LINK           0           0           0           0           0               0               0                           0
READDIR        0           0           0           0           0               0               0                           0
READDIRPLUS    0           0           0           0           0               0               0                           0
FSSTAT         0           0           0           0           0               0               0                           0
FSINFO         0           0           0           0           0               0               0                           0
PATHCONF       0           0           0           0           0               0               0                           0
COMMIT         0           0           0           0           0               0               0                           0
extendcmd      0           0           0           0           0               0               0                           0
ACL_NULL       0           0           0           0           0               0               0                           0
GETFACL        0           0           0           0           0               0               0                           0
SETFACL        0           0           0           0           0               0               0                           0
```

HUAWEI

Contents

Block Performance Tuning

- Stripe Depth of Storage Pool
  - Default 128KB applicable for random or small IO access, like VDI, OLTP, etc.
  - 512KB applicable for backup and video surveillance solution, etc.
- RAID Level
  - In RAID 2.0+, RAID level affects entire IOPS or bandwidth, but no big impact on common latency
- LUN Prefetch Policy
  - No Prefetch: Applicable for full random IO access, reduce system resource consumption
  - Intelligent Prefetch(default): Mixed application
  - Constant Prefetch/ Variable Prefetch : Single application and sequence read, recommend to prefetch size based on IO model

Filesystem Performance Tuning

- Reserve enough SSD space in storage pool, as the metadata allocate from SSD tier in higher priority
- Proper filesystem size, huge filesystem may reduce the read performance.
- Enable Lease V2 for CIFS
- Use Kerberos instead of NTLM

HUAWEI

Thank you.

Bring digital to every person, home and organization for a fully connected, intelligent world.