HDD IO Priority: Challenges and Questions

Yong Jiang
Storage Engineer

Oct-28-2019
Current Implementation

• Keep the queue low to achieve target SLA

• Command priority can be managed at application queue or scheduler queue, not in the HDD queue.

• This limits our capability to ensure end to end QoS for HDD
Priority Levels, SAS and SATA and vendors

• How many levels do we need?
  • “m to n” mapping instead of “m to 2”?
• Implementation needs to be consistent
  • SAS and SATA
    • Embedded in read/write IO command for SATA: 2 levels
    • SAS Frame priority? Multiple levels?
• Will this post challenge for implementing multiple priority levels at HDD FW?
• Across vendors and capacity point?

The priority is specified in the PRI0 field for SATA NCQ commands:
READ FPDMA QUEUED
WRITE FPDMA QUEUED

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>00b</td>
<td>Normal priority</td>
</tr>
<tr>
<td>01b</td>
<td>Isochronous deadline-dependent priority The device should complete isochronous requests prior to their associated deadline.</td>
</tr>
<tr>
<td>10b</td>
<td>High priority The device should attempt to provide better quality of service for the command. The device should complete high priority requests in a more timely fashion than normal and isochronous requests.</td>
</tr>
<tr>
<td>11b</td>
<td>Reserved</td>
</tr>
</tbody>
</table>
Latency target management

• Deadline combined with fast fail
  • If the host set a timeout limit (or latency target), can the HDD smartly tell if it can serve it or fast fail it?
  • Fast fail write?

• Multiple queue implementation?

• Does it make sense for SW/HW folks to co-design HDD queuing algorithm development?
  • Most of the time, SW engineers like simply HW implementation.