Performance monitoring of MetroCluster configurations

Active IQ Unified Manager 9.7

NetApp
June 10, 2024

Table of Contents

Performance monitoring of MetroCluster configurations

Unified Manager enables you to monitor the write throughput between clusters in a MetroCluster configuration to identify workloads with a high amount of write throughput. If these high-performing workloads are causing other volumes on the local cluster to have high I/O response times, Unified Manager triggers performance events to notify you.

When a local cluster in a MetroCluster configuration mirrors its data to its partner cluster, the data is written to NVRAM and then transferred over the interswitch links (ISLs) to the remote aggregates. Unified Manager analyzes the NVRAM to identify the workloads whose high write throughput is overutilizing the NVRAM, placing the NVRAM in contention.

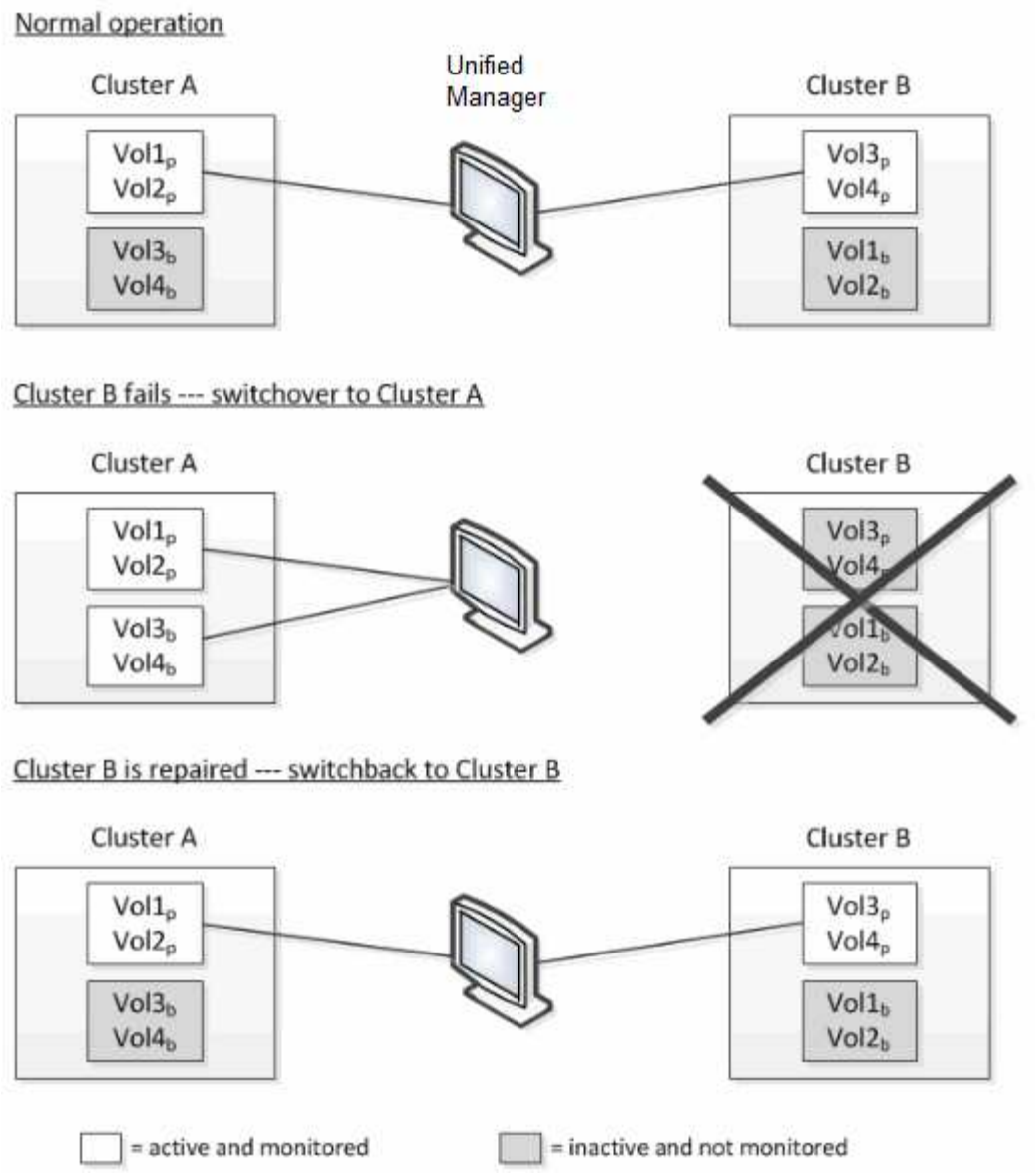
Workloads whose deviation in response time has exceeded the performance threshold are called *victims* and workloads whose deviation in write throughput to the NVRAM is higher than usual, causing the contention, are called *bullies*. Because only the write requests are mirrored to the partner cluster, Unified Manager does not analyze read throughput.

Unified Manager treats the clusters in a MetroCluster configuration as individual clusters. It does not distinguish between clusters that are partners or correlate the write throughput from each cluster.

Volume behavior during switchover and switchback

Events that trigger a switchover or switchback cause active volumes to be moved from one cluster to the other cluster in the disaster recovery group. The volumes on the cluster that were active and serving data to clients are stopped, and the volumes on the other cluster are activated and start serving data. Unified Manager monitors only those volumes that are active and running.

Because volumes are moved from one cluster to another, it is recommended that you monitor both clusters. A single instance of Unified Manager can monitor both clusters in a MetroCluster configuration, but sometimes the distance between the two locations necessitates using two Unified Manager instances to monitor both clusters. The following figure shows a single instance of Unified Manager:



The volumes with p in their names indicate the primary volumes, and the volumes with b in their names are mirrored backup volumes that are created by SnapMirror.

During normal operation:

- Cluster A has two active volumes: Vol1p and Vol2p.
- Cluster B has two active volumes: Vol3p and Vol4p.
- Cluster A has two inactive volumes: Vol3b and Vol4b.
- Cluster B has two inactive volumes: Vol1b and Vol2b.

Information pertaining to each of the active volumes (statistics, events, and so on) is collected by Unified Manager. Vol1p and Vol2p statistics are collected by Cluster A, and Vol3p and Vol4p statistics are collected by Cluster B.

After a catastrophic failure causes a switchover of active volumes from Cluster B to Cluster A:

- Cluster A has four active volumes: Vol1p, Vol2p, Vol3b, and Vol4b.

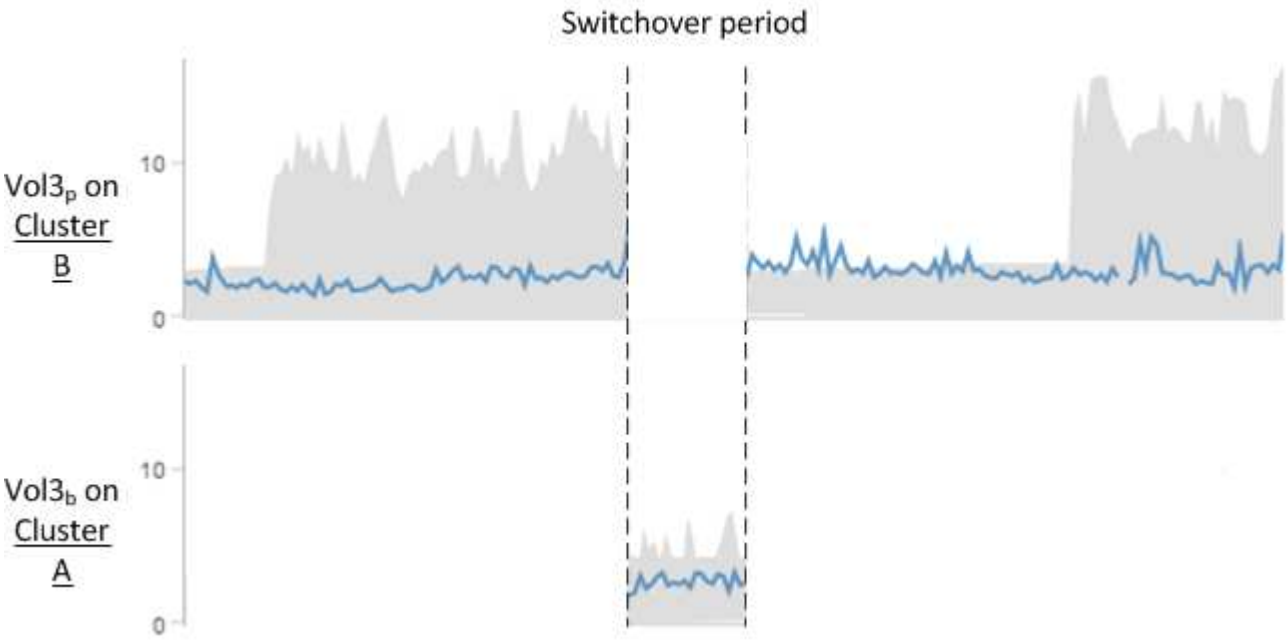
- Cluster B has four inactive volumes: Vol3p, Vol4p, Vol1b, and Vol2b.

As during normal operation, information pertaining to each of the active volumes is collected by Unified Manager. But in this case, Vol1p and Vol2p statistics are collected by Cluster A, and Vol3b and Vol4b statistics are also collected by Cluster A.

Note that Vol3p and Vol3b are not the same volumes, because they are on different clusters. The information in Unified Manager for Vol3p is not the same as Vol3b:

- During switchover to Cluster A, Vol3p statistics and events are not visible.
- On the very first switchover, Vol3b looks like a new volume with no historical information.

When Cluster B is repaired and a switchback is performed, Vol3p is active again on Cluster B, with the historical statistics and a gap of statistics for the period during the switchover. Vol3b is not viewable from Cluster A until another switchover occurs:



- MetroCluster volumes that are inactive, for example, Vol3b on Cluster A after switchback, are identified with the message “This volume was deleted”. The volume is not actually deleted, but it is not currently being monitored by Unified Manager because it is not the active volume.
- If a single Unified Manager is monitoring both clusters in a MetroCluster configuration, volume search returns information for whichever volume is active at that time. For example, a search for “Vol3” would return statistics and events for Vol3b on Cluster A if a switchover has occurred and Vol3 has become active on Cluster A.

Performance event analysis and notification

Performance events notify you about I/O performance issues on a workload caused by contention on a cluster component. Unified Manager analyzes the event to identify all workloads involved, the component in contention, and whether the event is still an issue that you might need to resolve.

Unified Manager monitors the I/O latency (response time) and IOPS (operations) for volumes on a cluster. When other workloads overuse a cluster component, for example, the component is in contention and cannot perform at an optimal level to meet workload demands. The performance of other workloads that are using the same component might be impacted, causing their latencies to increase. If the latency crosses the dynamic performance threshold, Unified Manager triggers a performance event to notify you.

Event analysis

Unified Manager performs the following analyses, using the previous 15 days of performance statistics, to identify the victim workloads, bully workloads, and the cluster component involved in an event:

- Identifies victim workloads whose latency has crossed the dynamic performance threshold, which is the upper boundary of the latency forecast:
  - For volumes on HDD or Flash Pool (hybrid) aggregates (local tier), events are triggered only when the latency is greater than 5 milliseconds (ms) and the IOPS are more than 10 operations per second (ops/sec).
  - For volumes on all-SSD aggregates or FabricPool aggregates (cloud tier), events are triggered only when the latency is greater than 1 ms and the IOPS are more than 100 ops/sec.
- Identifies the cluster component in contention.

  If the latency of victim workloads at the cluster interconnect is greater than 1 ms, Unified Manager treats this as significant and triggers an event for the cluster interconnect.

- Identifies the bully workloads that are overusing the cluster component and causing it to be in contention.
- Ranks the workloads involved, based on their deviation in utilization or activity of a cluster component, to determine which bullies have the highest change in usage of the cluster component and which victims are the most impacted.

An event might occur for only a brief moment and then correct itself after the component it is using is no longer in contention. A continuous event is one that reoccurs for the same cluster component within a five-minute interval and remains in the active state. For continuous events, Unified Manager triggers an alert after detecting the same event during two consecutive analysis intervals.

When an event is resolved, it remains available in Unified Manager as part of the record of past performance issues for a volume. Each event has a unique ID that identifies the event type and the volumes, cluster, and cluster components involved.

A single volume can be involved in more than one event at the same time.

Event state

Events can be in one of the following states:

- **Active**

  Indicates that the performance event is currently active (new or acknowledged). The issue causing the event has not corrected itself or has not been resolved. The performance counter for the storage object remains above the performance threshold.

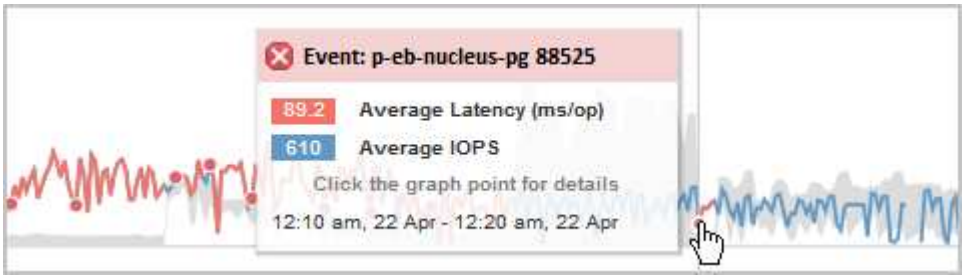
- **Obsolete**

  Indicates that the event is no longer active. The issue causing the event has corrected itself or has been

resolved. The performance counter for the storage object is no longer above the performance threshold.

Event notification

The events are displayed on the Dashboard page and on many other pages in the user interface, and alerts for those events are sent to specified email addresses. You can view detailed analysis information about an event and get suggestions for resolving it on the Event details page and on the Workload Analysis page.



In this example, an event is indicated by a red dot (🔴) on the Latency chart. Hovering your mouse cursor over the red dot displays a popup with more details about the event and options for analyzing it.

Event interaction

On the Event details page and on the Workload Analysis page, you can interact with events in the following ways:

- Moving the mouse over an event displays a message that shows the event ID and the date and time when the event was detected.

  If there are multiple events for the same time period, the message shows the number of events.

- Clicking a single event displays a dialog box that shows more detailed information about the event, including the cluster components that are involved.

  The component in contention is circled and highlighted red. You can click either the event ID or **View full analysis** to view the full analysis on the Event details page. If there are multiple events for the same time period, the dialog box shows details about the three most recent events. You can click an event ID to view the event analysis on the Event details page.

How Unified Manager determines the performance impact for an event

Unified Manager uses the deviation in activity, utilization, write throughput, cluster component usage, or I/O latency (response time) for a workload to determine the level of impact to workload performance. This information determines the role of each workload in the event and how they are ranked on the Event details page.

Unified Manager compares the last analyzed values for a workload to the expected range (latency forecast) of values. The difference between the values last analyzed and the expected range of values identifies the workloads whose performance was most impacted by the event.

For example, suppose a cluster contains two workloads: Workload A and Workload B. The latency forecast for Workload A is 5-10 milliseconds per operation (ms/op) and its actual latency is usually around 7 ms/op. The

latency forecast for Workload B is 10-20 ms/op and its actual latency is usually around 15 ms/op. Both workloads are well within their latency forecast. Due to contention on the cluster, the latency of both workloads increases to 40 ms/op, crossing the dynamic performance threshold, which is the upper bounds of the latency forecast, and triggering events. The deviation in latency, from the expected values to the values above the performance threshold, for Workload A is around 33 ms/op, and the deviation for Workload B is around 25 ms/op. The latency of both workloads spike to 40 ms/op, but Workload A had the bigger performance impact because it had the higher latency deviation at 33 ms/op.

On the Event details page, in the System Diagnosis section, you can sort workloads by their deviation in activity, utilization, or throughput for a cluster component. You can also sort workloads by latency. When you select a sort option, Unified Manager analyzes the deviation in activity, utilization, throughput, or latency since the event was detected from the expected values to determine the workload sort order. For the latency, the red dots (🔴) indicate a performance threshold crossing by a victim workload, and the subsequent impact to the latency. Each red dot indicates a higher level of deviation in latency, which helps you identify the victim workloads whose latency was impacted the most by an event.

Cluster components and why they can be in contention

You can identify cluster performance issues when a cluster component goes into contention. The performance of workloads that use the component slow down and their response time (latency) for client requests increases, which triggers an event in Unified Manager.

A component that is in contention cannot perform at an optimal level. Its performance has declined, and the performance of other cluster components and workloads, called *victims*, might have increased latency. To bring a component out of contention, you must reduce its workload or increase its ability to handle more work, so that the performance can return to normal levels. Because Unified Manager collects and analyzes workload performance in five-minute intervals, it detects only when a cluster component is consistently overused. Transient spikes of overusage that last for only a short duration within the five-minute interval are not detected.

For example, a storage aggregate might be under contention because one or more workloads on it are competing for their I/O requests to be fulfilled. Other workloads on the aggregate can be impacted, causing their performance to decrease. To reduce the amount of activity on the aggregate, there are different steps you can take, such as moving one or more workloads to a less busy aggregate or node, to lessen the overall workload demand on the current aggregate. For a QoS policy group, you can adjust the throughput limit, or move workloads to a different policy group, so that the workloads are no longer being throttled.

Unified Manager monitors the following cluster components to alert you when they are in contention:

- **Network**

  Represents the wait time of I/O requests by the external networking protocols on the cluster. The wait time is time spent waiting for “transfer ready” transactions to finish before the cluster can respond to an I/O request. If the network component is in contention, it means high wait time at the protocol layer is impacting the latency of one or more workloads.

- **Network Processing**

  Represents the software component in the cluster involved with I/O processing between the protocol layer and the cluster. The node handling network processing might have changed since the event was detected. If the network processing component is in contention, it means high utilization at the network processing node is impacting the latency of one or more workloads.

  When using an All SAN Array cluster in an active-active configuration, the network processing latency

value is displayed for both nodes so you can verify the nodes are sharing the load equally.

- **QoS Limit Max**

  Represents the throughput maximum (peak) setting of the storage Quality of Service (QoS) policy group assigned to the workload. If the policy group component is in contention, it means all workloads in the policy group are being throttled by the set throughput limit, which is impacting the latency of one or more of those workloads.

- **QoS Limit Min**

  Represents the latency to a workload that is being caused by QoS throughput minimum (expected) setting assigned to other workloads. If the QoS minimum set on certain workloads use the majority of the bandwidth to guarantee the promised throughput, other workloads will be throttled and see more latency.

- **Cluster Interconnect**

  Represents the cables and adapters with which clustered nodes are physically connected. If the cluster interconnect component is in contention, it means high wait time for I/O requests at the cluster interconnect is impacting the latency of one or more workloads.

- **Data Processing**

  Represents the software component in the cluster involved with I/O processing between the cluster and the storage aggregate that contains the workload. The node handling data processing might have changed since the event was detected. If the data processing component is in contention, it means high utilization at the data processing node is impacting the latency of one or more workloads.

- **Volume Activation**

  Represents the process that tracks the usage of all active volumes. In large environments where more than 1000 volumes are active, this process tracks how many critical volumes need to access resources through the node at the same time. When the number of concurrent active volumes exceeds the recommended maximum threshold, some of the non-critical volumes will experience latency as identified here.

- **MetroCluster Resources**

  Represents the MetroCluster resources, including NVRAM and interswitch links (ISLs), used to mirror data between clusters in a MetroCluster configuration. If the MetroCluster component is in contention, it means high write throughput from workloads on the local cluster or a link health issue is impacting the latency of one or more workloads on the local cluster. If the cluster is not in a MetroCluster configuration, this icon is not displayed.

- **Aggregate or SSD Aggregate Ops**

  Represents the storage aggregate on which the workloads are running. If the aggregate component is in contention, it means high utilization on the aggregate is impacting the latency of one or more workloads. An aggregate consists of all HDDs, or a mix of HDDs and SSDs (a Flash Pool aggregate). An “SSD Aggregate” consists of all SSDs (an all-flash aggregate), or a mix of SSDs and a cloud tier (a FabricPool aggregate).

- **Cloud Latency**

  Represents the software component in the cluster involved with I/O processing between the cluster and the cloud tier on which user data is stored. If the cloud latency component is in contention, it means that a large amount of reads from volumes that are hosted on the cloud tier are impacting the latency of one or

more workloads.

- **Sync SnapMirror**

  Represents the software component in the cluster involved with replicating user data from the primary volume to the secondary volume in a SnapMirror Synchronous relationship. If the sync SnapMirror component is in contention, it means that the activity from SnapMirror Synchronous operations are impacting the latency of one or more workloads.

Roles of workloads involved in a performance event

Unified Manager uses roles to identify the involvement of a workload in a performance event. The roles include victims, bullies, and sharks. A user-defined workload can be a victim, bully, and shark at the same time.

Role	Description
Victim	A user-defined workload whose performance has decreased due to other workloads, called bullies, that are over-using a cluster component. Only user-defined workloads are identified as victims. Unified Manager identifies victim workloads based on their deviation in latency, where the actual latency, during an event, has greatly increased from its latency forecast (expected range).
Bully	A user-defined or system-defined workload whose over-use of a cluster component has caused the performance of other workloads, called victims, to decrease. Unified Manager identifies bully workloads based on their deviation in usage of a cluster component, where the actual usage, during an event, has greatly increased from its expected range of usage.
Shark	A user-defined workload with the highest usage of a cluster component compared to all workloads involved in an event. Unified Manager identifies shark workloads based on their usage of a cluster component during an event.

Workloads on a cluster can share many of the cluster components, such as aggregates and the CPU for network and data processing. When a workload, such as a volume, increases its usage of a cluster component to the point that the component cannot efficiently meet workload demands, the component is in contention. The workload that is over-using a cluster component is a bully. The other workloads that share those components, and whose performance is impacted by the bully, are the victims. Activity from system-defined workloads, such as deduplication or Snapshot copies, can also escalate into “bullying”.

When Unified Manager detects an event, it identifies all workloads and cluster components involved, including the bully workloads that caused the event, the cluster component that is in contention, and the victim workloads whose performance has decreased due to the increased activity of bully workloads.

If Unified Manager cannot identify the bully workloads, it only alerts on the victim workloads and the cluster component involved.

Unified Manager can identify workloads that are victims of bully workloads, and also identify when those same workloads become bully workloads. A workload can be a bully to itself. For example, a high-performing workload that is being throttled by a policy group limit causes all workloads in the policy group to be throttled, including itself. A workload that is a bully or a victim in an ongoing performance event might change its role or no longer be a participant in the event.

Copyright information

Trademark information