

Root Cause Analysis: AZLCL01-CL, node AZL-N01 removed from the cluster on 1 October 2026

Cluster: AZLCL01-CL (Azure Local 2609, two nodes) Incident window analysed: 1 October 2026 01:40 UTC to 2 October 2026 01:40 UTC Evidence package: AZLCL01-CL_ClusterIssues_20261002_0146.zip (ClusterIssueIdentifier 0.27, Get-IncidentEvidence 0.59, CloudLabs.HealthCheck 0.92), collected on 2 October 2026 at 01:25 to 01:46 UTC Report date: 2 October 2026 All times are UTC. References such as E2463 are evidence IDs in the incident evidence file.

1. Summary

The operator reported: "Cluster node crashed unexpectedly".

What the evidence shows: **neither node crashed or rebooted**. On 1 October at 06:41:41, AZL-N02 removed AZL-N01 from the cluster membership because N01 had stopped answering cluster heartbeats on all three networks at the same time. Thirty seconds later the Cluster service on N01 stopped itself with a fatal error ("Changing Epoch failed"), the infrastructure volume was paused, and all clustered roles ran on N02. N01 rejoined the cluster at 06:43:02, about 80 seconds after it was removed.

Most probable root cause: a live kernel memory dump was being captured on AZL-N01. It started at 06:40:49 and suspended the node for longer than the cluster's heartbeat tolerance of 25 seconds. The heartbeats stopped in the same second the dump capture began, and N02 removed N01 in the same second the dump's final capture phase ended. No other event explains the moment, and the alternatives (network, hardware, storage, operating-system crash) are contradicted by the evidence.

What is not proven: who or what started the live dump, and why. The package does not contain that information (see section 7).

Confidence: **high** that the live dump capture caused the loss of membership; **unknown** for the reason the dump was taken.

2. What the operator reported versus what was measured

Statement	Evidence	Verdict
"A cluster node crashed"	Last boot of AZL-N01 30 September 15:48, of AZL-N02 30 September 15:45 (inventory, section "Caption / LastBootUpTime"). No event 41, 6008 or 1074 in the window (System/RebootShutdown, read). No crash dump or WER report found on either node (dump folders checked: 3 per node, files found: 0).	Not supported. The operating system kept running.
The cluster lost a node	N02 System 1135 at 06:41:41 (E2463); N01 Cluster service terminated 06:42:12 (E2572, E2576, E2577).	Supported. It was a loss of cluster membership and a Cluster service failure, not a node crash.

3. Timeline

Time (UTC)	Node	What happened	Evidence
06:39:36 to 06:40:14	both	Normal SMB client activity on both nodes (event 30952 every few seconds)	SMBCClient/Operational
06:40:14	N01	Last ordinary event logged by N01 before the gap	timeline
06:40:49	N01	"Live Dump Capture Dump Data API started" , followed in the same second by the sizing workflow (events 102, 108 to 111)	E2304 to E2313 , Kernel-LiveDump/Analytic
~06:40:49.7	N02	Last heartbeat response received from N01 on the management route. At 06:41:09.335 the route history records that response as 19.65 seconds old.	N02 cluster log, NETFT route history
06:41:00	N02	NetFT: more than 40% of consecutive heartbeats from N01 missed, on all three routes (management, Storage1, Storage2)	N02 cluster log and FailoverClustering/Operational 1650
06:41:10	N01	Live dump: sizing, mirroring phase 1 ended	E2318
06:41:11	both	"Lost communication ... reestablished" (1592), a reconnect with a new epoch	System, both cluster logs
06:41:15	N01	Live dump: capture pages workflow, mirroring started and phase 0 ended	E2319, E2320
06:41:22 to 06:41:39	N02	SMB resilient handles to N01 time out and cannot reconnect (30906, 30612, more than 100 events)	E2321 to E2440
06:41:39 to 06:41:40	N02	UDP heartbeat connections to N01 lost on all three routes	N02 cluster log
06:41:41	N01	Live dump: "Capture Pages Workflow: Mirroring Phase 1 ended" (last live dump event)	E2445
06:41:41.505	N02	"Cluster node 'AZL-N01' was removed from the active failover cluster membership" (1135); N01 enters the isolated state; its VMs become unmonitored	E2463 , N02 cluster log
06:41:41 to 06:41:50	N02	N02 brings up the cluster groups (Cluster Group through the Cloud Witness, SDDC, Orchestrator, Cloud Management)	N02 cluster log 1674, 1641, 1637
06:41:44	N01	CSV Infrastructure_1 paused (5120)	E2466
06:41:53 to 06:41:54	N02	Infrastructure_1: ownership unexpectedly lost (1038), resource failed (1069), ReFS metadata write failures (134), CSV paused (5120)	E2511 to E2523
06:41:55	N01	N01's own NetFT reports missed heartbeats from N02. The route history on N01 shows its last received heartbeat as 25.98 seconds old.	N01 cluster log
06:42:10	N01	N01 declares N02 removed from membership (1135)	E2559
06:42:12	N01	Cluster service terminated: "Changing Epoch failed, Error Code 3342385" (5417, 7024, 7031), Clusapi handles dead, ReFS write failure	E2572 to E2579

Time (UTC)	Node	What happened	Evidence
06:42:43	N01	Azure Stack HCI Orchestrator Service terminated unexpectedly (7034)	E2656
06:43:02	both	N01 exits the isolated state and rejoins the cluster (1680, 1682)	both cluster logs

4. Analysis: the hypotheses tested

H1. The node was suspended by a live kernel dump capture (supported)

- The Kernel-LiveDump channel on N01 shows a capture that started at 06:40:49 and whose last recorded phase ended at 06:41:41 (E2304 to E2445).
- The last heartbeat response N02 received from N01 is at about 06:40:49.7. That is derived from the route history N02 logged at 06:41:09.335, which records the response as 19.65 s old. It matches the start of the capture to within a second.
- N02 removed N01 at 06:41:41.505, the same second the capture's final phase ended. The heartbeat threshold is 25 missed heartbeats at a 1-second interval (SameSubnetThreshold 25, SameSubnetDelay 1000, from the collected settings). A gap of about 52 seconds is well beyond it.
- Between 06:40:49 and 06:41:41, N01 logged nothing except the live dump's own events. Its last ordinary event was at 06:40:14.
- N01 only noticed the missed heartbeats itself at 06:41:55, after the capture. Its route history then shows heartbeats 26 seconds old. That is how a node behaves that was paused rather than disconnected: it has no sense of the time that passed until it resumes.

Taken together: the node was not reachable for the duration of the capture, and the cluster acted exactly as configured.

H2. A network failure between the nodes (contradicted)

- All three routes (management, Storage1, Storage2) failed at the same moment and in both directions. These run over separate adapters.
- NDIS/Operational was read on both nodes and contains no events in the window. No link-down or adapter reset was recorded.
- N01 itself did not detect a network problem until after the live dump capture ended.

A network fault would not stop all three independent routes and leave no adapter event.

H3. Hardware fault (contradicted)

WHEA/System was read on both nodes: no events. No disk, storage-path or StorageSpaces-Driver events in the window.

H4. Storage failure as the trigger (contradicted)

The CSV pause, the ReFS write failures and the loss of disk ownership all come after the membership loss (06:41:44 and later). They are consequences of losing N01, not its cause. Storage channels show nothing before 06:41:41.

H5. Operating-system crash or reboot (contradicted)

See section 2: no reboot, no bugcheck, no crash dump.

H6. Resource exhaustion on the node (possible contributing factor, not provable from this package)

- Each node has 4 logical processors with SMT disabled (inventory, CPU section). Both nodes are virtual machines: model "Virtual Machine", Microsoft Corporation.
- N01 also hosts the Arc Resource Bridge VM, with 4 vCPUs and 8 GB, running and owned by N01.
- At collection time N01 had 7.9 GB of 32 GB memory available, against 15.1 GB on N02.
- A live dump on a small, busy node takes longer, so the suspension is longer. The package contains no processor or memory counters from 06:40 to 06:42, so how busy N01 was cannot be shown.

5. Root cause

Direct cause: a live kernel memory dump capture on AZL-N01, from 06:40:49 to about 06:41:41, suspended the node for about 52 seconds. That exceeded the 25-second heartbeat tolerance. AZL-N02 removed N01 from the membership, and the Cluster service on N01 terminated itself when it resumed into a membership that had moved on ("Changing Epoch failed").

Underlying cause: unknown. The evidence does not show what requested the live dump:

- **Event content:** event 1's flags field is not rendered (Flags: %1).
- **Dump file:** no live dump file was found in the standard folders on either node (C:\Windows\LiveKernelReports and the WER folders: 0 files). So the dump was either written to a non-standard location, or not kept.
- **Inference:** a dump that Windows triggers itself is normally kept in LiveKernelReports. Its absence points towards a dump that was **requested on purpose, with its own destination**, for example by an administrator or a diagnostic tool. That is an inference, not a measurement.

Contributing conditions (likely, not proven):

- small nodes, with 4 logical processors each;
- the Arc Resource Bridge VM on the node that was being dumped;
- a virtualised platform, which makes the suspension last longer.

6. Impact

- **Membership:** N01 was out of the cluster from 06:41:41 to 06:43:02, about 80 seconds.
- **Storage:** CSV Infrastructure_1 was paused and its disk resource failed and restarted on N02. ReFS reported metadata write failures on both nodes.
- **Roles:** all cluster groups were brought online on N02. VMs on N01 were unmonitored while N01 was isolated.
- **Services:** the Cluster service on N01 restarted, and the Azure Stack HCI Orchestrator Service on N01 terminated unexpectedly.
- **Not measured:** lasting damage. The package was collected 19 hours later and shows both nodes as members. Whether workloads were affected is not visible.

7. What is missing for a better report

Missing	Why it matters	Where to get it
Who or what started the live dump	It decides between "administrator action", "diagnostic tool" and "system-triggered", which is the actual root cause	Security log 4688 (process creation, if auditing is on), PowerShell/Operational and script block logging (4103/4104), Microsoft-Windows-Storage-* management logs, and the operator's own record of what was run at 06:40
The live dump file	Its location, size and the triggering component in the dump header	The destination folder used by the requesting tool; ask the operator
Rendered flags of LiveDump event 1	Shows the type of dump and the requesting interface	Re-read Kernel-LiveDump/Analytic with the message rendered on the node
CPU and memory counters for N01 around 06:40 to 06:42	Needed to judge H6 and why the suspension lasted 52 s	Performance counters or the hypervisor host's metrics (the nodes are VMs)
Host-side data of the virtualisation platform	The nodes are virtual machines; a pause of the VM itself would look the same from inside	The host's VM CPU and scheduling history for the two node VMs
The minutes before 06:42 in Health/Operational on N02	The log rolled and starts at 06:42:05, so the lead-up was lost	Larger channel size (see recommendation)
SMBClient/Connectivity on N01 before 04:54	Rolled; not relevant for this incident but a gap in the evidence	Larger channel size
The second Kernel-LiveDump record on N02 at 06:57:38 (event ID 0, no message)	Unexplained; it could be a second dump request or a channel artefact	Re-read the event with its full XML
The operator's timeline	What was being done on the cluster at 06:40	Interview

8. Recommendations

1. **Find out who requested the live dump** (section 7, first row), before closing this incident.
2. **Do not take live kernel dumps on a cluster node in production** without first pausing or draining it (`Suspend-ClusterNode -Drain`). On small nodes the capture can exceed the heartbeat tolerance and cost the node its membership.
3. **Size the nodes for their workload.** Four logical processors with the Arc Resource Bridge VM on the same node leaves little headroom.
4. **Enlarge the event channels that rolled** (Health/Operational, SMBClient/Connectivity) so that the minutes before an incident survive. Process-creation auditing would answer the main open question of this report next time.
5. **Do not raise the heartbeat thresholds to hide this.** The cluster did the right thing with a node that did not answer.

9. Separate observation, not part of this incident

From 2 October 00:43 on N01, and from 01:16 also on N02, the SDDC Management resource logs "Failed to get dedup jobs. Exception: Invalid namespace" every 30 seconds (FailoverClustering/Diagnostic 2051, for example E18522 and E21728). This started about 18 hours after the incident and has no visible link to it. It is worth a separate look.

10. Method and limits

- **Sources:** the incident evidence (27 channels, 17 read, 2 rolled and 8 absent or disabled; 29,046 timeline events), the cluster logs of both nodes (77 and 112 MB, covering the incident), and the health-check inventory.
- **Clock skew:** none was measured (the skew table is empty).
- **Not available:** CPU or memory history, and anything from outside the two nodes.
- Every statement in this report names its evidence. Where a statement is an inference, it says so.