| OpsLab | Infrastructure Audit Report |
|---|---|

Generated: 2026-06-01 07:28 UTC | Storage Infrastructure / Data Center | Confidential — for authorised recipients only

**CRITICAL ISSUES DETECTED**

| 1057 | 1270 min | 22363.9 h |
|---|---|---|
| Incidents | Avg. Resolution | Downtime |

## Summary

Over 88 days, 1057 critical drive failures accumulated 22363.87 hours of total downtime with a median repair time of 1345 minutes. All failures cluster at 02:00, and co-occurring drive failures suggest shared infrastructure triggers.

Incident rate decreased after December 1 but repair times remain extreme, with no incident resolved within 30 minutes.

## Key Issues

R1. Extreme mean time to repair across all drive failures — CRITICAL

R2. Correlated drive failures within 5-minute windows — CRITICAL

R3. Downtime moderately concentrated in top incidents — HIGH

## Recommended Actions

1) Deploy pre-staged hot spares for top-failing models [Immediate]

2) Inspect shared backplane and power for correlated failures [Immediate]

3) Implement proactive SMART-based drive retirement [Short-term]

---

Full technical analysis on the next page.

# TECHNICAL REPORT

## Risk Analysis

| # | Issue | Severity | Downtime |
|---|---|---|---|
| R1 | **R1 — Extreme mean time to repair across all drive failures**<br>No single failure was resolved quickly, meaning every drive loss degrades array redundancy for an entire day.<br>(2023-10-11 02:00:00) (2023-10-02 02:00:00) (2023-12-14 02:00:00)<br>services: WDC WUH722222ALE6L4, TOSHIBA MG08ACA16TA, TOSHIBA MG08ACA16TEY<br>MTTR mean=1270min with median=1345min and P95=2061min; 0% of incidents resolved within 30 minutes. This likely indicates insufficient hot-spare inventory or manual-only replacement workflows that delay restoration. Worst incident lasted 2512min (2023-10-11 02:00:00). | CRITICAL | 22363.9h |
| R2 | **R2 — Correlated drive failures within 5-minute windows**<br>Simultaneous multi-drive loss can exceed RAID parity protection and cause data loss.<br>services: ST8000NM0055, ST12000NM0008, HGST HUH721212ALN604, ST4000DM000<br>ST8000NM0055 → ST12000NM0008 → HGST HUH721212ALN604<br>Apriori-like lift analysis: ST8000NM0055 → ST12000NM0008 in 5-min window, lift=201.14, support=8.1%, n=86 (confidence 100%). ST8000NM0055 → HGST HUH721212ALN604, lift=206.03, support=7.5%, n=79. These simultaneous failures are consistent with shared backplane, power, or vibration faults triggering multi-drive loss. | CRITICAL | — |
| R3 | **R3 — Downtime moderately concentrated in top incidents**<br>Fixing a few bad drives will not help; the entire fleet contributes nearly equally to downtime.<br>(2023-10-11 02:00:00) (2023-10-02 02:00:00) (2023-12-14 02:00:00)<br>services: WDC WUH722222ALE6L4, TOSHIBA MG08ACA16TA, TOSHIBA MG08ACA16TEY<br>Gini=0.218, concentration=moderate; 65.0% of incidents cover 80% of downtime. The top 3 incidents account for only 0.5% of total downtime, meaning the problem is systemic rather than caused by outliers. This is consistent with a fleet-wide aging or environmental issue. | HIGH | — |
| R4 | **R4 — Incident rate decreased after December 1 but remains high**<br>A small improvement signals the root cause persists and further action is needed to prevent regression.<br>non-parametric comparison test (Mann-Whitney U): p=0.03286, effect size (Cliff's δ)=-0.283 (95% CI [-0.515, -0.049]); median daily incident count dropped from 12.000 to 10.000 (−16.7%), effect=small. The decrease is statistically significant but practically modest, likely reflecting partial mitigation or seasonal load reduction rather than a root-cause fix. | MEDIUM | — |

| # | Issue | Severity | Downtime |
|---|---|---|---|
| R5 | **R5 — Light tail risk but near-certain prolonged outages**<br>Nearly every failure guarantees over 6 hours of degraded storage, risking SLA breaches on every incident.<br>GPD fit (Peaks-Over-Threshold): P(downtime > 6.0 h) = 93.85%, 95% CI [92.34%, 95.17%]; tail=light. Despite a light tail (shape xi=-0.354, 95% CI [-1.134, -0.264]), the baseline durations are so high that almost every incident exceeds 6 hours. This is likely driven by logistics constraints on physical drive replacement. | HIGH | — |

## Event Chains

Apriori-like lift analysis (5-min co-occurrence window)

| Event A | → Event B | Lift | Support | Confidence | n |
|---|---|---|---|---|---|
| ST8000NM0055 | TOSHIBA MG08ACA16TA | 217.30 | 6.9% | 84.9% | 73 |
| HGST HUH721212ALN604 | ST4000DM000 | 207.40 | 8.1% | 76.1% | 86 |
| ST8000NM0055 | HGST HUH721212ALN604 | 206.03 | 7.5% | 91.9% | 79 |
| ST8000NM0055 | ST12000NM0008 | 201.14 | 8.1% | 100.0% | 86 |
| TOSHIBA MG07ACA14TA | TOSHIBA MG08ACA16TA | 198.47 | 6.5% | 77.5% | 69 |

## Tail Risk — Probability of a Very Long Outage

Peaks-Over-Threshold, Generalized Pareto Distribution (GPD) fit on long incidents

GPD shape $\xi$ = -0.354 (95% CI [-1.134, -0.264]); fitted on 105 events longer than 31.1 h. Tail behaviour: light — very long outages are not more likely than an exponential model implies.

| Outage longer than | Probability | 95% CI |
|---|---|---|
| 1 h | 100.00% | [100.00%, 100.00%] |
| 2 h | 100.00% | [100.00%, 100.00%] |
| 6 h | 93.85% | [92.34%, 95.17%] |

## Recommendations

### 1) Deploy pre-staged hot spares for top-failing models [Immediate · Medium]

Procure and rack hot spares for ST12000NM0008, HGST HUH721212ALN604, and TOSHIBA MG08ACA16TA in each affected enclosure. Configure automatic rebuild triggers so MTTR drops below 60 minutes for the most common drive models.

### 2) Inspect shared backplane and power for correlated failures [Immediate · High]

Audit power distribution units, vibration dampening, and backplane firmware for enclosures hosting ST8000NM0055 alongside HGST HUH721212ALN604 and TOSHIBA MG08ACA16TA. Run vendor diagnostics on shared SAS/SATA expanders to rule out controller-induced multi-drive faults.

### 3) Implement proactive SMART-based drive retirement [Short-term · Medium]

Enable weekly SMART attribute monitoring (reallocated sectors, pending sectors, current pending count) on all drives and automatically schedule replacement when thresholds exceed vendor recommendations, targeting the 02:00 batch-failure window for pre-emptive swaps during planned maintenance.

## Top Problematic Time Windows

| Time Window | Incidents | % of Total | Avg. Duration |
|---|---|---|---|
| Thu 02:00 | 193 | 18.3% | 1338 min |
| Mon 02:00 | 161 | 15.2% | 1295 min |
| Fri 02:00 | 160 | 15.1% | 1263 min |
| Wed 02:00 | 160 | 15.1% | 1243 min |
| Tue 02:00 | 137 | 13.0% | 1226 min |

| | |
|---|---|
| Data Quality Grade: B | Analysis is reliable with minor caveats — see observations above. |

### Data Quality Note

10 duplicate rows removed (1% of 1057). 86 time gaps exceeding 5 minutes detected. Periodicity module (Lomb-Scargle) returned no_significant_peaks: no peaks with FAP <= 0.1 (best FAP: 1). KMeans clustering silhouette=0.484 indicates moderate but not strong cluster separation (k=2). Change-related description matching found no matches across 1057 incidents scanned.

| | |
|---|---|
| Dual-Model Verification | Both analysis engines agree on the findings. |

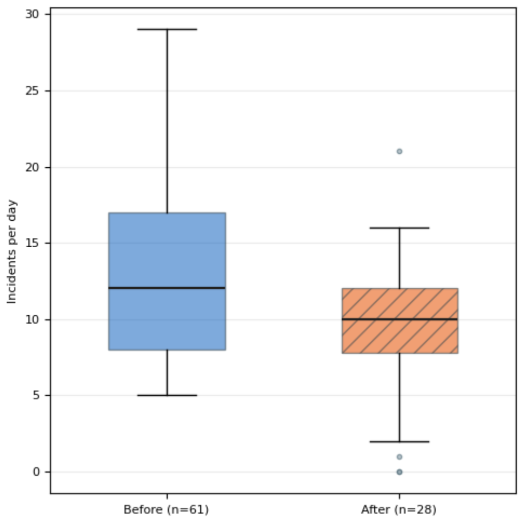
# METHODOLOGY

## Statistical Methods Applied

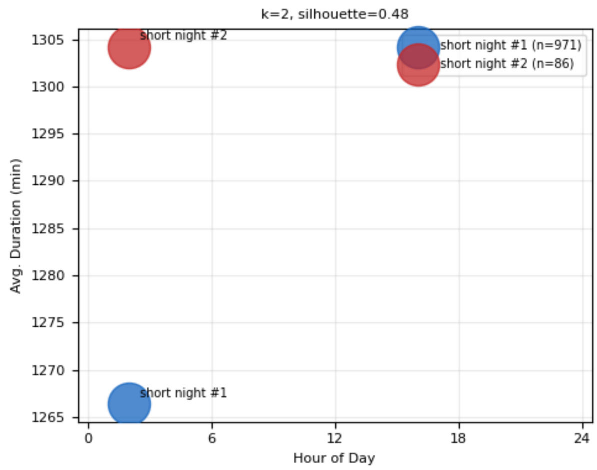
| Method | Purpose | α / Threshold |
|---|---|---|
| Lomb-Scargle | Detect recurring failure cycles in time-series data | FAP ≤ 0.10 |
| Mann-Whitney U (two-sided) | Compare incident rates / metrics before and after a change date | α = 0.05 |
| Cliff's δ + Bootstrap CI | Quantify the magnitude and uncertainty of the before/after change | 95% CI, 1 000 resamples |
| K-Means + Silhouette | Group incidents by behaviour (duration, time of day, weekday) | k □ [2..6], S ≥ 0.25 |
| Gini coefficient + Lorenz | Measure concentration of downtime among incidents | — |

Significance level α = 0.05 for all hypothesis tests. Tests run only when minimum data requirements are met (see Limitations section).

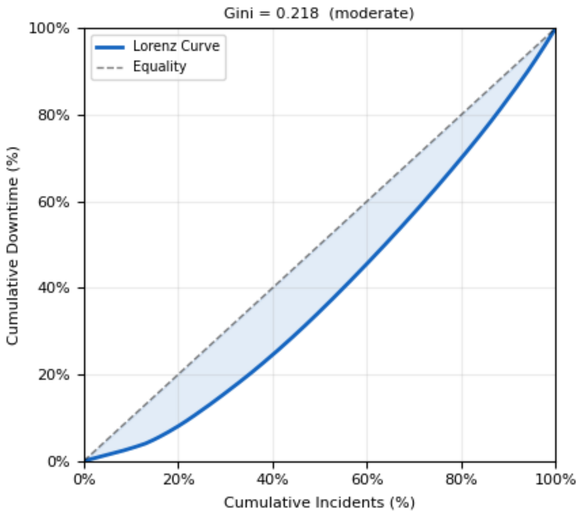
### Before vs After — Change Impact



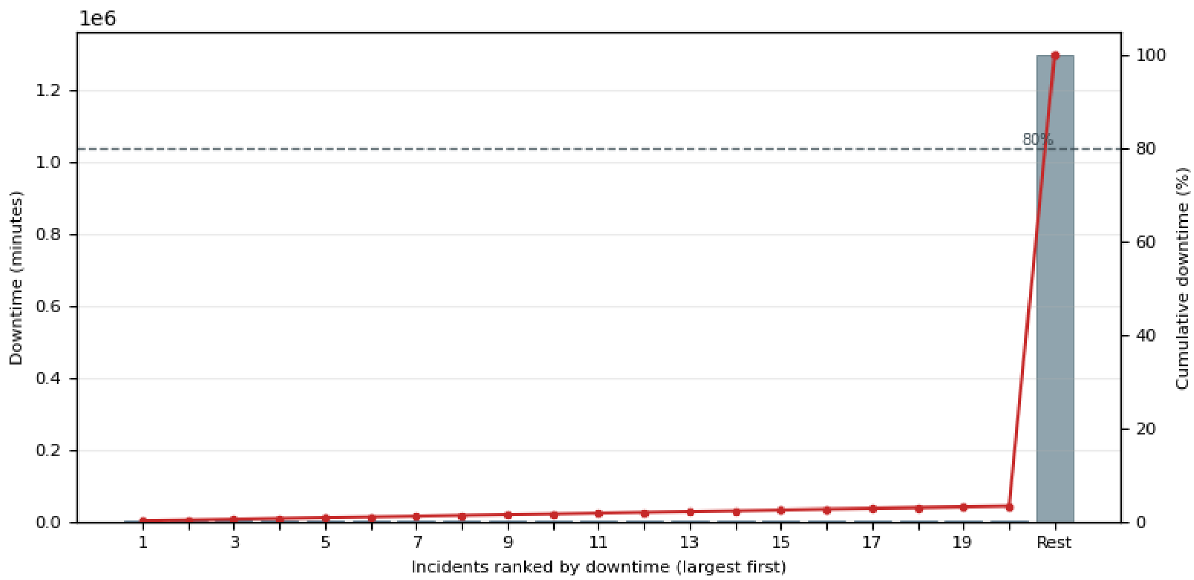

### Incident Clusters




### Lorenz Curve — Downtime Concentration



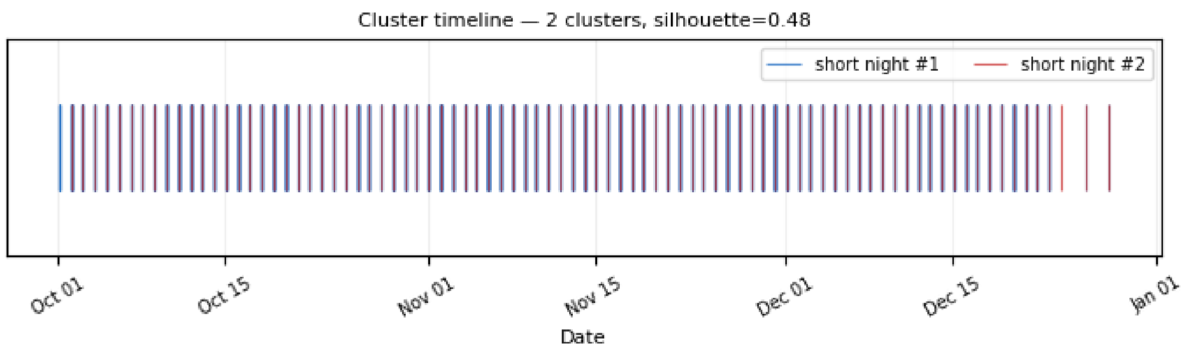

## Pareto Chart — Downtime by Incident




## Cluster Profile

| Cluster | Incidents | Median (min) | Peak Hour | Downtime share |
|---|---|---|---|---|
| short night #1 | ██████████ 971 (92%) | 1348 | 02:00 | ██████████ 92% |
| short night #2 | ██████████ 86 (8%) | 1314 | 02:00 | ██████████ 8% |

## Cluster Timeline



## Brief Glossary of Methods Used

| | |
|---|---|
| **p-value** | Probability of observing results at least as extreme as measured, assuming no real effect. $p < 0.05$ is statistically significant (< 5 % chance it is random). |
| **95% CI** | 95 % Confidence Interval — the range that contains the true value in 95 out of 100 repeated experiments. Wider CI = more uncertainty. |
| **Cluster** | A group of incidents with similar behaviour (duration, time of day, day of week). Incidents in the same cluster likely share a root cause. |
| **Gini coefficient** | Measures how unevenly downtime is distributed. 0 = all incidents equal; 1 = one incident caused all downtime. Gini > 0.6 = strong concentration. |
| **FAP** | False Alarm Probability — how likely a detected cycle is random noise. FAP < 0.01 = strong evidence of a real recurring pattern. |