# **하드웨어 장애 유형별 장애분석 가이드라인**

## **Case 1: ADU Report 분석 가이드라인**

ADU(Array Diagnostic Utility) 보고서는 **서버 스토리지 상태를 진단하는 중요한 도구**입니다. 다음 절차에 따라 분석하십시오.

1. **초기 증상 확인**:
   * 시스템 로그에서 "ext\_ext\_check\_block: bad header"와 같은 파일시스템 오류 메시지 확인
   * 사용자 보고 증상과 로그 메시지 간 연관성 분석
2. **하드웨어 상태 점검**:
   * ADU 보고서에서 Read/Write 관련 에러 항목 검사
   * Bad Target Count 확인 (0이면 물리적 디스크 손상 가능성 낮음)
   * SMART 데이터 분석으로 디스크 물리적 상태 확인
3. **파일시스템 오류 분석**:
   * ext4\_mb\_generate\_buddy, mb\_free\_blocks, ext4\_ext\_check\_block 관련 오류는 파일시스템 메타데이터 손상 가능성 시사
   * 해당 오류에 대한 벤더 지식베이스(RedHat 등) 참조
4. **스토리지 컨트롤러/RAID 점검**:
   * RAID 컨트롤러 펌웨어 버전 확인 및 최신 버전과 비교
   * 컨트롤러 로그에서 이벤트 패턴 분석
5. **권장 조치 수립**:
   * 하드웨어 교체보다 RAID 컨트롤러 펌웨어 업데이트 우선 시도
   * 벤더 제공 진단 도구(HPE Diagnostic Tool 등)로 심층 테스트 실행
   * 파일시스템 복구 작업 전 데이터 백업 확보

파일시스템 오류가 하드웨어 문제로 오인되는 경우가 많으므로, OS 로그와 하드웨어 진단을 상호 검증하는 과정이 필수적입니다.

## **Case 2: Network Ping Loss 분석 가이드라인**

네트워크 연결 문제는 다양한 계층에서 발생할 수 있으므로 체계적인 접근이 필요합니다:

1. **초기 현상 검증**:
   * ping 실패 패턴 분석 (간헐적/지속적, 특정 목적지/전체)
   * 패킷 손실률, 응답 시간 변동성 측정
2. **하드웨어 레벨 확인**:
   * 서버 이벤트 로그에서 NIC 관련 오류 메시지 검색
   * NIC 상태, 링크 상태, 듀플렉스 설정 확인
   * 케이블 및 물리적 연결 상태 점검
3. **OS 레벨 확인**:
   * 네트워크 인터페이스 통계 분석 (errors, drops, collisions)
   * 이벤트 로그에서 네트워크 스택 관련 오류 검색
   * TCP/IP 설정 및 라우팅 테이블 확인
4. **네트워크 인프라 확인**:
   * 스위치 포트 상태 및 오류 카운터 검사
   * 중간 네트워크 장비 로그 분석
   * 네트워크 경로 추적 및 병목 지점 식별
5. **권장 조치 수립**:
   * OS 로그 심층 분석으로 소프트웨어적 원인 식별
   * 중간 네트워크 장비 및 케이블 점검 우선 권고
   * 문제 지속 시 NIC 교체 고려

물리적 하드웨어 문제가 없는 경우가 많으므로, 네트워크 인프라와 OS 구성의 종합적 검토가 중요합니다.

## **Case 3: GPU 서버 장애 분석 가이드라인**

GPU 서버 장애는 고성능 컴퓨팅 환경에서 특히 중요하며, 다음과 같이 분석합니다:

1. **초기 오류 메시지 분류**:
   * "uncorrectable memory errors"는 GPU 메모리 오류 시사
   * "A corrupt InfoROM"은 GPU 메타데이터 저장소 손상 의미
2. **GPU 메모리 오류 분석**:
   * 메모리 오류 발생 횟수 및 패턴 분석
   * Remapping 발생 횟수 확인 (본 사례에서는 7번의 row remapping)
   * NVIDIA 공식 RMA 기준 적용 (동일 bank에 9번 이상 오류, 동일 row 중복 remapping, 512회 이상 총 remapping)
3. **InfoROM 손상 분석**:
   * GPU 드라이버 버전 확인 (특정 버전에서 알려진 이슈 확인)
   * NVIDIA-suspend 기능 사용 여부 확인
   * InfoROM 데이터 영역 손상 범위 평가
4. **권장 조치 수립**:
   * 메모리 오류의 경우: 초기에는 GPU reseat 시도, 기준 초과 시 교체
   * InfoROM 손상의 경우: 심각도에 따라 InfoROM 플래싱 또는 GPU 교체
   * 명확한 RMA 기준 제시로 불필요한 교체 방지

NVIDIA 공식 가이드라인을 참조하여 명확한 교체 기준을 적용하되, 초기 단계에서는 비침습적 해결책(reseat, 펌웨어 업데이트)을 우선 시도합니다.

## **Case 4: Infiniband 성능 저하 분석 가이드라인**

Infiniband 네트워크는 고성능 컴퓨팅에 중요하며, 성능 저하 원인을 다음과 같이 분석합니다:

1. **초기 성능 검증**:
   * ib\_send\_bw 테스트 결과 비교 분석
   * 저하된 대역폭 수치 및 패턴 확인
2. **물리적 연결 확인**:
   * Infiniband 어댑터 상태 및 포트 활성화 상태 점검
   * 케이블 재장착 및 물리적 손상 여부 확인
   * ibsw(Infiniband 스위치) 연결 상태 검증
3. **링크 구성 분석**:
   * 링크 속도 및 설정 검사 (4X HDR, Active/LinkUp 상태 확인)
   * 오류 카운터 점검으로 물리적 전송 오류 식별
4. **시스템 구성 요소 비교**:
   * 정상 시스템과 문제 시스템의 차이점 식별
   * PCIe relax order 설정 및 ibv\_wr\*API 활성화 여부 비교
   * OS 레벨의 Infiniband 설정 검토
5. **권장 조치 수립**:
   * PCIe relax order 및 ibv\_wr\*API 설정 조정
   * OS 레벨 Infiniband 구성 최적화
   * 하드웨어보다 소프트웨어/설정 이슈에 집중

Infiniband 성능 문제는 하드웨어 장애보다 구성 최적화 이슈인 경우가 많으므로, 시스템 설정 비교 분석이 핵심입니다.

## **공통 분석 프레임워크**

모든 하드웨어 장애 유형에 적용할 수 있는 공통 접근법:

1. **증상 검증**: 보고된 증상을 객관적으로 재현하고 측정합니다.
2. **로그 수집과 통합**: 하드웨어 로그, OS 로그, 애플리케이션 로그를 동일한 시간선에서 분석합니다.
3. **계층적 접근**: 물리 계층부터 애플리케이션 계층까지 단계적으로 검증합니다.
4. **벤더 가이드라인 참조**: 하드웨어 제조사의 공식 지침을 적용합니다.
5. **최소 침습 원칙**: 데이터 손실 위험을 최소화하는 조치부터 단계적으로 적용합니다.
6. **근본 원인 확인**: 증상이 아닌 근본 원인을 파악하여 재발을 방지합니다.