정보관리기술사 데일리 다이제스트

2026-10-23 · 국내/국제/공공 12대 도메인 매핑

1. 오늘의 브리핑

축	도메인	헤드라인	키워드
국내	DS AI SW 디지털서비스	**웹페이지 기능을 AI 에이전트가 직접 실행하는 WebMCP 적용** 에이전트 중심 웹 연동 방식 전환 · etnews.com/20261023000002	WebMCP · AI 에이전트 · 에이전틱 커머스 · 도구 호출
국제	CA DB AI 컴퓨터구조	**AI 추론이 남기는 KV 캐시 저장소 비용 부상** GPU 중심 비용 인식의 저장소 확장 · technologyreview.com/2026/10/22/1146948/ais-hidden-storage-bill-what-inference-leaves-behind/	KV 캐시 · 추론 스토리지 · 메모리 계층 · TCO
공공·정책	LAW SE 법·규제	**PHP 취약점 대응 개인정보 유출 예방조치 당부** **근거** 「개인정보 보호법」 안전조치의무 (제29조) · 취약점 악용 대비 법적 안전조치 환기 · pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=D010030000&nttId=12588	PHP RCE · 안전조치의무 · 긴급 패치 · 유출 예방

2. 오늘의 토픽 — LLM 추론 KV 캐시 계층화 스토리지

CA AI DB SE | 출제 예상: **2교시**

2-1. 개념 및 정의

LLM 추론 중 생성된 어텐션 키·값 텐서를 저장해 재사용하는 캐시 구조. 중복 연산 제거를 통한 지연시간 단축과 GPU 비용 절감.

2-2. 등장 배경

에이전트 장문 컨텍스트 확산에 따른 KV 캐시의 HBM 용량 초과. GPU 중심 비용 산정에서 메모리·저장소 포함 TCO 관점으로의 이동.

2-3. 핵심 기술요소

분류	주요 기술요소	역할 및 메커니즘
캐시 재사용	Prefix Caching, PagedAttention	공통 프롬프트 접두부 KV 블록 공유와 단편화 억제
계층 저장	HBM-DRAM-NVMe 오프로딩, CXL 메모리	접근 빈도별 KV 블록 계층 배치
운영 제어	캐시 인지 라우팅, 퇴출 정책, TTL	적중률 극대화와 보존 비용 균형

2-4. 시험·실무 관점

- [CA 연계] HBM 용량 한계와 CXL 메모리 풀링 결합
- [DB 연계] 캐시 일관성·퇴출·TTL 정책 설계
- [SE 연계] 테넌트 간 캐시 공유의 사이드채널 유출 위험

답안 필수 키워드 — KV 캐시 / PagedAttention / Prefix Caching / 메모리 계층화 / CXL