

AI 인프라 구축

GPU 집적화에 따른 AI 데이터센터 MEP 설계 구조 변화

글 | SK AX, AI&DC Solution CoE 손명선 CoE장

01·

초고밀도 가속기 집적화와
생명선(MEP)의 변화

02·

물리적·구조적 설계의 변화:
중형차 무게의 컴퓨터 캐비닛 수용하기

03·

전력(Electrical) 설계의 대변화:
고속도로망 전환과 800V DC 배전의 마법

04·

기계 및 배관(Mechanical & Piping) 설계:
에어컨 바람 대신 '온수'로 칩 식히기

05·

안전성 및 제어 시스템:
누액 감지, 격리 및 스마트 전원 차단 시스템

06·

국내외 프로젝트 실무 및 사업 추진 방식의 변화:
레고 블록 공법과 다자 설계 통합

07·

폐열 회수(Waste Heat Recovery)
및 친환경 데이터센터 구축

08·

결론

AI 팩토리 전환

- 생성형 AI와 대규모 언어모델 확산으로 데이터센터는 단순 서버 수용 시설에서 초고밀도 연산을 수행하는 **AI Factory**로 전환되고 있다. 특히 NVIDIA Blackwell·Rubin 계열과 같은 차세대 GPU 아키텍처는 랙 단위 전력·발열·중량을 기존 데이터센터 설계 한계 밖으로 끌어올리고 있다.
- 이에 따라 데이터센터 경쟁력은 더 이상 IT 장비 성능만으로 결정되지 않는다. **전력 공급, 냉각, 배관, 구조, 안전 제어가 통합된 MEP 역량**이 AI 인프라의 성능·안정성·운영비를 좌우하는 핵심 요소로 부상하고 있다.

초고밀도 랙 구조

- 기존 엔터프라이즈 데이터센터의 랙당 전력 밀도는 5~10kW 수준이었으나, 최신 AI 랙은 120~132kW, 피크 기준 192kW까지 요구한다. 랙 중량도 약 1.36톤에 달해 기존 이중 마루 구조로는 하중을 감당하기 어렵다.
- 따라서 AI 데이터센터는 표준 19인치 랙과 Raised Floor 중심 설계에서 벗어나, **OCP ORv3 기반 대형 랙, Slab-on-Grade 콘크리트 바닥, 고하중 구조 보강**을 전제로 설계되어야 한다. 이는 건축·구조 단계부터 IT 장비 사양을 반영해야 함을 의미한다.

중전압·800V HVDC

- 랙당 100kW 이상 전력을 공급하는 환경에서는 기존 저전압 배전 방식이 전류 증가, 케이블 비대화, 전력 손실, 공간 잠식 문제를 유발한다. 이를 해결하기 위해 AI 데이터센터는 6~35kV 중전압 배전과 34.5kV급 MV UPS를 서버 구역 인근까지 도입하는 방향으로 진화하고 있다.
- 또한 랙 내부에서는 기존 54V DC 체계가 한계에 도달하면서 **800V 고전압 직류 배전**이 핵심 대안으로 부상하고 있다. 800V HVDC는 전류량을 대폭 낮추고 전력 변환 단계를 줄여 구리 사용량, 열 손실, 전력실 면적을 동시에 절감할 수 있다.

직접 칩 액체냉각

- 초고밀도 GPU 랙의 발열은 기존 공랭 방식으로 처리하기 어렵다. 공기보다 열전달 능력이 월등히 높은 액체를 칩 표면에 직접 접촉 시키는 **DLC**(Direct Liquid Cooling)이 차세대 AI 데이터센터의 표준 냉각 방식으로 자리잡고 있다.
- DLC는 CDU를 중심으로 시설수 루프와 기술 냉각수 루프를 분리하는 이중 펌프 구조를 채택한다. 특히 45°C 수준의 온수 냉각을 허용하면 칠러(Chiller, 냉동기) 의존도를 낮추고, 건식 냉각기를 활용해 물 소비량을 0에 가깝게 줄이는 친환경 운영이 가능하다.

안전·모듈러·폐열 회수

- 액체냉각 확대는 누수, 합선, 장비 손상 리스크를 동반한다. 이에 따라 로프 센서, Redfish API, 자동 밸브 차단, 랙 전원 격리 등 **센서 기반 자율 제어 체계**가 필수 설계 요소가 되고 있다.
- 동시에 AI 인프라 구축 속도를 높이기 위해 전력 장비, CDU, 펌프, 배관을 공장에서 사전 조립하는 프리팹 모듈러 방식이 확산되고 있다. 나아가 55~70°C 수준의 회수 열을 지역난방이나 탄소포집 열원으로 활용하는 폐열 회수 모델은 AI 데이터센터의 지속가능성을 높이는 핵심 전략이다.

- ▶ 생성형 AI와 대규모 언어모델의 확산은 데이터센터의 설계 기준을 근본적으로 바꾸고 있다. 과거 데이터센터는 비교적 낮은 랙당 전력 밀도와 공랭 기반 냉각 체계로 운영 가능했으나, 최신 GPU 집적형 AI 인프라는 랙 하나가 기존 수십 가구에 해당하는 전력을 소비하고 중형 SUV 수준의 중량을 갖는 초고밀도 시스템으로 진화하고 있다.
- ▶ 이 변화는 MEP 설계를 데이터센터의 보조 인프라가 아닌 **AI 성능을 가능하게 하는 핵심 플랫폼**으로 격상시키고 있다. 전력 측면에서는 중전압 배전, MV UPS, 800V HVDC가 기존 저전압·다단계 변환 구조를 대체하고 있으며, 냉각 측면에서는 공랭에서 직접 칩 액체냉각으로의 전환이 빠르게 진행되고 있다.
- ▶ 특히 45°C 온수 기반 액체냉각은 칠러 의존도를 낮추고 건식 냉각기를 활용함으로써 에너지 효율과 물 사용량 절감을 동시에 달성할 수 있다. 이는 AI 데이터센터가 막대한 전력을 소비하면서도 지속가능성을 확보해야 하는 시장 요구와 맞닿아 있다.
- ▶ 향후 AI 데이터센터의 경쟁력은 GPU 확보 능력뿐 아니라, 고전력·고발열·고하중 환경을 안정적으로 수용하는 통합 MEP 설계 역량, 프리팹(Prefab) 모듈러 기반 빠른 구축 역량, 폐열 회수까지 포함한 친환경 운영 모델에 의해 좌우될 것이다.

01 초고밀도 가속기 집적화와 생명선(MEP)의 변화

정보기술(IT) 산업 전반에 걸쳐 대규모 언어 모델(LLM)의 사전학습(Pre-training)과 실시간 생성형 AI 서비스가 기하급수적으로 확장됨에 따라, 전통적인 데이터센터 설계 문법은 근본적인 한계에 직면해 있다.

과거의 데이터센터는 일반적인 인터넷 홈페이지를 보여주거나 파일을 저장하는 가벼운 작업을 처리했기 때문에, 개별 컴퓨터 캐비닛(랙)이 소모하는 전력이 낮아 일반 가정용 에어컨과 비슷한 공기 냉각(Air Cooling) 방식으로도 충분히 열을 식힐 수 있었다. 이는 비유하자면 일반 사무실에 개인용 컴퓨터 여러 대를 켜두고 선풍기를 돌리는 수준이었다.

하지만 단일 실리콘 칩의 성능 한계를 넘어 랙 단위에서 수천 개의 GPU(그래픽 처리 장치)를 일체화된 하나의 초거대 컴퓨터처럼 작동시키는 엔비디아(NVIDIA)의 Blackwell 및 Rubin 아키텍처의 도입은 데이터센터를 하나의 거대한 'AI 팩토리(AI Factory)'로 탈바꿈시키고 있다. 이제 데이터센터는 사무실이 아니라 거대한 용광로를 가동하는 중화학 공장에 가까워졌다.

이에 따라 데이터센터의 핵심 생명선이라고 할 수 있는 MEP, 즉 기계(Mechanical - 냉각), 전기(Electrical - 전력), 배관(Plumbing - 물길) 설계는 완전히 새로운 기술적 임계점을 맞이하고 있다. 델로로(Dell'Oro)의 분석에 따르면, 컴퓨터 칩을 액체로 직접 식히는 직접 칩 액체냉각(DLC, Direct Liquid Cooling) 시장 규모는 2025년 30억 달러에서 2029년 70억 달러 수준으로 폭발적으로 성장할 전망이다. 이는 차세대 데이터센터의 성패가 이 뜨겁고 거대한 연산 엔진에 전기와 물을 얼마나 효율적으로 공급하고 식히느냐에 달려 있음을 보여준다.

02 물리적·구조적 설계의 변화: 중형차 무게의 컴퓨터 캐비닛 수용하기

전통적인 데이터센터의 컴퓨터 캐비닛인 '랙(Rack)' 한 대가 쓰는 전력은 5kW에서 10kW 수준에 머물렀다. 이는 가정에서 쓰는 드라이기 서너 개를 동시에 켜두는 정도의 전기 용량이다. 반면, 최신 AI 데이터센터에서는 랙 한 대당 전력 밀도가 최소 40kW에서 최대 130kW 이상으로 폭증하고 있다.

특히 엔비디아의 차세대 가속기 랙(GB200 NVL72)은 랙 단 한 대에서만 무려 120kW에서 132kW의 전기를 지속적으로 소모하며, 순간적으로 전기 힘을 모아 작동할 때는 최대 192kW의 전력을 요구한다. 이는 일반 가정집 약 50~60가구가 동시에 쓰는 엄청난 전력량을 캐비닛 단 한 대가 소모하는 격이다.

설계 차원	전통적인 엔터프라이즈 데이터센터	초고밀도 AI 데이터센터 (GB200 NVL72 준거)
랙당 전력 밀도	5kW ~ 10kW	120kW ~ 132kW (피크 192kW)
물리적 랙 규격	표준 19인치 규격 (EIA-310-D)	비표준 특수 대형 규격 (OCP ORv3)
단일 랙당 무게	약 300kg ~ 500kg	약 1.36메트릭톤 (중형 SUV 1대 무게)
바닥 하중 설계	이중 마루 (Raised Floor, 150~250 lbs/sq ft)	콘크리트 슬래브(Slab-on-Grade) 또는 고하중 강화 공법
기본 냉각 방식	에어컨 바람을 이용한 공기 냉각 (CRAC/CRAH)	직접 칩 액체냉각(DLC) 및 매니폴드 루프 통

이러한 전력 밀도의 폭증은 랙의 무게와 크기도 극적으로 늘려놓았다. GB200 NVL72 랙은 수십 개의 GPU와 CPU, 무거운 직류 전원 공급 장치, 그리고 액체 냉각용 수직 물길 파이프(매니폴드)를 단일 프레임에 뿔뿔하게 밀집시키면서 총 무게가 약 1.36톤(3,245 lbs)에 육박하게 되었다.

과거 데이터센터에서는 바닥 밑으로 전선이나 바람이 지나갈 수 있도록 격자형 다리로 띄워 놓은 '이중 마루(Raised Floor)' 공법을 주로 사용했다. 하지만 이중 마루의 허용 하중은 대개 평방피트당 150~250파운드 수준에 불과하다. 여기에 중형 SUV 차량 한 대 무게에 달하는 1.36톤의 AI 랙을 올릴 경우, 얇은 임시 무대 위에 무거운 피아노를 올려둔 것처럼 바닥 뼈대가 휘어지거나 주저앉는 대참사가 발생할 수 있다.

따라서 현대의 AI 데이터센터는 아래가 붕 떠 있는 이중 마루 시스템을 과감히 제거하고, 건물 최하단 바닥과 같은 단단하고 두꺼운 콘크리트 통바닥 직접 시공(Slab-on-Grade) 방식을 채택하거나, 건물의 철골 기초 프레임을 극도로 보강하는 설계를 표준화하고 있다.

더 많은 내용을 보시려면

파일 다운받기

버튼을 눌러주세요