

NVIDIA DGX 고객 정보 세션 FAQ

DGX H100을 사용한 데이터 센터 모범 사례

Q. 한 SU (Scalable unit) 의 권장 작업 시간은 어떻게 되나요?

A. 데이터 센터가 이미 DGX SuperPOD에 적합하다고 가정할 경우 (즉, 추가 전력이나 냉각을 추가할 필요가 없을 때) 일반적으로 배치 전 작업 (현장조사, 장비조달 등) 에 소요되는 기간은 ~12주입니다. 이후 배치부터 실제 작업 소요 시간은 일반적으로 4-6주가 추가 소요됩니다

Q. Rack에 있는 스토리지가 기타 장치를 위해 추가 전력이 필요합니까?

A. 스토리지 및 기타 인프라 항목은 DGX SuperPOD의 개별 Rack 에 있습니다. DGX H100 Rack 과는 다른 전력 요구사항을 가지고 있습니다.

Q. DGX를 설치하려면 hot one (핫존) 에는 어느 정도의 공간이 필요합니까?

A. 필요한 최소 공간이 어떻게 되는지는 [여기](#)에 있는 도표를 확인해주시기 바랍니다. 또한, 서버리프트를 작동시키기에 충분한 공간이 필요하며 참고로 서버 리프트는 핫 존이 (hot zone) 아닌콜드 존 (cold zone) 에서만 작동합니다

Q. 상면 상온 온도 24C 인 40kW Rack 에 표시되는 풍량(CFM) 의 양은 얼마입니까?

A. 해당 정보는 여기 DGX H100 사이트 조사에서 확인하실 수 있습니다.

Q. Rack 에서 하단 2대의 DGX H100 시스템과 상단 2대 사이에 간격이 있는 이유가 있나요?

A. B해당 간격은 필요시 수평 Rack PDU 추가를 위한 것입니다. 또한 Rack 내부 공기흐름을 원활하게 하기 위한 용도이기도 합니다.

Q. DGX H100에서 인접한 공간으로 전달되는 소음과 진동을 줄이기 위한 모범 사례는 무엇인가요?

A. 시스템에서 발생하는 소음의 양은 팬이 얼마나 세게 작동해야 하는지에 따라 결정됩니다.공급 공기가 차가울수록 팬이 덜 작동하고 시스템이 더 조용해집니다. 사양에 명시된 101데시벨은 은 0°C의 공급 공기에 대한 것으로, 이는 상당히 높은 수준이며 이 수준에서는 팬이80로 작동합니다. 작업자는 서버 주변에서 작업할 때 귀마개와 귀마개 등 적절한 개인 보호장비를 착용해야 합니다. 인접한 방의 경우, 데이터 센터 룸 외부의 소음을 제한하는 소음 저감 전략을 수행하려는 경우, 건물 구조에 따라 다를 수 있습니다.

Q. Rack 에 있는 4개의 시스템을 지원하려면 3개의 PDU로 전원 연결이 가능합니까?

A. Rack PDU가 적절한 수의 출력을 가진 유형이라면 3개의 PDU로 충분합니다.

Q. Rack 당 평균 및 피크 전력 소비량은 얼마인가요?

A. 전체 4 서버 Rack 프로파일의 경우 40.8kW, 2 서버 Rack 프로파일의 경우 20.4kW, 단일 서버 Rack 프로파일의 경우 10.2kW입니다.

Q. 20.4kW 에서 2 DGX H100 설계를 생각하기 위한 일반적인 모범 사례는 무엇인가요?

A. 몇 가지 일반적인 냉각 최적화 기법에는 적절한 공기 차단 및 격리 Rack 사이에 공간을 배치하여 냉각 집중하는 등의 방법이 있습니다. 냉각 설계는 데이터센터 디자인에 따라 달라질 수 있습니다.

Q. H100과 H100 NVL의 전력에 차이가 있나요?

A. DGX SuperPOD는 H100 SXM 폼 팩터 GPU가 포함된 DGX H100 서버를 사용합니다. 또한 PCIe 폼 팩터인 H100 및 H100 NVL GPU도 있습니다. SXM 버전은 GPU 간의 통신에 최적화 되어 있으며, 다른 버전들보다 더 높은 전력/clock 으로 실행됩니다. 차이점은 [여기](#)서 확인 할 수 있습니다.

Q. DGX H100 사이트의 전력 및 냉각에 대한 표준 가이드가 필요합니다.

A. [여기](#)에 표시된 "데이터센터 설계 가이드"에 모범 사례를 참고 하십시오.

Q. NVIDIA는 특정 Rack 규격 지정합니까?

A. 예. NVIDIA는 사용할 Rack의 특정 브랜드나 모델을 지정하지는 않지만, 다음과 같은 사항을 명시 하고 있습니다.

Rack은 19인치 EIA 마운팅 이 있는 밀폐형 Rack의 경우 EIA-310 표준을 준수해야 합니다. Rack의 규격은 최소 24" x 44"(600mm x 1100mm) 크기, 최소 48U 높이여야 합니다. NVIDIA는 30" x 48" x 52U (700mm x 1200mm) Rack을 권장합니다.

Q. 패치 패널 또는 Rack 스위치 상단을 사용하여 InfiniBand 케이블 길이 제한을 연장할 수 있습니까?

A. 아니요. InfiniBand는 매우 고성능 아키텍처이며, 케이블 길이의 제한은 케이블의 일부 구간이 아닌 전체 신호 경로의 신호 감쇠 및 지연 시간을 기준으로 합니다. 패치패널은 신호감쇠를 악화시키고 Rack 스위치의 중간 상단에 있는 스위치는 Latency 를 증가시킵니다.

Q. NVIDIA는 특정 Rack PDU 유형을 지정하나요?

A. 예. Rack PDU는 각 데이터센터에서 사용할 수 있는 전원 프로비저닝 을 준수해야 하며, 각 시장 지역에서 사용 가능한 브랜드를 따라야 하므로 NVIDIA는 특정 브랜드나 모델을 지정하지 않습니다. 하지만, NVIDIA SuperPOD 설계가 통합된 스마트 모듈, 네트워크 인터페이스, RestAPI 인터페이스, 온도 및 센서 프로브용 포트, 잠금 리셉터클, PDU 레벨 측정, 리셉터클별 원격 콘센트 스위칭, 시각적 회로 구분을 위한 빨간색, 파란색, 회색 외관 (옵션) 을 포함하는 rPDU를 포함 할 것을 권장합니다.

Q. 확장형 유닛을 몇 개까지 배포할 수 있나요?

A. 일반적인 SuperPOD에는 최대 4개의 확장형 장치가 포함될 수 있지만, 추가적인 장치를 배치할 수도 있습니다.

Q. DGX Rack의 빈 공간을 다른 IT 장비에 활용할 수 있나요?

A. DGX 스케일러블 유닛 (Scalable Unit) 은 설계된 솔루션이므로 공유 Rack에서 관련 없는 장비와 함께 사용해서는 안 됩니다. 또한 냉각 용량을 통합하기 위해 Rack의 간격을 띄워 배치한 경우에도 DGX H100 Rack 사이의 공간에 관련 없는 장비를 배치를 지양합니다.

Q. DGX H100 시스템 아키텍처가 포함된 DGX SuperPgD에 대한 추가 정보는 어디에서 찾을 수 있나요?

A. 여기에서 다운로드할 수 있는 문서를 찾을 수 있습니다 NVIDIA DGX SuperPOD 데이터센터 디자인 레퍼런스 가이드 및 NVIDIA DGX SuperPOD AI 리더십을 위한 차세대 확장 가능한 인프라 레퍼런스 아키텍처.

