# SECTOR UPDATE

2026. 1. 7

Tech팀

**이종욱 팀장**
jwstar.lee@samsung.com

**김경빈 Research Associate**
kyoungbeen.kim@samsung.com

리서치센터 리포트
바로가기

# Tech (OVERWEIGHT)

## 젠슨황, 메모리 병목 현상을 언급

- 엔비디아, 베라 루빈 생산 시작과 함께 BlueField-4를 통해 메모리 스토리지 운영을 언급
- Agent AI와 Physical AI 추론 과정에서 메모리의 병목 현상이 급증되며 HBM 뿐 아니라 DRAM과 NAND의 동반 성장 시사

## WHAT'S THE STORY?

**엔비디아 CES 기조연설 - 26년에도 AI 투자 사이클 지속:** 엔비디아 CEO 젠슨황은 1월 6일 CES 2026 기조 연설에서 AI 인프라 사이클이 26년에도 강하게 진행될 것이라고 전망했다. 먼저, 현재 진행 중인 AI 인프라 투자를 신규 수요 창출 뿐만이 아닌 컴퓨터 인프라의 현대화(modernization) 단계로 규정했다. 거대한 규모의 투자 배경에는 기존 데이터센터가 LLM, Agent AI, Physical AI 워크로드에 구조적으로 부적합해 서버와 스토리지 전반의 재설계와 추가투자가 포함되어 있다. 다수의 오픈 모델의 등장, 자율주행용 Physical 추론 모델 알파마요(AlphaMayo)를 공개하며 생태계 확장이 계속됨을 보여주었다. 차세대 GPU 플랫폼인 베라루빈의 양산이 시작되었고 새로운 디자인의 서버와 메모리 아키텍처가 도입될 예정이다.

**트렌치 구조가 26년 AI 메모리의 핵심:** 이번 기조 연설이 메모리에 주는 가장 중요한 함의는 엔비디아의 AI 서버에서 GPU 중심의 트렌치 구조가 공식화되었다는 점이다. 베라루빈과 함께 BlueField-4 DPU를 통한 AI 메모리 스토리지 시스템이 제시되었는데, 이는 HBM 중심 구조에서 벗어나 DRAM과 NAND가 캐시 계층으로 본격 편입되는 전환점이다. Agent AI가 확산되면서 추론 단계에서 연산 호출 횟수와 시퀀스의 길이가 급격하게 증가하였다. 젠슨황은 KV Cache를 처리하기 위한 메모리의 대역폭과 용량이 기술적 병목현상이라고 언급했다. 이에 따라 엔비디아는 HBM - LPDDR(DRAM) - NVMe(NAND)의 연속된 메모리 계층을 통합하고, BlueField 반도체가 GPU에 직접 연결하고 컨트롤하는 방식을 통해 대용량의 KV Cache를 처리할 것이다. 이 과정에서 DPU를 프로그래밍하는 DOCA(Data-center On-a-Chip Architecture), 스토리지를 GPU와 연결하는 GDS(GPU Direct Storage), 그리고 BlueField 반도체(DPU+ethernet)로 이어지는 소프트웨어-하드웨어 통합 아키텍처가 마치 CUDA처럼 또 다른 구조적 해자로 작동할 가능성이 높다.

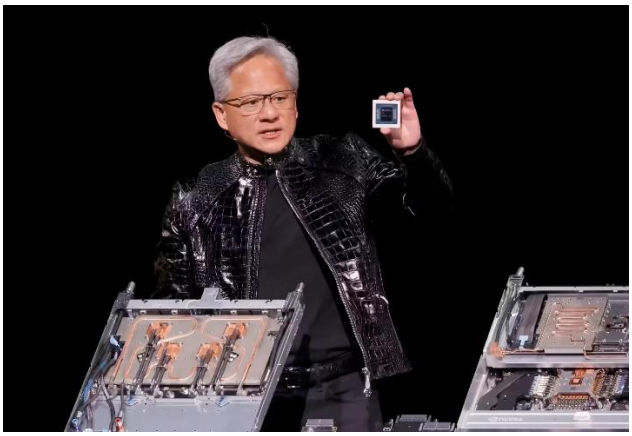
*(다음 페이지에 계속)*

**데이터센터의 중심은 당분간 메모리:** 당분간 AI 서버당 메모리 컨텐츠 증가 속도가 GPU 컴퓨팅 증가 속도를 상회할 가능성이 높다고 생각된다. 엔비디아가 전한 메시지는 베라루빈 시스템을 도입할 때 KV Cache 수용 능력을 최대한 확장하는 것이 우선 과제임을 시사한다. Agent AI와 Physical AI 워크로드는 추론 중심이 될 것이고, 학습 과정 대비 메모리 상주 데이터 비중이 높아서 GPU 1대당 연결되는 메모리와 스토리지의 탑재량 증가가 예상된다. 이는 HBM의 증가 뿐 아니라 일반 DRAM과 고용량 NVMe NAND의 동반 수요 확대로 이어질 것이다. 26년은 데이터센터향 메모리 bit 수요 성장률이 가파르게 올라가는 시작점이다.

엔비디아: BlueField-4

자료: Nvidia, 삼성증권

CES2026에서 기조 연설 중인 젠슨황



자료: 언론보도, 삼성증권

**〈엔비디아 투자자미팅 Q&A 요약〉**

Q) 베라 루빈(Vera Rubin)의 양산은 무엇을 의미하나?

베라 루빈은 현재 양산 단계이며 2H26 출시 계획. 약 9개월의 리드타임 존재.

Q) Groq(그록) 라이센스 발표의 전략적 의미는?

엔비디아 아키텍처가 주류이지만, 스마트글래스처럼 극도로 낮은 지연시간이 필요한 영역에서 Groq 팀의 기술이 우수함.

Q) 자율 주행, 로보틱스 시장의 범위와 타이밍, 수익모델?

자율주행 관련 데이터센터센터 시장은 이미 수십억 달러 규모이며 더욱 성장할 것.

자율주행 트럭, 로보택시 모두 10년 내 거대한 사업으로 성장 전망.

우리 칩을 쓰지 않더라도 생태계가 번성하는 것이 중요

Q) Anthropic(앤트로픽)이 엔비디아 반도체를 사용한 배경?

초창기 구글과 아마존이 Anthropic에 펀딩하며 구글과 아마존에 의존했으나 성능 중심으로 벤더 확장. 엔비디아 플랫폼이 고품질 토큰 생성에 유리.

엔비디아는 세계 거의 모든 모델이 사용하고 있음.

Q) 디램 가격 및 수급 상황, 전력 공급 문제

엔비디아는 전세계 칩 회사 중 유일하게 디램을 대량 구매. HBM 파트너와도 긴밀히 협력. 마찬가지로 AMAT, ASML 등의 반도체 장비 업체, 부지 확보, 전략망 구축 등 업스트림에 관여하고 있음. 공급망을 직접 관리하는 것이 안정적인 공급에 유리.

Q) 중국 시장 전략

중국향 제품은 별도의 공급망으로 대응하여 다른 지역 물량을 침해하지 않을 것.

현재 양국 정부 라이선스 절차를 기다리는 중.

Q) DGX Cloud가 고객사와 경쟁하는 서비스인가?

초기에 CSP들이 AI 아키텍처에 생소했고, 우리가 자체 시스템 용으로 구축한 뒤 2만개 이상의 AI 네이티브 기업을 CSP로 유치. 우리가 DGX Cloud를 통해 CSP의 영업사원 역할을 한 셈. 이제는 DGX Cloud라는 대여 사업의 중요도 하락.

Q) 메모리/스토리지 컨트롤러 사업이란?

AI 팩토리의 추론량이 너무 커서 KV cache(캐시)가 매우 무거움. 이를 처리하기 위해 컴퓨팅 패브릭 안에 바로 연결되는 새로운 계층의 스토리지가 필요. BlueField 4(블루필드)의 DOCA 소프트웨어가 그 역할을 수행. CUDA처럼 성장할 것.

Q) 베라 루빈의 양산 속도와 수익성

베라 루빈은 모든 구성 칩을 새롭게 개발. 서버 조립 시간을 2시간으로 단축. NVLink의 유지 보수 효율 향상. 베라 루빈의 도입 가치는 높고, 수익성도 유지될 것.

Q) AI 모델은 소수의 승자로 통합될 것인가, 아니면 파편화된 상태로 유지될 것인가?

범용 AI는 충분히 스마트하지만 특정 도메인에서는 Vertical agent가 필요. 많은 기업들이 오픈 소스 모델을 기반으로 자신만의 AI를 구축하게 될 것.

Q) Agent AI와 Physical AI의 변곡점은 언제?

이미 Agent AI와 Physical AI는 우리 곁에 존재. 사용은 더욱 쉬워질 것

Q) 네오 클라우드의 역할은?

기술 변화가 빠르고 전력 확보가 어려운 상황에서 지역별로 특화된 플레이어들이 나타난 것. 엔비디아의 중요한 파트너.

Q) 수익성의 지속가능성

수익성은 우리가 제공하는 가치와 직접적 관련이 있음. 모델 훈련 시간 단축, 추론 비용 절감, 데이터센터 단위의 수익성 증대라는 점에서 우리의 가치는 압도적. 500억 달러에 이르는 기가와트급 데이터센터를 건설 할 때 엔비디아의 제품이 경쟁사대비 더 많은 수익을 창출하게 도와준다면, 고객은 기꺼이 비용을 지불

Q) 다음 게임체인저는? 토큰 수요의 성장 지속성은?

컨텍스트 길이(context length)의 확대와 지속적 학습(continuous learning)이 다음 과제. Agent system 덕분에 토큰 생성량 50배 확대

Q) R&D 투자와 M&A 철학?

우리는 칩, 모델, 어플리케이션 등 전반에 투자. R&D와 외부 투자를 구분하지 않으며, 우리와 생각이 같고 기술적으로 빠른 팀이 있다면 인수 검토.

Q) SRAM이나 자체 기술을 통해 HBM 의존도를 낮출 수 있을까?

CPX 등 HBM 의존도를 낮춘 제품은 특정 작업에서 효율이 좋을 수는 있음. 그러나 AI 워크로드는 계속 변화하는 것이 문제. 데이터센터 전체의 전력은 한정되어 있으므로, 어떤 모델이든 최고의 성능을 낼 수 있는 통합 아키텍처가 결국 전체 소유 비용(TCO) 면에서 유리.

## Compliance notice

- 본 조사분석자료의 애널리스트는 2026년 1월 6일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 2026년 1월 6일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자가 등 제3자에게 사전 제공된 사실이 없습니다.

# 삼성증권

**삼성증권주식회사**
서울특별시 서초구 서초대로74길 11(삼성전자빌딩)
Tel: 02 2020 8000 / www.samsungpop.com

**삼성증권 Family Center:** 1588 2323

**고객 불편사항 접수:** 080 911 0900

NCSI KS-SQI KCSI KCPI 소비자중심 Best Korea Brands 2024
Member of Dow Jones Sustainability Indices Powered by the S&P Global CSA