中泰证券 ZHONGTAI SECURITIES

证券研究报告
2024年8月29日

【中泰电子】AI全视角 - 科技大厂财报专题英伟达Q2解读：Q2业绩新高，数据中心拉动增长

分析师：

王芳 S0740521120002

杨旭 S0740521120001

李雪峰 S0740522080004

中泰证券研究所
专业 | 领先 | 深度 | 诚信

FY25Q2业绩概览

单位：亿美元	FY25Q2 实际	yoy	qoq	市场预期	是否超预期
营收	300	+122%	+15%	286	超预期
——数据中心	263	154%	+16%	249	超预期
——游戏	29	+16%	+9%	28	超预期
——专业可视化	5	+20%	+6%	5	-
——汽车	3	+37%	+5%	3	
——其他	1	+33%	+13%	-	-
净利润（GAAP）	166	+168%	+12%	147	超预期
毛利率（GAAP）	75.1%	+5.0%	-3.3%	75.5%	–
Diluted EPS（GAAP，美元）	0.67	+168%	+12%	-	–

单位：亿美元	FY25Q3指引 下限	上限	中值	yoy	qoq
营收	319	332	325	+79%	+8%
毛利率（GAAP）	75.4%	74.9%	74.4%	+0.4%	-0.7%
毛利率（Non-GAAP）	76.0%	75.5%	75.0%	+0.0%	-0.7%
运营费用（GAAP）	-	-	43	+44%	+9%
运营费用（Non-GAAP）	-	-	30	+48%	+7%

来源：英伟达财报，彭博，中泰证券研究所

FY25Q2业绩要点

- 【**FY25Q2业绩**】FY25Q2公司收入300.4亿美元，yoy+122%，qoq+15%，此前市场预期286亿美元，超预期4.9%。本季度四大业务均同环比增长，其中数据中心收入同比增速最高。毛利率方面， FY25Q2毛利率为75.1%，yoy+5.0pct，qoq-3.3pct，市场预期75.5%，基本符合预期。

- 【**数据中心**】FY25Q2数据中心收入262.7亿美元，创历史新高，yoy+154%，qoq+16%，实现连续6个季度环比大幅增长，此前市场预期248.6亿美元，超预期5.7%。数据中心高增，主要得益于Hopper GPU计算平台的强劲需求。

- 【**H&B系列GPU**】**Hopper：** FY25Q2 H200平台开始增长，主要向大型云厂商、互联网公司发货。目前Hopper需求强劲，公司预计2025财年下半年出货量增加；**Blackwell：** FY25Q2公司向客户交付样品，为提高生产良率，对Blackwell GPU掩膜进行改动。Blackwell的生产爬坡计划于第四季度开始，一直持续到FY2026，并预计FY25Q4 Blackwell开始贡献数十亿美元的收入。

- 【**回购**】公司宣布董事会批准了额外500亿美元的股票回购计划，无到期日期。2025财年上半年，英伟达通过回购股票和现金股息的方式已向股东返还了154亿美元。截至第二财季末，英伟达的股票回购计划中还剩75亿美元。

- 【**FY25Q3指引**】预计营收325亿美元，上下浮动2%；毛利率（GAAP）为74.4%，上下浮动0.5%。

- 【**风险提示**】行业需求不及预期的风险、大陆厂商技术进步不及预期、中美贸易摩擦加剧、研报使用的信息更新不及时的风险。

来源：彭博，英伟达财报，华尔街见闻，凤凰网，中泰证券研究所

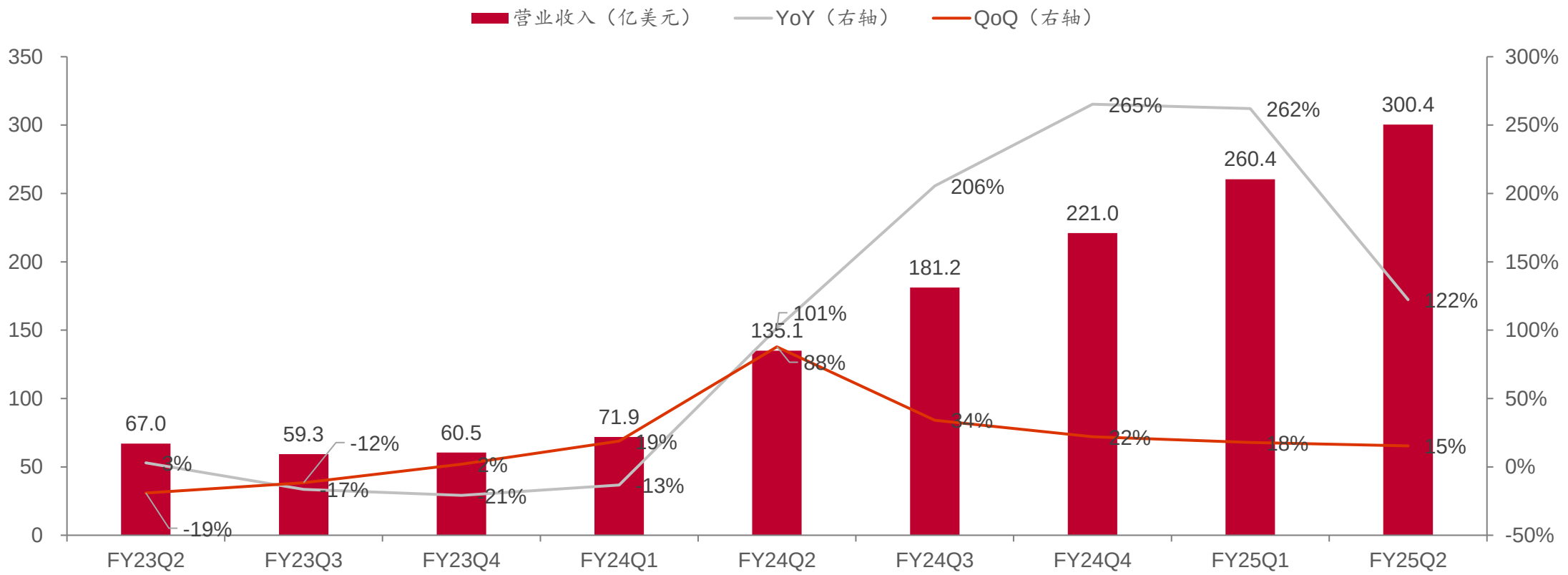
目 录

1、FY25Q2 财务情况：业绩超预期，数据中心拉动增长

■ **数据中心高增推动FY25Q2收入持续增长。**

➢ 收入：FY25Q2单季度收入300.4亿美元，yoy+122%，qoq+15%，连续7个季度环比增长，此前市场预期286亿美元，超预期4.9%。主要得益于数据中心收入高增推动。

图表：季度收入及增速情况



备注：FY25Q2为2024年5月至7月

来源：彭博，英伟达财报，中泰证券研究所

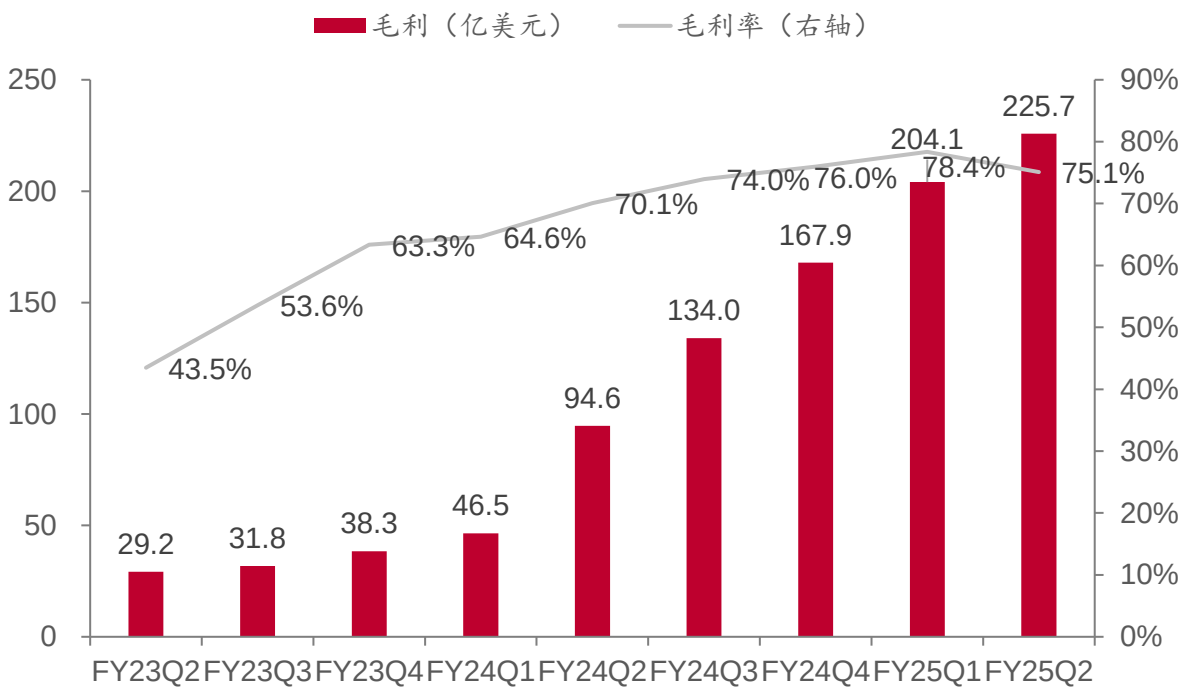
1、FY25Q2 财务情况：业绩超预期，数据中心拉动增长

■ **毛利率同比增长符合预期，研发费用持续增长。**

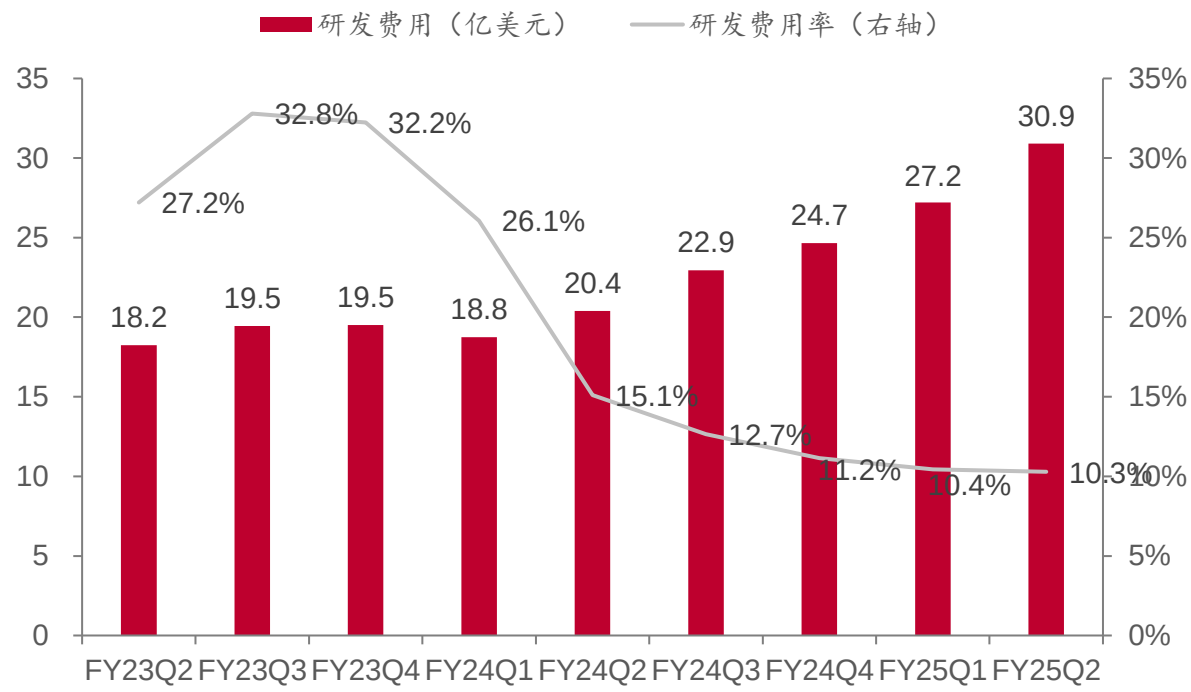
➢ 毛利率：FY25Q2毛利率为75.1%，yoy+5.0pct，qoq-3.3pct，市场预期75.5%，基本符合预期。

➢ 研发费用：本季度研发费用30.9亿美元，yoy+51.5%，研发费用率10.3%。

图表：毛利及毛利率



图表：研发费用及研发费用率



来源：彭博，英伟达财报，中泰证券研究所

1、FY25Q2 财务情况：业绩超预期，数据中心拉动增长

➢ **分业务看：FY25Q2各业务均有所增长，其中数据中心增速最高，占比达87%。**FY25Q2数据中心收入262.7亿美元，yoy+154%，qoq+16%，占比87%；游戏收入28.8亿美元，yoy+16%，qoq+9%，占比10%；可视化收入4.5亿美元，yoy+20%，qoq+6%，占比2%；汽车收入3.5亿美元，yoy+37%，qoq+5%，占比1%。

图表：分业务业绩情况

		FY23Q2	FY23Q3	FY23Q4	FY24Q1	FY24Q2	FY24Q3	FY24Q4	FY25Q1	FY25Q2
数据中心（亿美元）		38.1	38.3	36.2	42.8	103.2	145.1	184.0	225.6	262.7
	YoY	61%	31%	11%	14%	171%	279%	409%	427%	154%
	QoQ	1%	1%	-6%	18%	141%	41%	27%	23%	16%
游戏（亿美元）		20.4	15.7	18.3	22.4	24.9	28.6	28.7	26.5	28.8
	YoY	-33%	-51%	-46%	-38%	22%	81%	56%	18%	16%
	QoQ	-44%	-23%	16%	22%	11%	15%	0%	-8%	9%
专业可视化（亿美元）		5.0	2.0	2.3	3.0	3.8	4.2	4.6	4.3	4.5
	YoY	-4%	-65%	-65%	-53%	-24%	108%	105%	45%	20%
	QoQ	-20%	-60%	13%	31%	28%	10%	11%	-8%	6%
汽车（亿美元）		2.2	2.5	2.9	3.0	2.5	2.6	2.8	3.3	3.5
	YoY	45%	86%	135%	114%	15%	4%	-4%	11%	37%
	QoQ	59%	14%	17%	1%	-15%	3%	8%	17%	5%
其他（亿美元）		1.4	0.7	0.8	0.8	0.7	0.7	0.9	0.8	0.9
	YoY	-66%	-69%	-56%	-51%	-53%	0%	7%	1%	33%
	QoQ	-11%	-48%	15%	-8%	-14%	11%	23%	-13%	13%
数据中心占比		57%	65%	60%	60%	76%	80%	83%	87%	87%
游戏占比		30%	27%	30%	31%	18%	16%	13%	10%	10%
可视化占比		7%	3%	4%	4%	3%	2%	2%	2%	2%
汽车占比		3%	4%	5%	4%	2%	1%	1%	1%	1%
其他占比		2%	1%	1%	1%	0%	0%	0%	0%	0%

来源：彭博，英伟达财报，中泰证券研究所

1、FY25Q2 财务情况：业绩超预期，数据中心拉动增长

➢ **分地区看：各地区贡献营收均高速增长，美国贡献营收大头。**FY25Q2 美国地区收入130亿美元，yoy+115%，qoq-4%，占比43%；中国大陆（含香港）收入37亿美元，yoy+34%，qoq+47%，占比12%；中国台湾收入57亿美元，yoy+102%，qoq+31%，占比19%；其他国家收入76亿美元，yoy+304%，qoq+34%，占比25%。

图表：分地区业绩情况

	FY23Q2	FY23Q3	FY23Q4	FY24Q1	FY24Q2	FY24Q3	FY24Q4	FY25Q1	FY25Q2
美国（亿美元）	20	21	22	24	60	63	122	135	130
YoY	100%	91%	52%	23%	204%	193%	450%	466%	115%
QoQ	3%	8%	4%	7%	153%	4%	94%	10%	-4%
中国大陆（含香港）（亿美元）	16	11	10	16	27	40	19	25	37
YoY	-7%	-43%	-52%	-24%	71%	251%	104%	57%	34%
QoQ	-23%	-28%	-17%	67%	72%	47%	-52%	28%	47%
中国台湾（亿美元）	12	12	19	18	28	43	44	44	57
YoY	-39%	-47%	-29%	-35%	136%	276%	140%	143%	102%
QoQ	-57%	-4%	61%	-3%	58%	53%	2%	-1%	31%
其他国家（亿美元）	19	15	10	14	19	35	35	57	76
YoY	4%	-16%	-36%	-5%	-1%	133%	241%	300%	304%
QoQ	28%	-22%	-31%	39%	33%	83%	1%	63%	34%
美国占比	30%	36%	37%	33%	45%	35%	55%	52%	43%
中国大陆（含香港）占比	24%	19%	16%	22%	20%	22%	9%	10%	12%
中国台湾占比	18%	19%	31%	25%	21%	24%	20%	17%	19%
其他国家占比	28%	25%	17%	20%	14%	19%	16%	22%	25%

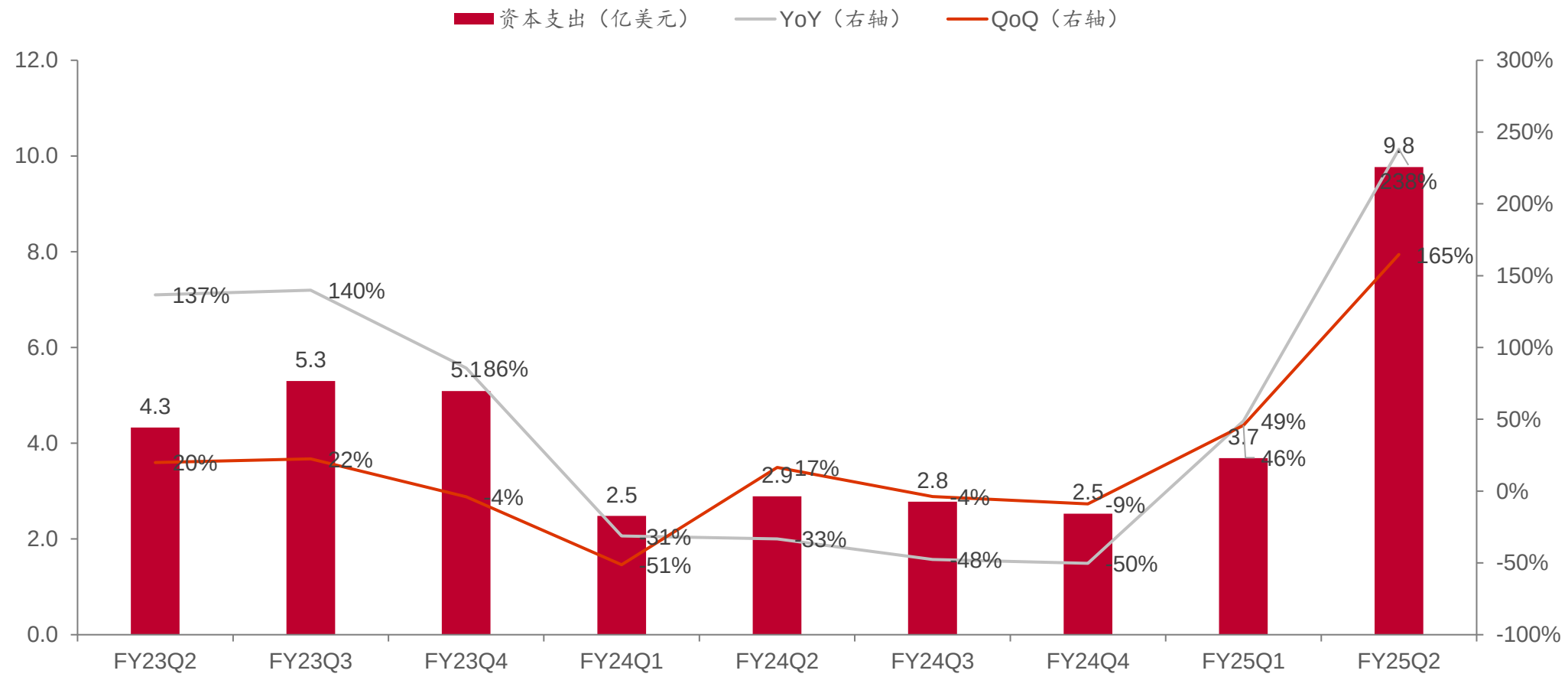
来源：彭博，英伟达财报，中泰证券研究所

1、FY25Q2 财务情况：业绩超预期，数据中心拉动增长

■ **资本支出超预期大幅增加。**

➢ 资本支出：FY25Q2资本支出9.8亿美元，yoy+238%，qoq+165%，此前预期4.3亿美元，超预期127%。

图表：资本支出及增速



来源：彭博，中泰证券研究所

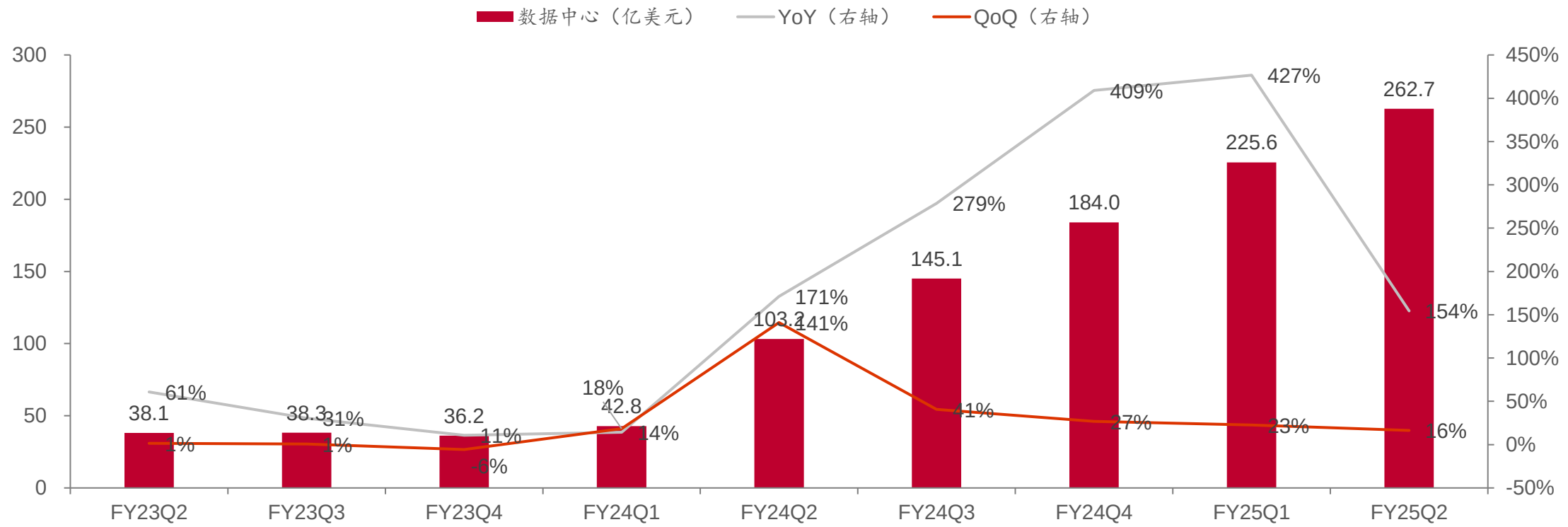
目 录

2、分业务：Hopper需求强劲推动数据中心收入高增

■ **Hopper GPU需求强劲推动数据中心收入高增。**

➢ 数据中心：该业务借助NVIDIA加速计算平台对硬件和软件进行集成，为企业数据中心提供服务。FY25Q2收入262.7亿美元，创历史新高，yoy+154%，qoq+16%，实现连续6个季度环比大幅增长，此前市场预期248.6亿美元，超预期5.7%。同环比增长，主要得益于Hopper GPU计算平台的强劲需求。其中云厂商贡献收入在数据中心收入中占到45%。目前对于H200和Blackwell需求远超供给，公司预计这种情况会持续到明年。

图表：季度收入及增速



来源：彭博，英伟达财报，中泰证券研究所

2、分业务：数据中心收入高增，计算贡献主要增长

■ **数据中心收入高增，计算贡献主要增长。**

➢ 数据中心分业务情况：FY25Q2计算收入为226.0亿美元，yoy+162%，qoq+17%，占比86%；网络收入为36.7亿美元，yoy+114%，qoq+16%，占比14%。其中网络收入增长，主要得益于InfiniBand和AI以太网收入增长，其中AI以太网收入环比翻倍。

图表：数据中心收入拆分

		FY23Q2	FY23Q3	FY23Q4	FY24Q1	FY24Q2	FY24Q3	FY24Q4	FY25Q1	FY25Q2
数据中心营收（亿美元）		38.1	38.3	36.2	42.8	103.2	145.1	184.0	225.6	262.7
计算营收（亿美元）		29.2	28.0	25.6	33.6	86.1	118.7	150.7	193.9	226.0
	YoY	-	-	-	-	195%	324%	488%	478%	162%
	QoQ	-	-4%	-8%	31%	157%	38%	27%	29%	17%
网络营收（亿美元）		8.9	10.3	10.5	9.3	17.1	26.5	33.3	31.7	36.7
	YoY	-	-	-	-	94%	155%	217%	242%	114%
	QoQ	-	17%	2%	-12%	85%	52%	28%	-5%	16%
计算营收占比		77%	73%	71%	78%	83%	82%	82%	86%	86%
网络营收占比		23%	27%	29%	22%	17%	18%	18%	14%	14%

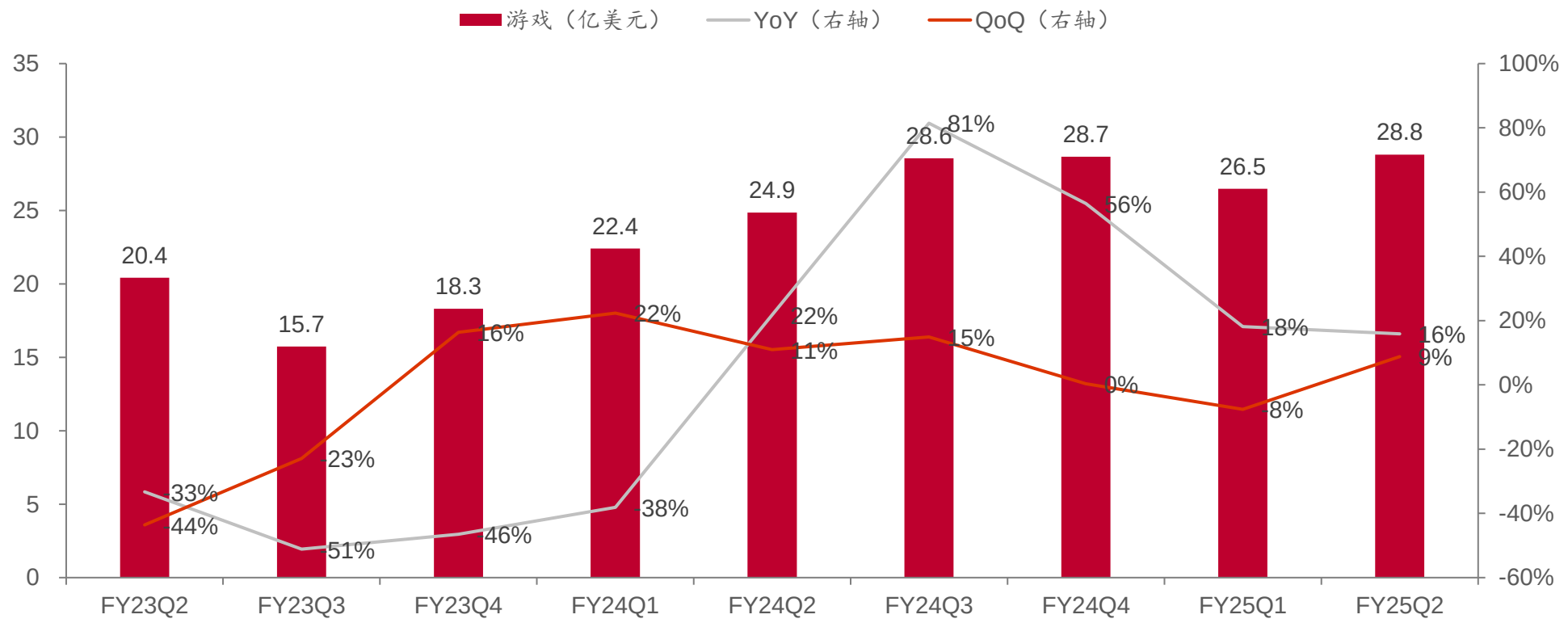
来源：彭博，英伟达财报，中泰证券研究所

2、分业务：GeForce RTX40和游戏机SoC推动增长

■ **GeForce RTX 40系列GPU和游戏机SoC需求增加推动游戏业务收入同环比增长。**

➢ 游戏：该业务为游戏开发提供端到端企业解决方案。FY25Q2收入28.8亿美元，yoy+16%，qoq+9%，此前市场预期28.0亿美元，超预期3%。同环比增长得益于GeForce RTX 40系列GPU和游戏机SoC销量增加。目前整个产品的系列的终端需求和渠道库存保持健康。其中需求方面，英伟达和微软达成合作，此前宣布LLMs在NVIDIA GeForce RTX AI PC上的运行速度提高了3倍。此外，网易游戏、腾讯游戏开发商正在使用NVIDIA ACE来创造逼真的游戏角色，以提升玩家的体验。

图表：季度收入及增速



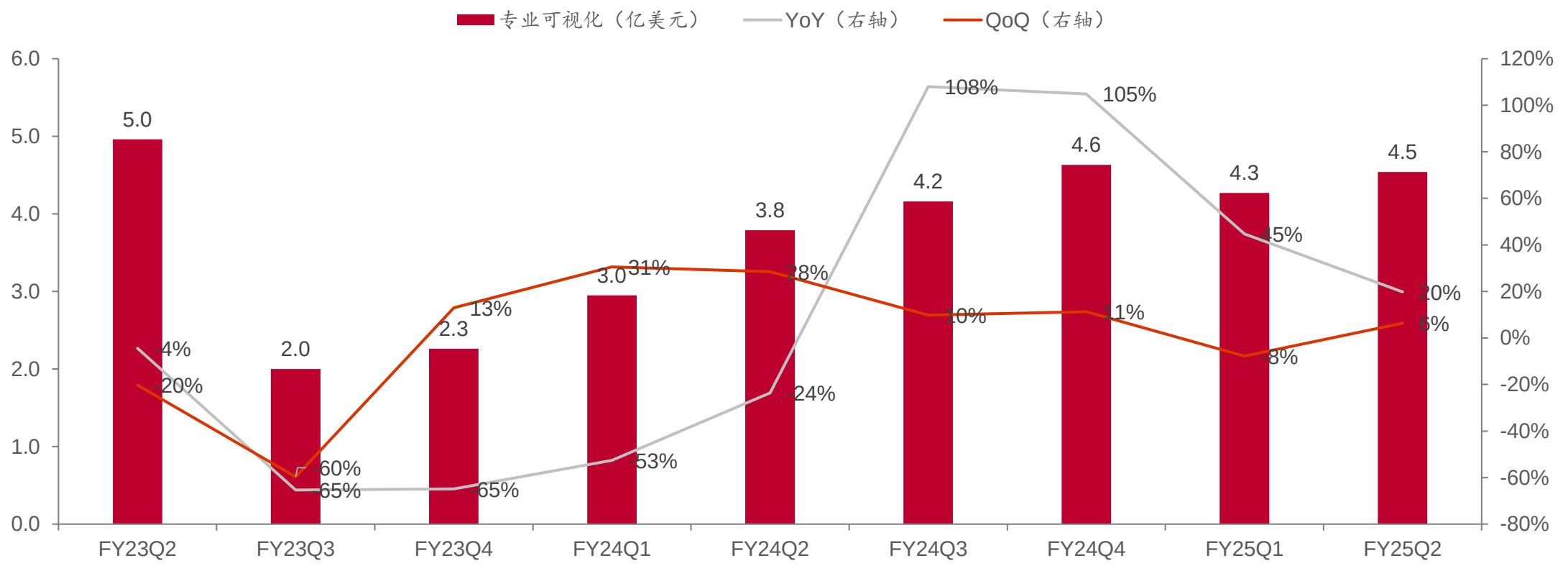
来源：彭博，英伟达财报，中泰证券研究所

2、分业务：RTX GPU需求增长

■ **RTX GPU需求增长推动专业可视化收入。**

➢ 专业可视化：该业务涵盖专业图形渲染、云端XR应用等多个方面，产品包含专用图形显卡、计算卡等。FY25Q2收入4.5亿美元，yoy+20%，qoq+6%，此前市场预期4.5亿美元，超预期0.5%。同比增长主要得益于RTX GPU需求的持续增长。

图表：季度收入及增速



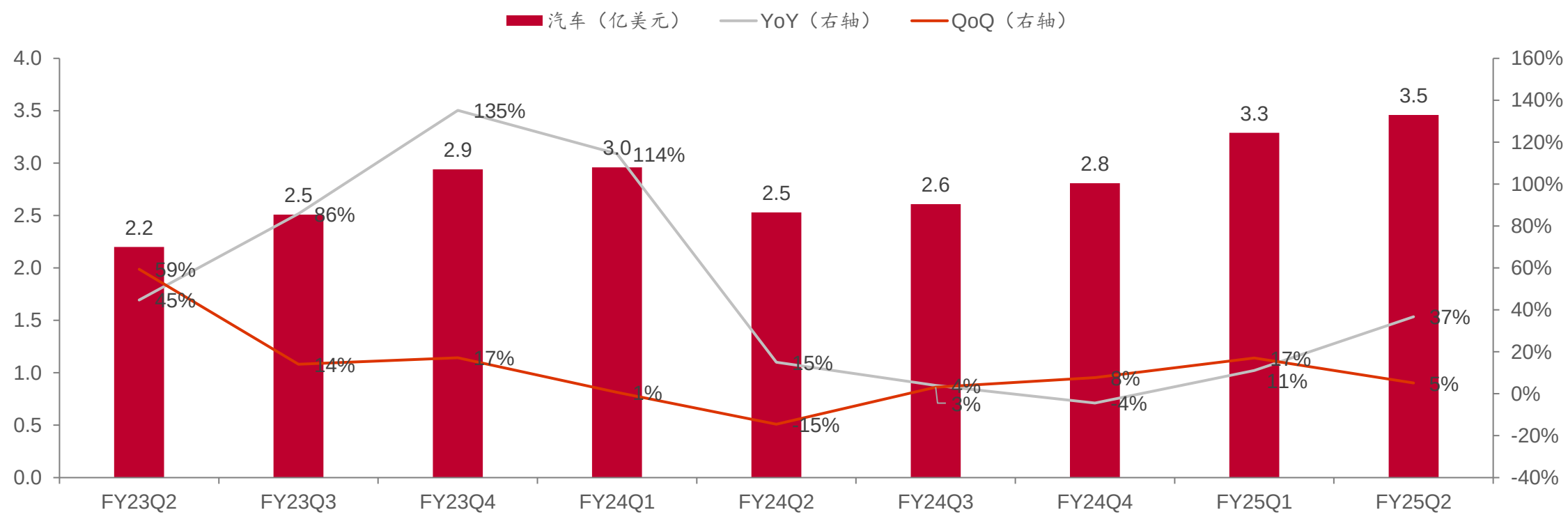
来源：彭博，英伟达财报，腾讯新闻，中泰证券研究所

2、分业务：自动驾驶平台推动汽车收入同环比增长

■ **自动驾驶平台推动汽车收入同环比增长。**

➢ 汽车：该业务主要聚焦于向用户提供自动驾驶汽车的开发和部署的端到端平台，即NVIDIA DRIVE平台。FY25Q2汽车收入3.5亿美元，yoy+37%，qoq+5%，实现连续4个季度环比增长，此前市场预期3.5亿美元，符合预期。其中同比增长主要系自动驾驶平台推动。目前，客户中，NVIDIA DRIVE Orin帮助小米推出首款电动汽车SU7。并且，基于NVIDIA DRIVE Thor芯片，与比亚迪、小鹏、广汽等多家领先的电动汽车制造商进行合作，预计DRIVE Thor将于明年开始量产。

图表：季度收入及增速



来源：彭博，英伟达财报，腾讯新闻，中泰证券研究所

目 录

3、Blackwell和数据中心平台助力实现万亿参数AI

■ **Blackwell：**FY25Q2公司已向客户交付样品，为提高生产良率，对Blackwell GPU掩膜进行了改动。Blackwell的生产爬坡计划于第四季度开始，一直持续到FY2026，并预计FY25Q4 Blackwell开始贡献数十亿美元的收入。

图表：NVIDIA数据中心GPU参数对比

	A100 PCIe/SXM	A2	A10	A16	A30	A40	H100 SXM	H100 Pcie	H100 NVL	L4	L40s	H200 SXM	H200 NVL	B100	B200	GB200
FP32（TFLOPS）	19.5	4.5	31.2	18	10.3	37.4	67	51	134	30.3	91.6	67	67	-	-	-
TF32 Tensor Core（TFLOPS）	156 | 312	9 | 18	62.5 | 125	36	82 | 165	74.8 | 149.6	989	756	1979	120	183 I 366	989	989	0.9 I 1.8 PFLOPS	1.12 I 2.25 PFLOPS	2.5 I 5 PFLOPS
FP16 Tensor Core（TFLOPS）	312 | 624	18 | 36	125 | 250	71.6 | 143.6	165 | 330	149.7 | 299.4	1979	1513	3958	242	362.05 I 733	1979	1979	1.8 I 3.5 PFLOPS	2.25 I 4.5 PFLOPS	5 I 10 PFLOPS
FP8 Tensor Core（TFLOPS）	-	-	-	-	-	-	3958	3026	7916	485	733 I 1,466	3958	3958	3.5 I 7 PFLOPS	4.5 I 9 PFLOPS	10 I 20 PFLOPS
GPU 显存	80GB HBM2e	16GB GDDR6	24GB GDDR6	64GB GDDR7	24GB HBM2	48GB GDDR6	80GB	80GB	188GB	24GB	48GB GDDR6	141GB	141GB	192GB	192GB	384GB
GPU 显存带宽	1935/2039 GB/s	200GB/s	600 GB/s	800 GB/s	933 GB/s	696 GB/s	3.35TB/s	2TB/s	7.8TB/s	300GB/s	864GB/s	4.8TB/s	4.8TB/s	8TB/s	8TB/s	16TB/s
最大热设计功耗 (TDP)	300/400瓦	60瓦	150瓦	250瓦	165瓦	300瓦	700 瓦	350 瓦	400瓦	72瓦	350瓦	700瓦	600瓦	700瓦	1000瓦	2700瓦
外形规格	PCIe双插槽风冷式或单插槽液冷式/SXM	单插槽，半高PCIe	单插槽FHFL	全高、全长（FHFL）双插槽	双插槽、全高、全长（FHFL）	4.4 吋（高）x 10.5 吋（长），双插槽	SXM	PCIe，双插槽，风冷式	2×Pcle，双插槽，风冷式	单插槽半高，PCIe	4.4 吋（高）x 10.5 吋（长），双插槽	SXM	Pcle	-	-	-
互连技术	NVLink：600GB/s+PCIe 4：64GB/s	PCIe 4 x8	PCIe 4.0：64GB/s	-	第4代 PCIe：64GB/s+第3代 NVLINK：200GB/s	NVLink：112.5GB/s (双向)+ PCIe Gen4: 64GB/s	NVLink：900GB/s +PCIe 5: 128GB/s	NVLink：600GB/s +PCIe 5: 128GB/s	NVLink: 600GB/s +PCIe 5:128GB/s	16×PCIe 4：64GB/s	16×PCIe 4：64GB/s（双向）	NVLink: 900 GB/s +PCIe 5: 128 GB/s	2-way or 4-way NVLink: 900 GB/s+PCIe 5: 128 GB/s	NVLink：1.8TB/s	NVLink：1.8TB/s	NVLink：2×1.8TB/s
发布时间	2020	2021	2021	2021	2021	2022	2022	2022	2022	2023	2023	2023	2023	2024	2024	2024

来源：英伟达官网，百度，腾讯新闻，smzdm网，IT之家，CSDN，澎湃新闻，华尔街见闻，中泰证券研究所

3、Blackwell和数据中心平台助力实现万亿参数AI

- **NVIDIA Blackwell能够在万亿参数的大型语言模型上实现实时生成式人工智能，其TCO（整体成本）和能耗比上一代低25倍，Blackwell提供比H100快4倍的训练速度和30倍的推理速度。**
  - GB200 NVL72是一个多节点、液冷、机架规模的系统，结合了36个GB200超级芯片作为单个GPU。每个GB200连接两个B200 GPU和一个Grace CPU。其中B200 GPU由两个Die组成，通过10TB/s的高宽带接口NV-HBI连接成一个GPU，拥有2080亿个晶体管。此外，8个B200 GPU通过NV-Link可组成用于x86系统的HGX B200。目前，Amazon, Google, Meta, Microsoft, OpenAI, Oracle, Tesla以及其他的AI公司都希望采用Blackwell。

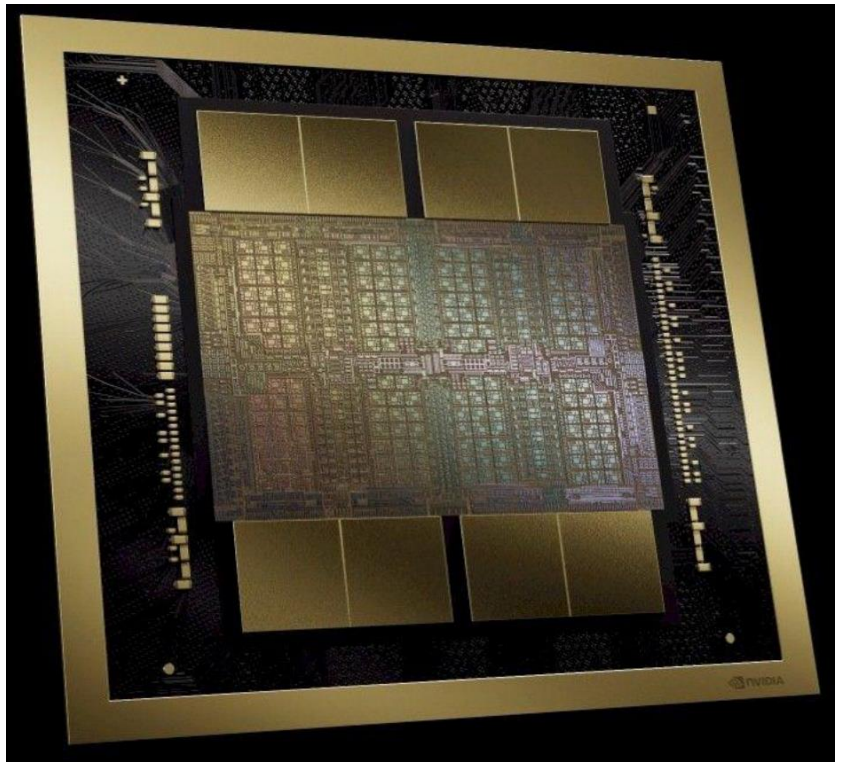
图表：NVIDIA Blackwell

来源：英伟达官网，中泰证券研究所

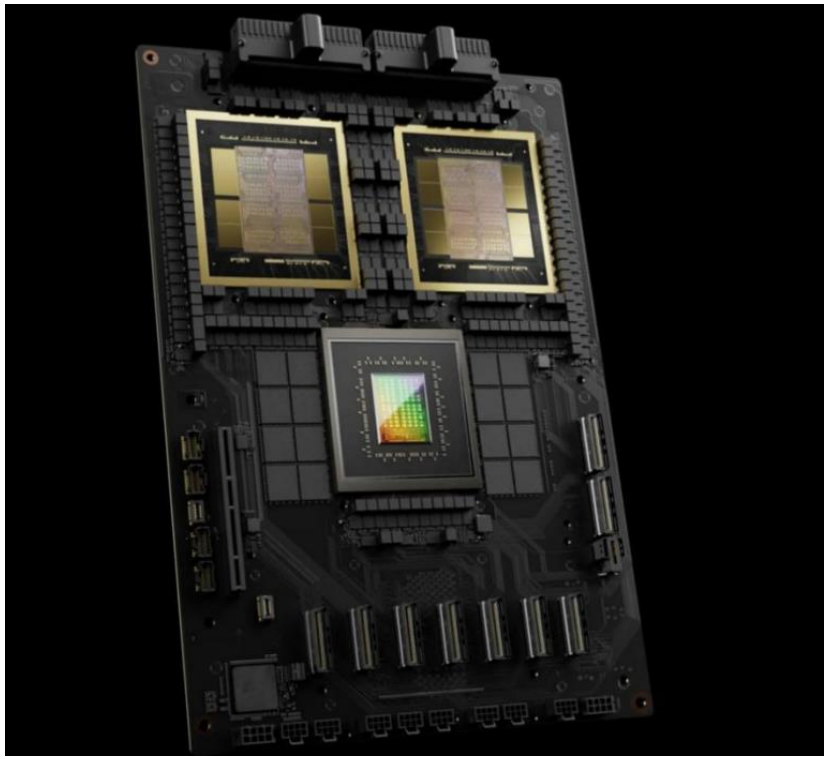
3、基于B102、B100芯片生成3款服务器主板

■ **基于B102、B100芯片生成3款服务器主板：B200A、B200、GB200。**目前英伟达针对Blackwell架构的基础芯片有两款B102与B100，对应服务器主板有三款：B200A、B200、GB200。其中B102是“重新设计”的芯片，是所有Blackwell芯片的基础 ,它由一个GPU die+4个HBM3e组成，B100则是由两个B102组成 ，即两个GPU die+8个HBM3e组成。B200由1个B100芯片组成，B200A由1个B102芯片组成，GB200由两颗B100+一颗Grace GPU组成。此外，基于不同主板和系统配件（液冷、铜缆等）组成不同款式服务器系统。

图表：B100结构图



图表：GB200结构图



来源：腾讯科技，华尔街见闻，英伟达官网，腾讯新闻，中泰证券研究所

3、B100（HGX）和B200（HGX）被B200A替代

- **中级增强版：**B200 Ultra采用CoWoS-L封装，包含高达288GB的12层HBM3E。此外，FLOPS性能方面提升了高达50%。B200A Ultra则会有更高的FLOPS，但在显存上不会进行升级。
- **GB NVL：**在原来的GB系列基础上增加了GB200A Ultra NVL36，GB200A NVL36是一款风冷40kW/机架服务器，配备36个通过NVLink互连的GPU。每个机架配备9个计算托盘和9个NVSwitch托盘。每个计算托盘为2U，包含1个Grace CPU和4个700W的B200A。每个1U NVSwitch托盘有1个交换机ASIC，每个交换机ASIC的带宽为28.8 TB/s。

图表：NVIDIA Blackwell芯片对比

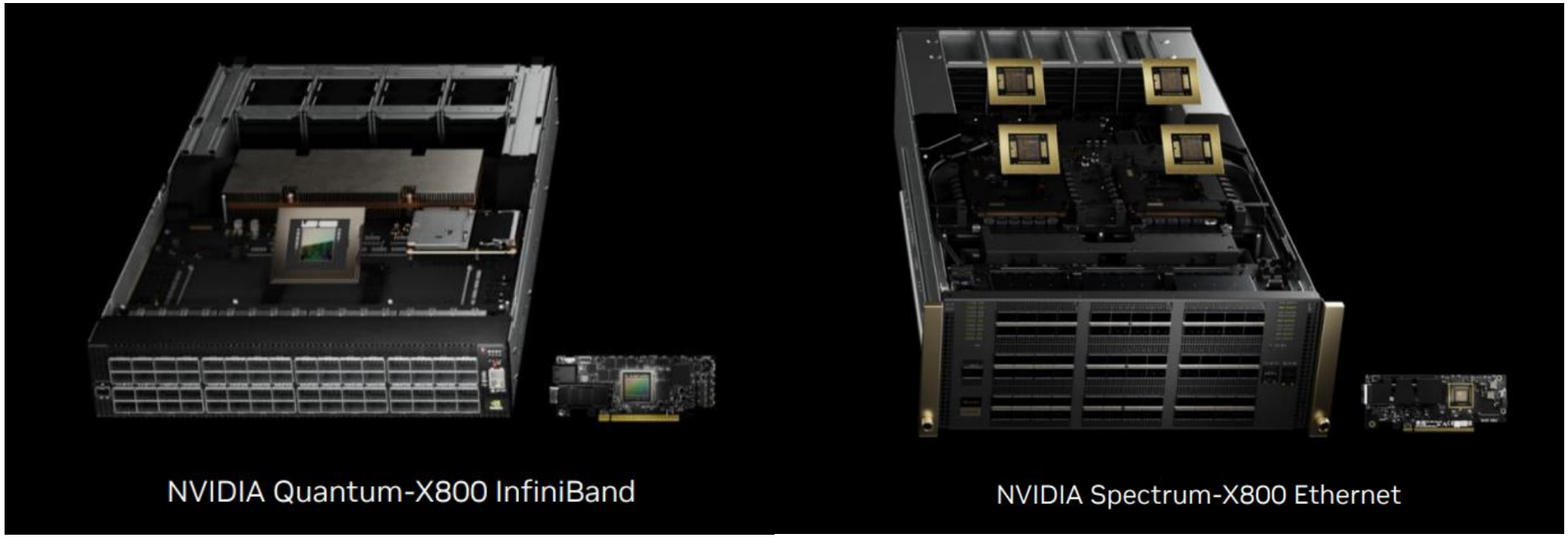
	B100 Blackwell	B200 Blackwell	B200A Blackwell	B200 Blackwell Ultra	B200A Blackwell Ultra	B20 Blackwell
Codename	B100	B200	B102/B200A	B210	B102 Ultra/B210A	B102/B200A
Packaging	CoWoS-L	CoWoS-L	CoWoS-S	CoWoS-L	CoWoS-S	CoWoS-S
HBM(GB)	Up to 192	192	Up to 144	288	144	96
Logic Dies	2	2	1	2	1	1
Power(W)	700	1000(HGX) 1200(GB NVL)	700/1000	1000(HGX) 1200(GB NVL)	1000(HGX) 700(GB NVL)	300
8-GPU HGX	Small Volume	Small Volume	Yes	Yes	Yes	Yes
GB NVL	N/A	72/36	N/A	72/36	36(air-cooled)	Yes

来源：36氪，中泰证券研究所

3、为人工智能设计的新一代NVIDIA网络

- **NVIDIA Quantum-X800 InfiniBand和NVIDIA Spectrum-X800以太网是世界上第一个能够实现端到端800Gb/s吞吐量的网络平台，这些平台上的软件可在各类数据中心加速AI、云、数据处理和HPC应用程序，以及新的Blackwell产品。**
  - GB200供电的系统可以与NVIDIAQuantum-X800 InfiniBand或Spectrum-X800以太网交换机连接，以获得最高的人工智能性能。 此外，HGX B200通过NVIDIA Quantum-2 InfiniBand和Spectrum-X以太网网络平台支持高达400Gb/s的网络速度。
  - 该产品下游客户优质，Quantum InfiniBand和Spectrum-X的使用客户包括Microsoft Azure、Oracle Cloud Infrastructure和CoreWeave。

图表：NVIDIA Quantum-X800 InfiniBand& Spectrum-X800 Ethernet展示



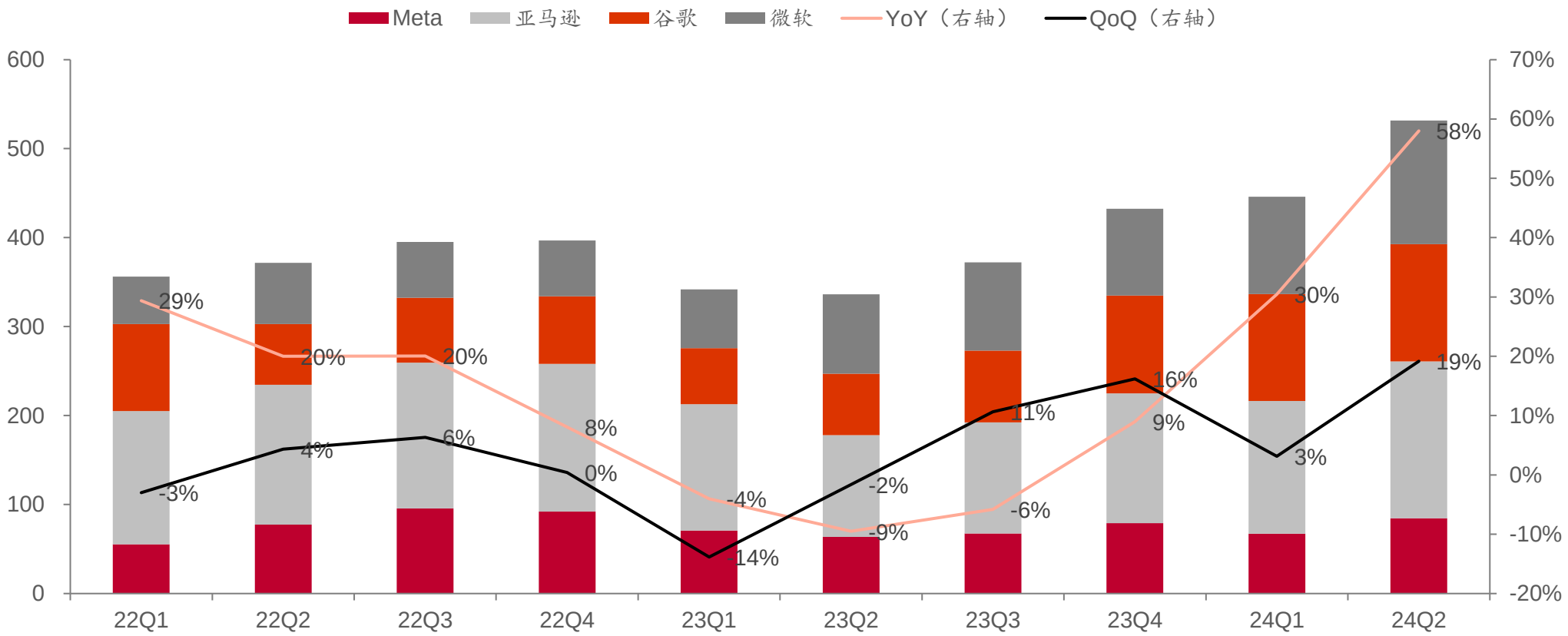
来源：英伟达官网，中泰证券研究所

3、四大云厂商资本支出均同环比高增

■ **四大云厂商资本支出合计532亿美元，创历史新高，yoy+58%，qoq+19%。**

➢ Meta24Q2资本支出（含融资租赁）为85亿美元，yoy+33%，qoq+26%；亚马逊资本支出176亿美元，yoy+54%，qoq+18%；谷歌资本支出132亿美元，yoy+91%，qoq+10%；微软资本支出139亿美元，yoy+55%，qoq+27%。

图表：云厂商资本支出（亿美元）



来源：wind，Meta财报，亚马逊财报，彭博，中泰证券研究所

3、全球云厂商24年资本支出有望大幅上升

- **Meta上修24年资本支出中值，预计24年资本支出370-400亿美元，中值385，此前为350-400亿美元，中值375亿美元。亚马逊称，24H1在网络服务云部门的数据中心等资本支出上已花费350亿美元，预计下半年资本支出会高于350亿美元。微软预计FY25Q1（CY24Q3）资本支出环比上升，并预计25财年资本支出将高于24财年。谷歌称，投资不足的风险远远大于过度投资的风险。**
- **四大云厂商Q2资本开支超预期，预计2024年资本支出有望大幅增长。**我们预计美国四大互联网公司整体24年的资本支出会从23年近1500亿美元增长至24年的超2000亿美元。考虑到普通服务器增速放缓，预计增长将主要由AI拉动，体现为AI硬件方面投资。

图表：资本支出预测

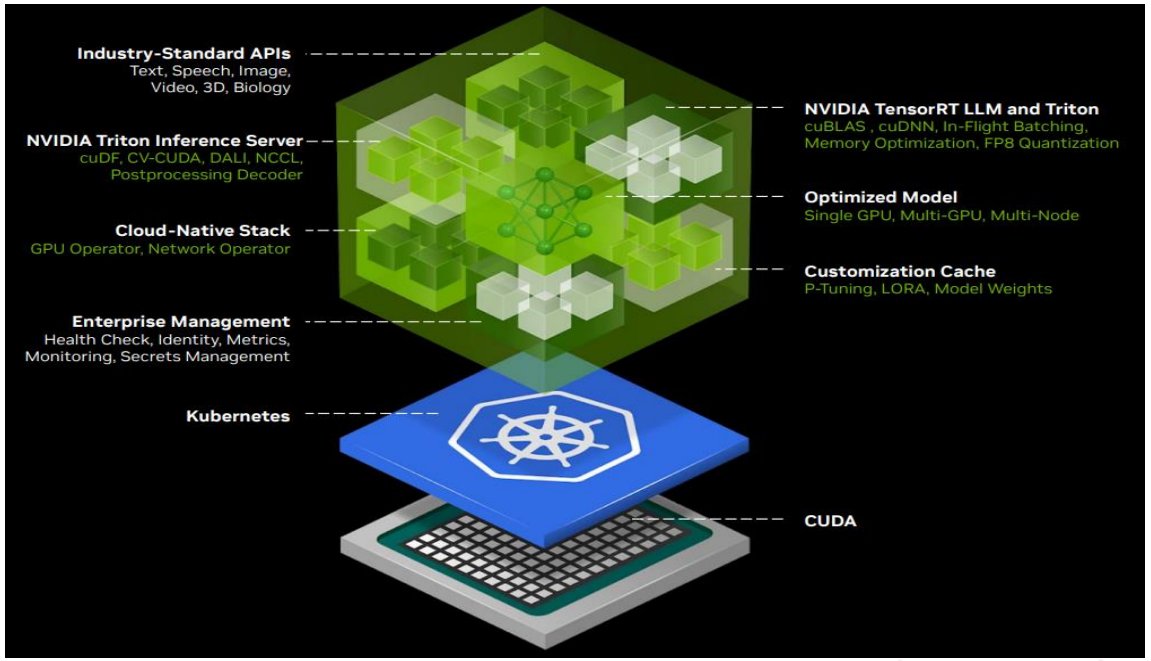
公司		资本开支							
		23Q1	23Q2	23Q3	23Q4	2023	24Q1	24Q2	2024E
微软（亿美元）		66	89	99	97	352	110	139	500
	YOY	24%	30%	58%	55%	42%	66%	55%	42%
谷歌（亿美元）		63	69	81	110	323	120	132	500
	YOY	-36%	1%	11%	45%	2%	91%	91%	55%
Meta（亿美元）		71	64	68	79	281	67	85	370-400
	YOY	26%	-18%	-30%	-15%	-13%	-6%	31%	37%
亚马逊（亿美元）		142	115	125	146	527	149	176	650
	YOY	-5%	-27%	-24%	-12%	-17%	5%	54%	23%
合计（亿美元）		**342**	**336**	**372**	**432**	**1483**	**446**	**532**	**2035**
注释：微软以CY统计	YOY	**-4%**	**-9%**	**-6%**	**9%**	**-2%**	**30%**	**58%**	**37%**

来源：wind，彭博、Meta财报，亚马逊财报，华尔街见闻，msn.cn，36氪，新格致产业，中泰证券研究所

3、NVIDIA NIM推理和CUDA-X微服务助力开发人员

- **NVIDIA NIM微服务和CUDA-X微服务可帮助开发人员快速构建和部署AI和加速应用程序。**
  - NIM推理微服务提供由NVIDIA推理软件(包括Triton和TensorRT-LLM)支持的预构建容器，以及用于医疗等领域的行业标准API。
  - CUDA-X微服务可在数据处理、AI 和 HPC 应用中轻松集成、定制和部署以下这些微服务：NVIDIA Riva（用于可自定义的语音和翻译 AI）、NVIDIA Earth-2（用于高分辨率气候和天气仿真）、NVIDIA cuOpt（用于路由优化）以及 NVIDIA NeMo Retriever（用于为企业提供响应式检索增强型生成 (RAG) 功能）。
  - 企业可以使用NVIDIA AI Enterprise 5.0部署生产级NIM微服务。目前，Adobe、Cadence、CrowdStrike、Getty Images、SAP、ServiceNow和Shutterstock都是首批使用新微服务的公司。

图表：NVIDIA NIM 推理微服务



来源：英伟达官网，中泰证券研究所

3、在不同行业中扩大使用NVIDIA AI和加速计算

■ **在不同行业中扩大使用NVIDIA AI和加速计算。**

- 企业软件和数据平台：将加速计算和通用人工智能功能引入企业软件和数据平台，如Box、Cloudera、Cohesity、Dropbox、NetApp、SAP等。
- 医疗保健：通过AWS和微软Azure集成超20个NVIDIA医疗保健微服务，用于药物发现、医疗技术和数字健康。与强生医疗科技公司合作，将NVIDIA IGX和Holoscan医疗器械边缘人工智能平台用于其连接的外科数字生态系统。
- 工业数字化：推出全新的Omniverse Cloud API，用于数字孪生和仿真应用开发，被Ansys、Cadence、达索系统、西门子采用，将于今年晚些时候在微软Azure上推出。
- EDA和CAE：EDA和CAE应用程序通过NVIDIA加速后，提供超过10倍的速度，目前被Ansys、Cadence和Synopsys采用。
- 计算光刻：作为半导体制造过程中最密集的计算工作量，可通过NVIDIA cuLitho加速了40-60倍。目前已于台积电和新思科技投入生产。

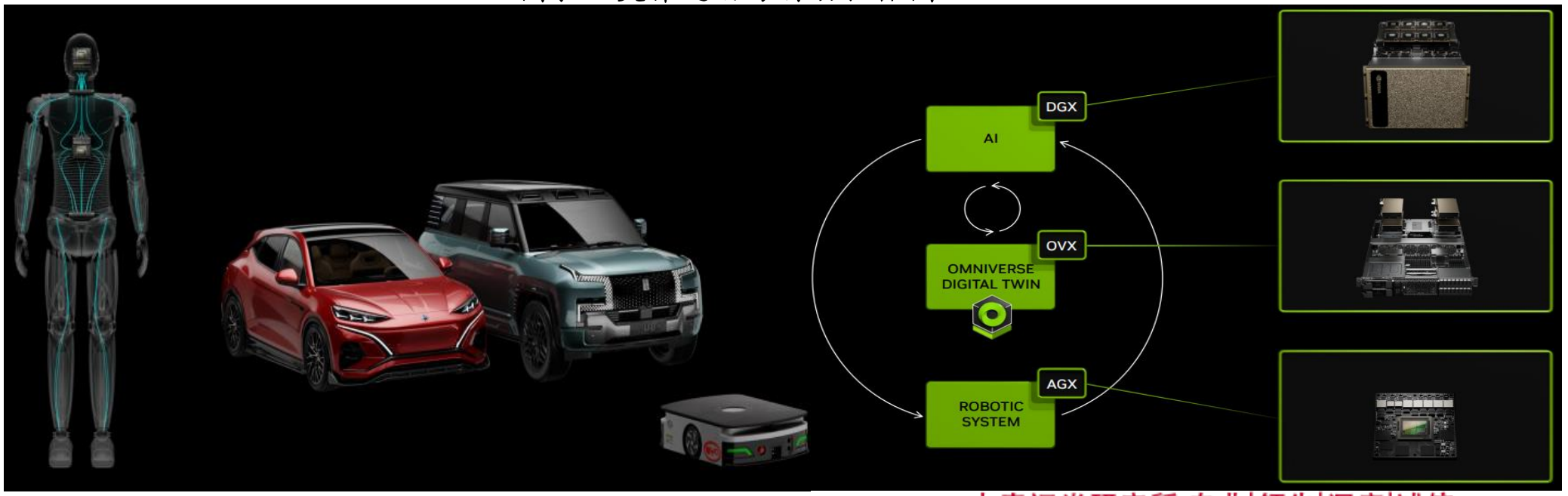
图表：不同行业工作时使用NVIDIA AI和加速计算

来源：英伟达官网，中泰证券研究所

3、英伟达助力下一代自动机器

- **全球领先的汽车公司选择NVIDIA DRIVE Thor为其下一代消费者和商业车队提供动力，从新能源汽车和卡车到出租车、巴士和自动送货车辆。**
  - 汽车方面：DRIVE Thor是一个基于Blackwell的集中式车载计算平台。它提供了丰富的驾驶舱功能以及安全可靠的高度自动化和自动驾驶。目前，BYD、广汽、小鹏、Plus、Nuro、Waab以及蔚来汽车等都采用DRIVE Thor，预计DRIVE Thor最早于25年投产。
  - 机器人方面，英伟达的GROOT项目是一个通用的基础模型，可以让机器人能够理解自然语言，并通过观察人类的行为来模仿动作。目前，已对NVIDIA Isaac机器人平台进行重大升级，以支持领先的人形机器人公司，包括1X Technologies, Agility Robotics, Apptronik等。

图表：英伟达助力自动机器图

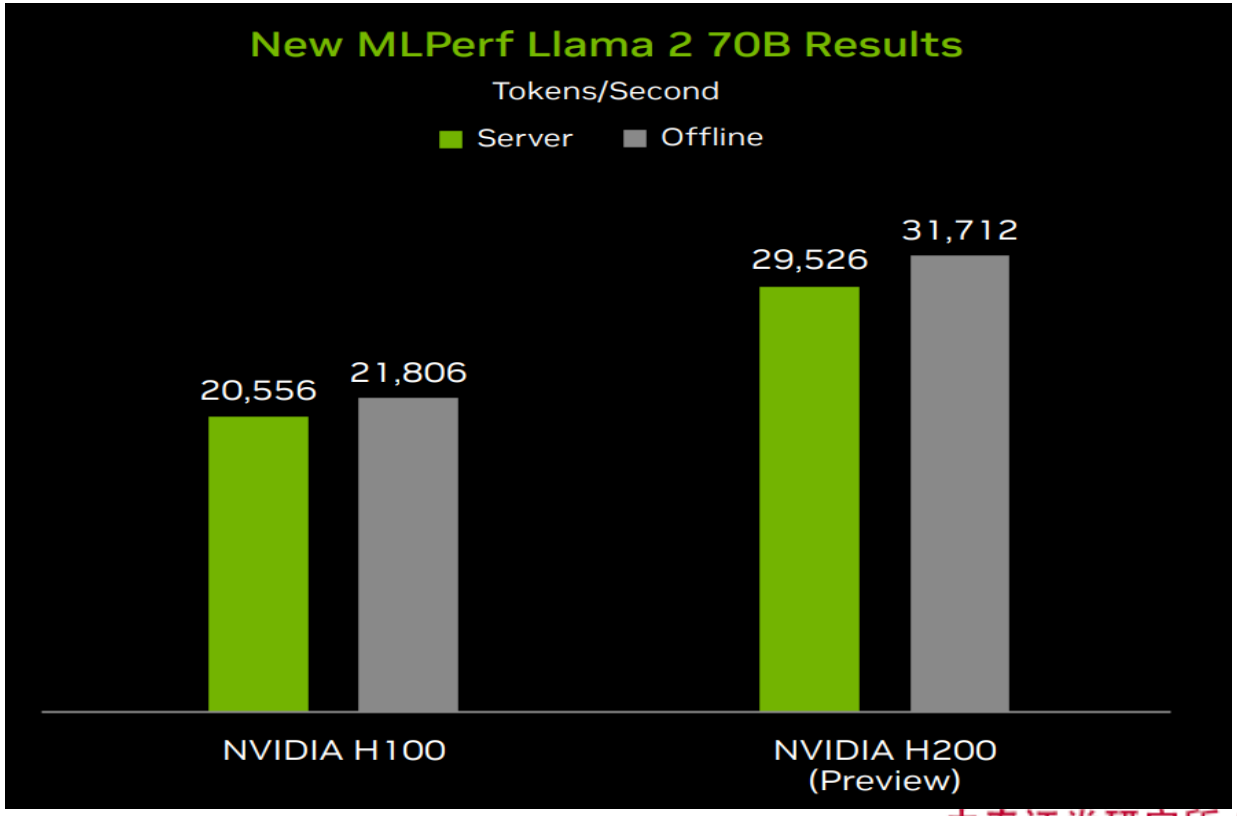


来源：英伟达官网，中泰证券研究所

3、NVIDIA Hopper推理性能再次提升

- **在最新的人工智能推理MLPerf基准测试中，NVIDIA Hopper在人工智能推理的每一次测试中都获得领先，展示了NVIDIA的芯片、系统和软件平台在处理运行通用人工智能面对苛刻要求时的强大能力。**
  - NVIDIA TensorRT-LLM：该软件加速和简化了大型语言模型上复杂的推理工作，提高了NVIDIA Hopper在GPT-J LLM上的性能，比六个月前的结果提高了近3倍。并且，NVIDIA H200 GPU在其MLPerf首次亮相时，在Llama270b基准上创下了纪录。

图表：在MLPerf中的Llama 270b推理速度提高45%



来源：英伟达官网，中泰证券研究所

4、风险提示

- 行业需求不及预期的风险：若包括手机、PC、可穿戴等终端产品需求不及预期，则产业链相关公司的业绩增长可能不及预期。
- 大陆厂商技术进步不及预期、中美贸易摩擦加剧、研报使用的信息更新不及时的风险、报告中各行业相关业绩增速测算未剔除负值影响，计算结果存在与实际情况偏差的风险、行业数据或因存在主观筛选导致与行业实际情况存在偏差风险。

重要声明

- **中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。**
- 本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。
- 市场有风险，投资需谨慎。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。
- 投资者应注意，在法律允许的情况下，本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。
-