电子

周观点：谷歌发布 TPU v8，光互联需求升级

谷歌 Cloud Next 2026：发布 TPU v8，光互联价值凸显。分为专为模型训练设计的 TPU 8t 与专为推理优化的 TPU 8i，围绕网络架构和互联架构进行优化，互联在集群中地位凸显，带动光模块、OCS 需求提升。

1）TPU 8t：大规模预训练旗舰芯片，引入 Virgo 互联架构提升训练带宽。采用 3D Torus 网络拓扑结构，单个 Pod 最多可集成 9600 块 TPU 8t 芯片，配备 2 PB 高带宽内存，每 Pod 计算性能达 121 exaflops，较上一代 Ironwood 提升约 3 倍，同等价格下性能提升 2.8 倍。引入 Virgo 互联架构，训练场景数据中心网络带宽最高提升至前代 4 倍，单套网络可互联 13.4 万颗 TPU 8t 芯片，芯片间互联带宽较上一代提升 2 倍。

2）TPU 8i：专注于 AI 推理场景，采用全新 Broadfly 架构降低集群推理延迟。配备容量最高片上 SRAM，是原来的 3 倍；单个 Pod 可扩展至 1152 块芯片，较 Ironwood 同等价格下性能提升 80%，每瓦性能较上一代提升 117%。采用全新的 Boardfly 层级互联拓扑架构，4 颗芯片环形互联构成基础单元，8 块板卡通过铜缆全互联构成本地算力组，36 个算力组通过 OCS 连接构成最高 1024 颗芯片的集群。在这种结构下，任意两枚芯片之间的通信最多只需经过 7 次跳转，全对全通信延迟改善最高 50%。

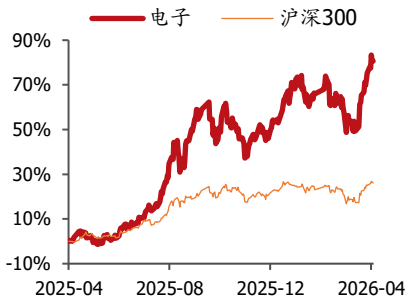
DeepSeek v4 开源发布，国产大模型+国产高端算力深度绑定。DeepSeek-V4 的预览版本正式上线并同步开源，在 Agent 能力、世界知识和推理性能上均实现国内与开源领域的领先。该系列包括 DeepSeek-V4-Pro（1.6T 参数，49B 激活）和 DeepSeek-V4-Flash（284B 参数，13B 激活），两者均支持一百万 Tokens 的上下文长度，在大规模语言模型的效率上迈出了重要一步，能够有效处理超长序列，从而为复杂的长时间跨度任务开辟了新的可能性。根据国家统计局，2026 年 4 月，中国日均 AI Token 调用量突破 140 万亿，两年增长超千倍，国产算力需求正经历指数级增长。国内 AI 算力产业完成了“大模型-芯片-框架”的协同适配验证，我们看好具有供给优势的设计服务厂商、第三方芯片厂商及算力底座的代工厂商及配套产业链。

英特尔 26Q1 业绩超预期，AI 服务器 CPU 需求激增。26Q1，英特尔实现营收 136 亿美元（超出指引上限 127 亿美元），同比增长 7%；GAAP 毛利率 39.4%（大幅超出指引 32.3%），同比提升 2.5 个百分点。分业务来看，数据中心和人工智能产品实现营收 51 亿美元，同比增长 22%；代工业务实现营收 54 亿美元，同比增长 16%。并指引 26Q2 营收指引为 138 亿美元至 148 亿美元，中值同比增长 11%。AI 算力的叙事逻辑正在从“GPU 单极主导”向“CPU+GPU 协同共生”演进。

SK 海力士 26Q1 业绩创新高，重视存储板块业绩释放机遇。26Q1，公司营收达 52.58 万亿韩元，环比增长 60%，同比增幅高达 198%；营业利润为 37.61 万亿韩元，环比增长 96%，同比激增 405%，营业利润率 72%；净利润为 40.35 万亿韩元，净利润率 77%。在 AI 基础设施投资扩大的推动下，市场需求持续强劲。公司通过扩大 HBM、大容量服务器 DRAM 模块、企业级固态硬盘（eSSD）等高附加值产品的销售，延续了业绩上升势

增持（维持）

行业走势



作者

分析师 余凌星
执业证书编号：S0680525010004
邮箱：shelingxing1@gszq.com
分析师 钟琳
执业证书编号：S0680525010003
邮箱：zhonglin1@gszq.com

相关研究

1、《电子：周观点：产业链龙头共振，AI 需求持续高景气》 2026-04-18

2、《电子：周观点：模型升级不断，看好硬件投资机遇》 2026-04-11

3、《电子：周观点：OCS 产业化进入加速期，看好光上游投资机遇》 2026-04-04

头。展望第二季度，预计 DRAM 出货量环比增长高个位数百分比，NAND 出货量（含 Solidigm）环比增长百分之十几（mid teen）。

周观点：见相关标的。

风险提示：下游需求不及预期、研发进展不及预期、地缘政治风险。

重点标的

股票代码	股票名称	投资评级	EPS（元）				PE			
			2025A	2026E	2027E	2028E	2025A	2026E	2027E	2028E
002384.SZ	东山精密	买入	0.76	4.11	8.24	13.19	246.70	45.75	22.80	14.24
300476.SZ	胜宏科技	买入	4.94	11.41	18.59	24.78	52.00	27.65	16.98	12.74
300475.SZ	香农芯创	买入	1.17	12.69	16.78	21.78	133.50	12.77	9.66	7.44
688521.SH	芯原股份	买入	-0.16	0.11	0.51	-	-	2,110.21	448.81	-
688981.SH	中芯国际	买入	0.67	0.85	-	-	145.50	130.09	-	-
688807.SH	优迅股份	买入	1.10	1.39	1.64	1.89	226.90	281.08	238.17	206.62
688167.SH	炬光科技	买入	-0.46	1.08	2.69	-	-	326.09	130.70	-
688012.SH	中微公司	买入	3.59	5.51	7.73	-	62.70	58.70	41.83	-
002371.SZ	北方华创	买入	10.18	13.05	16.49	-	39.00	36.23	28.66	-
300672.SZ	国科微	买入	-1.02	2.28	3.74	-	-	69.65	42.41	-

资料来源：Wind，国盛证券研究所

内容目录

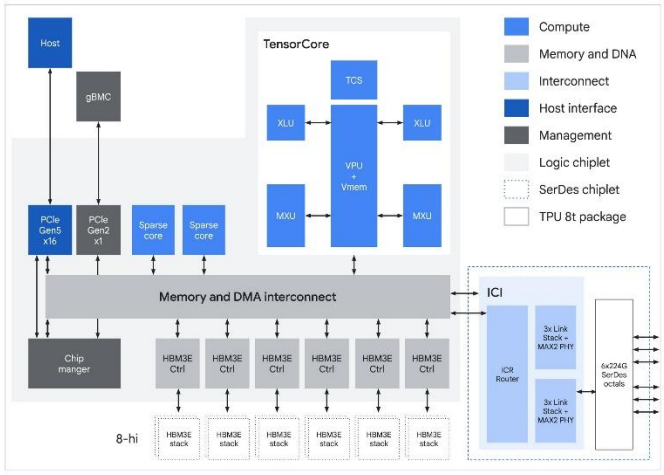
图表目录

/

一、谷歌 Cloud Next '26：推出第八代 TPU

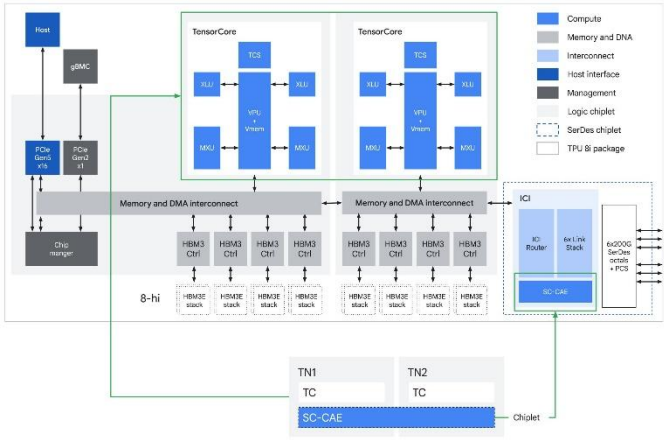
当地时间 2026 年 4 月 22 日，在拉斯维加斯举行的 Google Cloud Next '26 大会上，谷歌正式发布了第八代 TPU，首次将 AI 训练与推理任务拆分至两款独立芯片——TPU 8t（负责模型训练）与 TPU 8i（负责推理优化），标志着其 AI 硬件战略的重大转向。

图表1：TPU 8t ASIC 框图



资料来源：谷歌，国盛证券研究所

图表2：TPU 8i ASIC 框图



资料来源：谷歌，国盛证券研究所

图表3：TPU 8t 相比上一代的核心升级

	Ironwood (2025)	TPU 8t (2026)
Pod size	9,216	9,600
FP4 EFlops per pod	42.5	121
Bidirectional scale-up bandwidth (Tb/s per chip)	9.6	19.2
Scale-out networking bandwidth (Gb/s per chip)	100	400

资料来源：谷歌，国盛证券研究所

图表4：TPU 8i 相比上一代的核心升级

	Ironwood (2025)	TPU 8i (2026)
Pod size	256	1,152
FP8 EFlops per pod	1.2	11.6
Total HBM capacity per pod (TB)	49.2	331.8
Bidirectional scale-up bandwidth (Tb/s per chip)	9.6	19.2

资料来源：谷歌，国盛证券研究所

一、TPU 8t：专为大规模预训练和嵌入密集型工作负载进行了优化

采用 TPU 中常用的 3D Torus 网络拓扑结构，单个超级节点中集成规模从 9216 个扩展到 9600 个芯片。TPU 8t 旨在实现数百个超级节点的最大吞吐量，确保训练运行按计划进行。相较于上一代 TPU，TPU 8t 的核心突破包括：

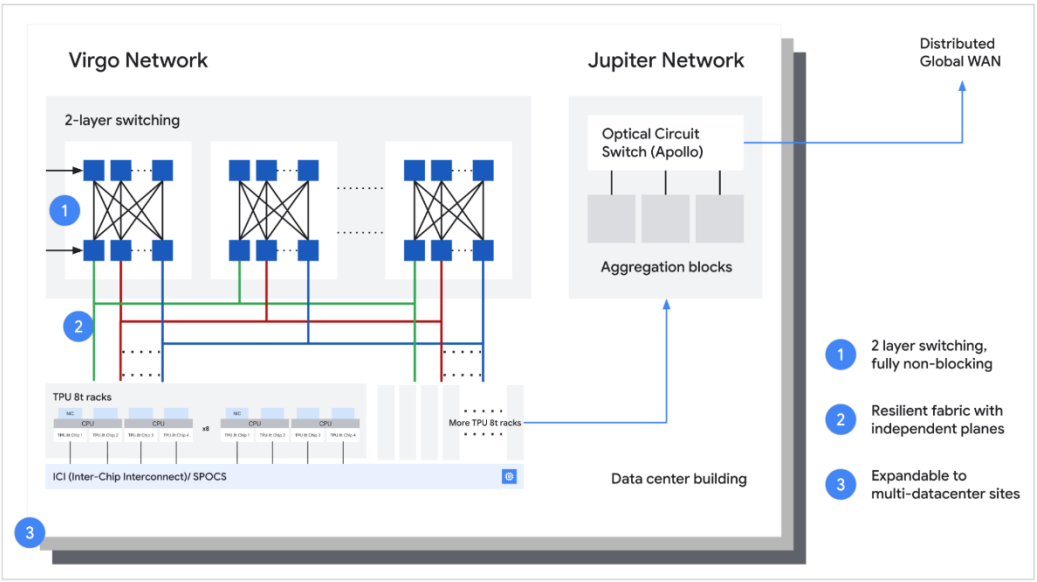
- **SparseCore 的优势**：专门设计的加速器，用于处理嵌入查找中不规则的内存访问模式。矩阵乘法单元（MXU）负责处理矩阵运算，而 SparseCore 则负责卸载数据相关的全集操作以及其他一些集中操作，从而避免了通用芯片中常见的零操作瓶颈。
- **VPU/MXU 重叠和均衡扩展**：通过实现更均衡的向量处理单元（VPU）扩展，该架构最大限度地减少了向量运算时间。使得量化、softmax 和 LayerNorm 等操作与 MXU 的矩阵乘法运算更好地重叠，帮助芯片保持繁忙状态，而不是等待顺序向量任务。
- **原生 FP4**：引入原生 4 位浮点运算（FP4），克服内存带宽瓶颈，在保持大型模型精度（即使在低精度量化下）的同时，将 MXU 吞吐量提升一倍。通过减少每个参数的

位数，该平台最大限度地减少了耗能的数据传输，并允许更大的模型层适应本地硬件缓冲区，从而实现最佳计算利用率。

- **ICI：**与上一代产品相比，TPU 8t 的芯片间互连（ICI）带宽提升了 2 倍，原始 DCN 横向扩展带宽提升了高达 4 倍，从而显著降低了数据瓶颈。

Virgo Network 拓扑结构，实现 4 倍的 DC 网络带宽提升：这种全新的网络架构使 TPU 8t 训练的数据中心网络（DCN）带宽提升高达 4 倍。Virgo Network 是一种横向扩展架构，专为满足现代 AI 工作负载的极端需求而设计。它基于高基数交换机构建，通过增加每个交换机的端口数量来减少网络层数，并采用扁平化的两层无阻塞拓扑结构。与传统数据中心网络相比，这种架构通过最大限度地减少网络层数，显著降低了延迟。它采用多平面设计，并具有独立的控制域来连接 TPU 8t 芯片。TPU 8t 机架还连接到 Jupiter 南北向网络架构，以访问计算和存储服务。这种精简的架构共同提供了海量的二分带宽和确定性的低延迟，从而能够支持全球最大规模的高可用性训练集群。

图表5：TPU 8t 机架级与 Virgo 光纤通道的连接



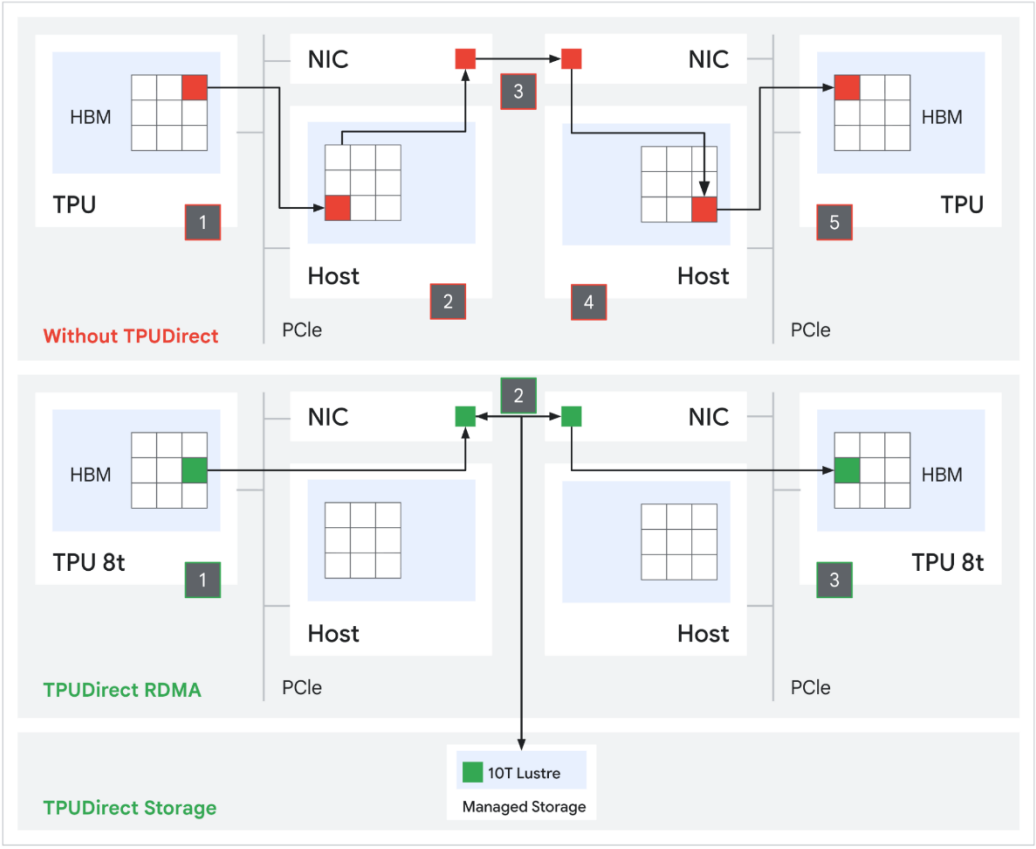
资料来源：谷歌，国盛证券研究所

更快的存储访问：在 TPU 8t 中引入 TPUDirect RDMA 和 TPUDirect Storage。

1）TPU Direct RDMA 支持 TPU 内存（HBM）与网络接口卡（NIC）之间的直接数据传输，绕过主机 CPU 和 DRAM。这降低了延迟和主机系统瓶颈，提高了 TPU 间通信的有效带宽。

2）TPUDirect Storage 通过实现 TPU 与高速托管存储（例如 10T Lustre）之间的直接内存访问，绕过了 CPU 主机瓶颈，有效地将海量数据传输的带宽提高了一倍。这种架构使芯片能够以线速摄取训练数据，确保即使在处理大型多模态数据集时，MXU 也能保持满负荷运行。

图表6：TPU 8t 中引入 TPUDirect RDMA 和 TPU Direct Storage



资料来源：谷歌，国盛证券研究所

二、TPU 8i：针对训练后处理和高并发推理进行了优化

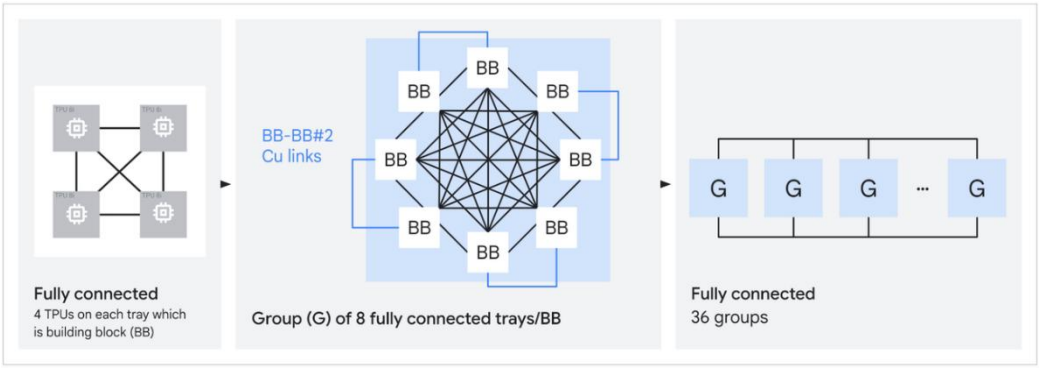
搭载了最大的 SRAM、新的集体加速引擎（CAE）以及新型服务优化网络拓扑 Boardfly。

- **大容量片上 SRAM：**与上一代相比，TPU 8i 的片上 SRAM 容量增加了 3 倍，因此可以完全在硅片上容纳更大的 KV 缓存，从而显著减少长上下文解码期间内核的空闲时间。
- **集体加速引擎（CAE）：**为了解决采样瓶颈，TPU 8i 采用 CAE，它能够以近乎零延迟聚合跨核心的结果。每个 TPU 8i 片都包含两个位于核心芯片上的张量核心（TC）和一个位于芯片组芯片上的 CAE，取代了上一代 Ironwood TPU 中位于核心芯片上的四个稀疏核心（SC）。通过集成专用 CAE，TPU 8i 将集体操作的片上延迟进一步降低了 5 倍。每次集体操作的延迟越低，等待时间就越短，从而直接提高了运行数百万个代理所需的吞吐量。

Boardfly ICI 拓扑结构：虽然 3D 环面结构允许连接数千个芯片以协同工作，但大型网状结构会增加芯片间的跳数，从而导致更高的全连接延迟。TPU 8i 中改变了芯片的连接方式，采用全连接板，然后将这些板聚合成组。利用高基数设计，最多可以连接 1152 个芯片，从而减小网络直径，并减少数据包穿越系统所需的跳数。Boardfly 在通信密集型工作负载下实现了高达 50%的延迟降低。

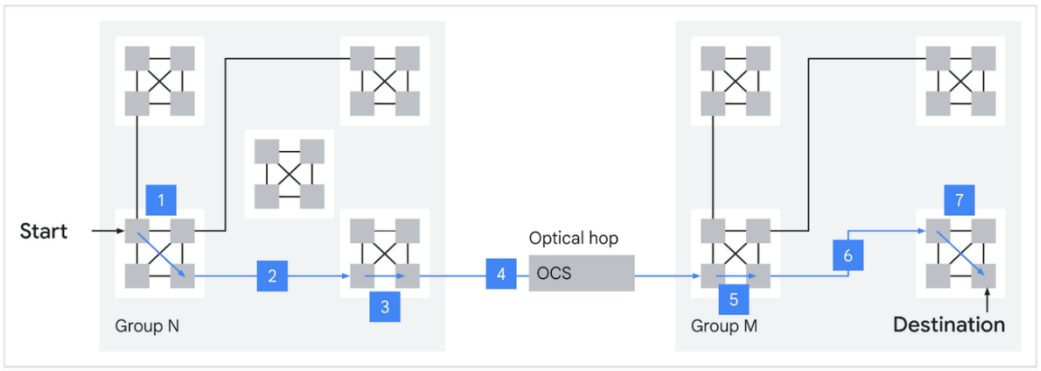
- **构建模块（BB）：**每个托盘使用内部 ICI 连接形成一个四芯片环，提供 16 个外部连接，以实现更广泛的联网。
- **组（G）：**8 个 BB 构成一个 Group，通过铜缆完全连接，利用 11 个可用的外部链路进行组内通信。
- **Pod 结构：**最终架构可扩展至 36 个组（最多支持 1024 个活跃芯片），通过光路交换机（OCS）连接，确保任何芯片间通信的最大延迟为 7 跳。

图表7：TPU 8i 分层式 Boardfly 拓扑结构



资料来源：谷歌，国盛证券研究所

图表8：Pod 架构与 OCS 光交换



资料来源：谷歌，国盛证券研究所

基于 2nm 制程，2027 年底量产。两款第八代 TPU 芯片均搭载了谷歌自研的 Arm 架构 Axion CPU 作为主控，解决数据预处理延迟导致的主机算力瓶颈。芯片采用台积电 2nm 制程工艺制造，目标在 2027 年底量产，并由公司第四代液冷技术支持散热。在软件生态方面，第八代 TPU 支持 JAX、PyTorch、Keras 及 vLLM 等主流框架，原生 PyTorch 支持现已进入预览阶段，用户可直接迁移模型而无需修改代码。

二、DeepSeek V4 双版本上线，打破最强闭源垄断

DeepSeek-V4 拥有百万字超长上下文，在 Agent 能力、世界知识和推理性能上均实现国内与开源领域的领先。2026 年 4 月 24 日，DeepSeek-V4 的预览版本正式上线并同步开源。模型按大小分为两个版本：DeepSeek-V4-Pro（1.6T 参数，49B 激活）和 DeepSeek-V4-Flash（284B 参数，13B 激活），两者均支持一百万 Tokens 的上下文长度，旨在提升超长上下文场景下的性能。

图表9：DeepSeek-V4 拥有百万字超长上下文

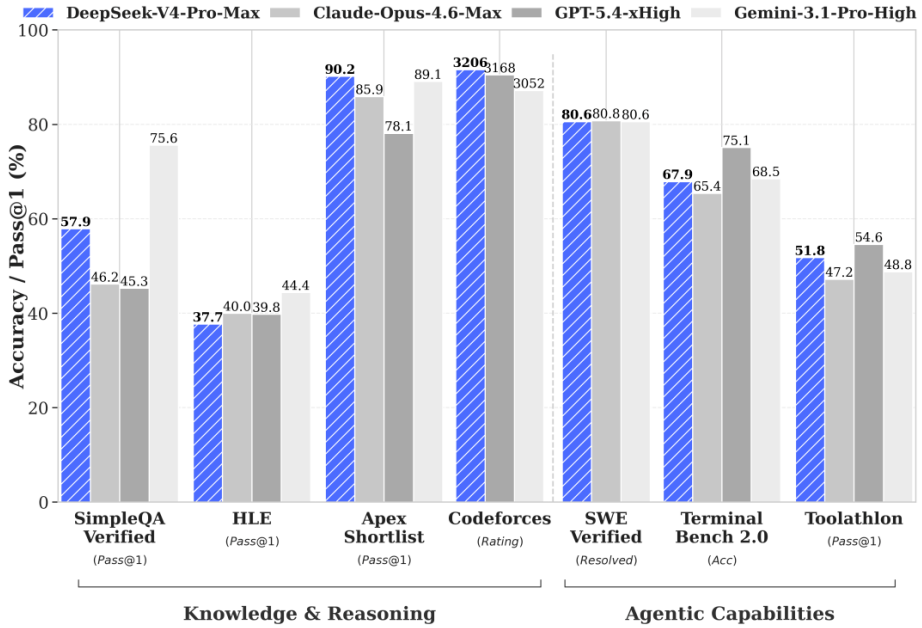
模型	参数	激活	预训练数据	上下文长度	开源	API 服务	网页端/APP 访问方式
deepseek-v4-pro	1.6T	49B	33T	1M	✓	✓	专家模式
deepseek-v4-flash	284B	13B	32T	1M	✓	✓	快速模式

资料来源：DeepSeek，国盛证券研究所

DeepSeek-V4-Pro：性能比肩顶级闭源模型

- **Agent 能力大幅提高：**在 Agentic Coding 评测中，V4-Pro 已达到当前开源模型最佳水平，并在其他 Agent 相关评测中同样表现优异。目前 DeepSeek-V4 已成为公司内部员工使用的 Agentic Coding 模型，据评测反馈使用体验优于 Sonnet 4.5，交付质量接近 Opus 4.6 非思考模式，但仍与 Opus 4.6 思考模式存在一定差距。
- **丰富的世界知识：**DeepSeek-V4-Pro 在世界知识测评中，大幅领先其他开源模型，仅稍逊于顶尖闭源模型 Gemini-Pro-3.1。
- **世界顶级推理性能：**在数学、STEM、竞赛型代码的测评中，DeepSeek-V4-Pro 超越当前所有已公开评测的开源模型，取得了比肩世界顶级闭源模型的优异成绩。

图表10：DeepSeek-V4-Pro 性能比肩顶级闭源模型



资料来源：DeepSeek，国盛证券研究所

DeepSeek-V4-Flash：更快捷高效的经济之选

- 相比 DeepSeek-V4-Pro，DeepSeek-V4-Flash 在世界知识储备方面稍逊一筹，但展现出了接近的推理能力。而由于模型参数和激活更小，相较之下 V4-Flash 能够提供更加快捷、经济的 API 服务。
- 在 Agent 测评中，DeepSeek-V4-Flash 在简单任务上与 DeepSeek-V4-Pro 旗鼓相当，但在高难度任务上仍有差距。

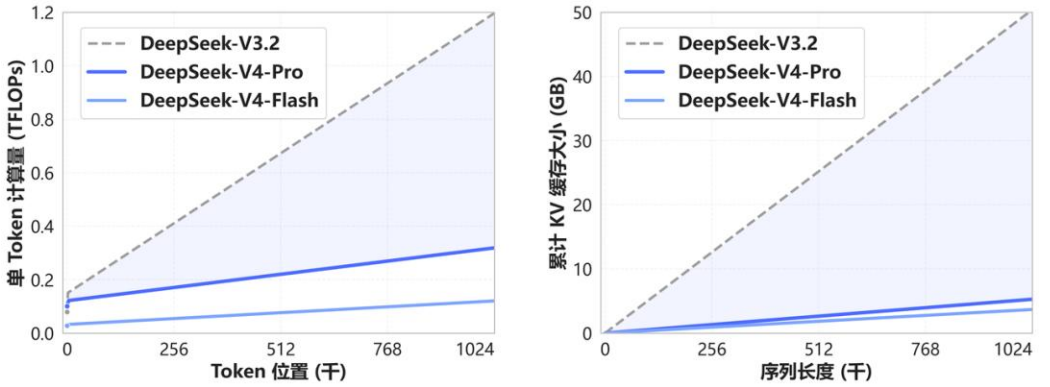
图表11：DeepSeek-V4 测试表现

Benchmark (metric) Reasoning Effort		DS-V4-Pro Max	DS-V4-Flash Max	K2.6 Thinking	GLM-5.1 Thinking	Opus-4.6 Max	GPT-5.4 xHigh	Gemini-3.1-Pro High
Knowledge & Reasoning	MMLU-Pro (EM)	87.5	86.2	87.1	86.0	89.1	87.5	91.0
	SimpleQA-Verified (Pass@1)	57.9	34.1	36.9	38.1	46.2	45.3	75.6
	Chinese-SimpleQA (Pass@1)	84.4	78.9	75.9	75.0	76.2	76.8	85.9
	GPQA Diamond (Pass@1)	90.1	88.1	90.5	86.2	91.3	93.0	94.3
	HLE (Pass@1)	37.7	34.8	36.4	34.7	40.0	39.8	44.4
	LiveCodeBench (Pass@1)	93.5	91.6	89.6	-	88.8	-	91.7
	Codeforces (Rating)	3206	3052	-	-	-	3168	3052
	HMMT 2026 Feb (Pass@1)	95.2	94.8	92.7	89.4	96.2	97.7	94.7
	IMOAnswerBench (Pass@1)	89.8	88.4	86.0	83.8	75.3	91.4	81.0
	Apex (Pass@1)	38.3	33.0	24.0	11.5	34.5	54.1	60.9
	Apex Shortlist (Pass@1)	90.2	85.7	75.5	72.4	85.9	78.1	89.1
Long Context	MRCR 1M (MMR)	83.5	78.7	-	-	92.9	-	76.3
	CorpusQA 1M (ACC)	62.0	60.5	-	-	71.7	-	53.8
Agentic	Terminal Bench 2.0 (Acc)	67.9	56.9	66.7	63.5	65.4	75.1	68.5
	SWE Verified (Resolved)	80.6	79.0	80.2	-	80.8	-	80.6
	SWE Pro (Resolved)	55.4	52.6	58.6	58.4	57.3	57.7	54.2
	SWE Multilingual (Resolved)	76.2	73.3	76.7	73.3	77.5	-	-
	BrowseComp (Pass@1)	83.4	73.2	83.2	79.3	83.7	82.7	85.9
	HLE w/tools (Pass@1)	48.2	45.1	54.0	50.4	53.1	52.0	51.6
	GDPval-AA(Elo)	1554	1395	1482	1535	1619	1674	1314
	MCPAtlas Public(Pass@1)	73.6	69.0	66.6	71.8	73.8	67.2	69.2
	Toolathlon (Pass@1)	51.8	47.8	50.0	40.7	47.2	54.6	48.8

资料来源：DeepSeek，国盛证券研究所

百万上下文实现标配。一年前，1M 上下文为 Gemini 独家的王牌；其他所有闭源模型一般为 128K 或 200K 上下文；开源模型没有做到这个量级的。DeepSeek-V4 开创了一种全新的注意力机制，在 token 维度进行压缩，结合 DSA 稀疏注意力（DeepSeek Sparse Attention），实现了全球领先的长上下文能力，并且相比于传统方法大幅降低了对计算和显存的需求。从现在开始，1M 上下文将是 DeepSeek 所有官方服务的标配。

图表12：DeepSeek-V4 和 DeepSeek-V3.2 的计算量和显存容量随上下文长度的变化



资料来源：DeepSeek，国盛证券研究所

Agent 能力专项优化。DeepSeek-V4 针对 Claude Code、OpenClaw、OpenCode、CodeBuddy 等主流的 Agent 产品进行了适配和优化，在代码任务、文档生成任务等方面表现均有提升。API 这边，V4-Pro 和 V4-Flash 同步上线，支持 OpenAI ChatCompletions 接口和 Anthropic 接口两套，受限于高端算力，目前 DeepSeek-V4-Pro 的服务吞吐十分有限，预计下半年昇腾 950 超节点批量上市后，其价格会大幅下调。

昇腾 950 超节点支撑 DeepSeek V4 毫秒级推理。昇腾 950 超节点实现 DeepSeek V4-Pro 20ms 和 DeepSeek V4-Flash 10ms 低时延推理。这得益于昇腾 950 代际底层架构的三大升级：

1）原生精度加速，其全面支持 FP8、MXFP8、MXFP4 等数据格式，在保证模型精度的同时，可实现内存占用降低 50%+，计算能力翻倍。

2）稀疏访存优化，针对 MoE 模型的离散访存特征，他们通过**大幅提升硬件级稀疏访存能力**，解决了专家路由过程中的带宽瓶颈。

3）Vector 与 Cube 共享 Memory，其采用创新存储架构设计，实现了向量单元（Vector）与矩阵单元（Cube）的 Memory 共享，消除大量片上数据搬运开销，降低了端到端推理时延。

根据华为官方信息，昇腾 950 超节点还从基础器件、协议算法到光电互联，实现了系统级突破，支持用户以 64 卡为步长按需扩展，可实现 8192 卡无收敛全互联，提供业界最大 Scale Up 能力。

多款国产高端芯片率先完成适配与深度兼容。智源研究院众智 FlagOS 社区宣布将对 DeepSeek-V4 模型进行全量适配，目前其已完成 DeepSeek-V4-Flash 在 8 款以上 AI 芯片上的全量适配与推理部署，包括**海光、沐曦、华为昇腾、摩尔线程（FP8）、昆仑芯、平头哥真武、天数、英伟达（FP8）**等芯片，正在推进 DeepSeek-V4-Pro 模型在多个芯片的迁移适配。

三、英特尔 26Q1 业绩强劲，CPU 需求大幅提升

英特尔发布 2026 年第一季度财报，数据中心业务增长迅猛，人工智能驱动需求。26Q1，英特尔实现营收 136 亿美元（超出指引上限 127 亿美元），同比增长 7%；GAAP 毛利率 39.4%（大幅超出指引 32.3%），同比提升 2.5 个百分点。分业务来看，数据中心和人工智能产品实现营收 51 亿美元，同比增长 22%，成为所有业务板块中增长最快的板块；客户端计算组营收 77 亿美元，同比仅增长 1%；代工业务实现营收 54 亿美元，同比增长 16%。新一轮的人工智能浪潮将推动智能更贴近终端用户，其演进路径从基础模型走向推理，再迈向智能体阶段。这一转变正显著增加市场对英特尔 CPU、晶圆及先进封装产品的需求。

图表13：英特尔 26Q1 业绩概览

	通用会计准则			非通用会计准则		
	2026 第一季度	2025 第一季度	相比2025 第一季度	2026 第一季度	2025 第一季度	相比2025 第一季度
营收	$136亿	$127亿	上升7%			
毛利率	39.4%	36.9%	上升2.5个百分点	41.0%	39.2%	上升1.8个百分点
研发和营销管理费用	$44亿	$48亿	下降8%	$39亿	$43亿	下降9%
营业利润率（亏损）	(23.1)%	(2.4)%	下降20.7个百分点	12.3%	5.4%	上升6.9个百分点
税率	(8.5)%	(51.4)%	上升42.9个百分点	11.0%	12.0%	下降1个百分点
英特尔净收入（亏损）	$(37)亿	$(8)亿	*无比较意义	$15亿	$6亿	上升156%
英特尔稀释后每股收益（亏损）	$(0.73)	$(0.19)	*无比较意义	$0.29	$0.13	上升123%

资料来源：英特尔中国，国盛证券研究所

图表14：英特尔 26Q1 分业务业绩

业务部门收入及趋势	2026 第一季度[1]	相比2025 第一季度
英特尔产品：		
客户端计算事业部（CCG）	$77亿	上升1%
数据中心和人工智能事业部（DCAI）	$51亿	上升22%
英特尔产品总收入	$128亿	上升9%
英特尔代工	$54亿	上升16%
其他所有	$6亿	下降33%
部门间抵消	$(53)亿	
总净收入	$136亿	上升7%

资料来源：英特尔中国，国盛证券研究所

持续扩展客户端产品组合，推出多款针对性产品。英特尔扩展了客户端产品组合（面向工作站的英特尔至强 600 系列处理器、面向桌面和移动设备的英特尔酷睿超极本 200S Plus 和 200HX Plus 处理器、面向医疗健康和生命科学边缘计算的英特尔酷睿 2 系列处

理器、基于第三代英特尔酷睿 Ultra 系列处理器的英特尔 vPro 平台）；推出了基于 Intel 18A 制程的第三代英特尔酷睿系列处理器，首次将 Intel 18A 制程、最新 IP、现代特性及全天电池续航引入主流市场。

深化了与核心合作伙伴的合作，巩固 AI 领域地位。1）与 Google 宣布达成为期多年的合作，将在 Google 针对工作负载优化的实例中持续部署英特尔® 至强® 处理器，包括采用最新英特尔® 至强® 6 处理器的 C4 和 N4 实例。该合作还包括，双方将共同开发定制 ASIC 基础设施处理器（IPU），从而提高利用率、降低复杂性并更高效地扩展 AI 工作负载。2）英特尔至强 6 处理器将作为主控 CPU，应用于 NVIDIA DGX Rubin NVL8 系统，这巩固了英特尔在领先 AI 基础设施部署中的核心地位。3）与 SambaNova 宣布了异构硬件解决方案蓝图，以应对企业和云服务提供商面临的性能、效率和软件兼容性挑战。该设计将结合 GPU、SambaNova RDU 以及作为主机和任务 CPU 的英特尔® 至强® 6 处理器。4）作为战略合作伙伴参与 Terafab 项目，与 SpaceX、xAI 和 Tesla 携手。英特尔大规模设计、制造及封装超高性能芯片的能力，将有助于重构硅晶圆制造技术。

面向全球对封装解决方案日益增长的需求，英特尔代工扩大了马来西亚槟城工厂的封装测试产能，为客户产品提供支持，同时增强了全球半导体供应链的韧性。回购了与爱尔兰 Fab 34 晶圆厂相关的合资企业中 49%的少数股权权益。该协议反映了英特尔持续的业务发展势头，这得益于 CPU 在 AI 时代日益重要且不可或缺的作用，以及英特尔显著增强的资产负债表。

发布 2026 年第二季度业绩指引。26Q2，营收指引为 138 亿美元至 148 亿美元，中值为 143 亿美元，同比增长 11%；GAAP 毛利率指引为 37.5%，同比增长 10 个百分点。

图表15：英特尔 26Q2 业绩指引

2026 第二季度	通用会计准则	非通用会计准则
营收	$138 - 148亿	
毛利率	37.5%	39%
税率	4%	11%
英特尔稀释后每股收益	$0.08	$0.20

资料来源：英特尔中国，国盛证券研究所

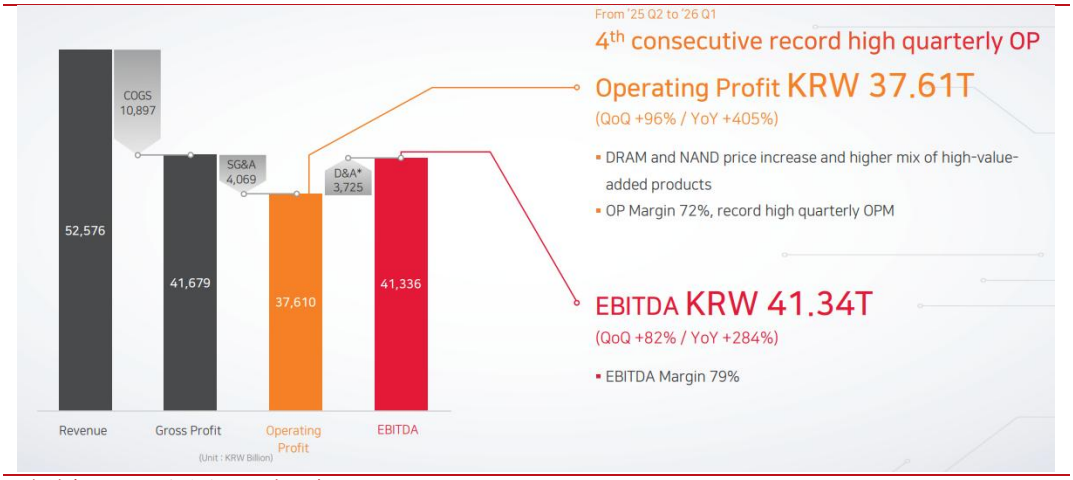
四、SK 海力士 26Q1 创新高，重视存储板块业绩释放机遇

SK 海力士 26Q1 业绩：AI 需求持续强劲，高附加值产品销售扩大，创下单季度历史最高业绩。26Q1，公司营收达 52.58 万亿韩元，环比增长 60%，同比增幅高达 198%；营业利润为 37.61 万亿韩元，环比增长 96%，同比增 405%，营业利润率 72%，显著展现了盈利能力的强劲改善趋势。净利润为 40.35 万亿韩元，净利润率 77%。在 AI 基础设施投资扩大的推动下，市场需求持续强劲。公司通过扩大 HBM、大容量服务器 DRAM 模块、企业级固态硬盘（eSSD）等高附加值产品的销售，延续了业绩上升势头。

图表16：SK 海力士 26Q1 收入情况

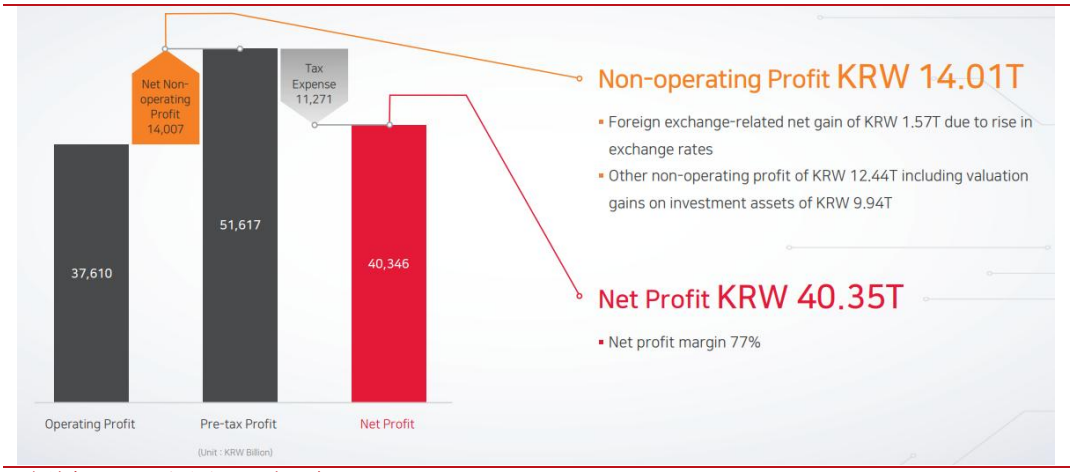
资料来源：SK 海力士，国盛证券研究所

图表17：SK 海力士 26Q1 营业利润情况



资料来源：SK 海力士，国盛证券研究所

图表18：SK 海力士 26Q1 净利润情况



资料来源：SK 海力士，国盛证券研究所

1）在 HBM 方面，公司将进一步加强整合性能、良率、质量及供应稳定性的综合执行能力。**2）在 DRAM 方面，**公司将全面启动全球首款采用第六代 10 纳米级（1c）工艺的 LPDDR6 DRAM，以及基于同一工艺并于本月开始量产的 192GB SOCAMM2 的供应。**3）在 NAND 闪存方面，**公司开始供应基于 CTF*且采用 321 层 QLC**技术的客户端固态硬盘（cSSD）“PQC21”的同时，将通过在 eSSD 的全领域构建涵盖高性能 TLC 和大容量 QLC 的产品阵容，灵活应对人工智能领域的整体需求。此外，公司将依托在大容量 QLC eSSD 领域拥有优势的 Solidigm 产生的协同效应，进一步加强在 AI 数据中心和 AI PC 存储器市场的竞争力。**展望第二季度，预计 DRAM 出货量环比增长高个位数百分比，NAND 出货量（含 Solidigm）环比增长百分之十几（mid teen）。**

五、相关标的

【光产业链】
光模块&光芯片：东山精密（索尔思）、源杰科技、中际旭创、新易盛、天孚通信、长光华芯、永鼎股份、仕佳光子等。
电芯片：优迅股份。
OCS：炬光科技、豪威集团、腾景科技、光库科技、赛微电子、光迅科技等。
光器件：太辰光、长飞光纤等。

【国产算力产业链】
代工：中芯国际、华虹半导体。
芯片：芯原股份、沐曦股份、海光信息、寒武纪、天数智芯、壁仞科技、摩尔线程、灿芯股份、翱捷科技等。
昇腾产业链：杰华特、华丰科技、深南电路、兴森科技等。
超节点：盛科通信、华勤技术、澜起科技等。
封测：盛合晶微、长电科技、通富微电等。
国产存储：香农芯创、兆易创新、国科微、佰维存储、江波龙、澜起科技、聚辰股份等。

风险提示

1. 下游需求不及预期。
2. 研发进展不及预期。
3. 地缘政治风险。

免责声明

国盛证券股份有限公司（以下简称“本公司”）具有中国证监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。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。

本报告的信息均来源于本公司认为可信的公开资料，但本公司及其研究人员对该等信息的准确性及完整性不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，可能会随时调整。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息及资料保持在最新状态，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。

本公司力求报告内容客观、公正，但本报告所载的资料、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。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也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。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的任何观点均精准地反映了我们对标的证券和发行人的个人看法，结论不受任何第三方的授意或影响。我们所得报酬的任何部分无论是在过去、现在及将来均不会与本报告中的具体投资建议或观点有直接或间接联系。

投资评级说明

投资建议的评级标准		评级	说明
评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准。	股票评级	买入	相对同期基准指数涨幅在 15%以上
		增持	相对同期基准指数涨幅在 5%~15%之间
		持有	相对同期基准指数涨幅在-5%~+5%之间
		减持	相对同期基准指数跌幅在 5%以上
	行业评级	增持	相对同期基准指数涨幅在 10%以上
		中性	相对同期基准指数涨幅在-10%~+10%之间
		减持	相对同期基准指数跌幅在 10%以上

国盛证券研究所

北京
地址：北京市东城区永定门西滨河路 8 号院 7 楼中海地产广场东塔 7 层
邮编：100077
邮箱：gsresearch@gszq.com

上海
地址：上海市浦东新区南洋泾路 555 号陆家嘴金融街区 22 栋
邮编：200120
电话：021-38124100
邮箱：gsresearch@gszq.com

南昌
地址：南昌市红谷滩新区凤凰中大道 1115 号北京银行大厦
邮编：330038
传真：0791-86281485
邮箱：gsresearch@gszq.com

深圳
地址：深圳市福田区福华三路 100 号鼎和大厦 24 楼
邮编：518033
邮箱：gsresearch@gszq.com