

AI 算力需求强劲，英伟达 Q1 业绩大超预期

计算机行业

事件概述：

2023年5月25日，英伟达召开2024财年一季度业绩会，一季度营收71.9亿美元，环比增长19%，同比下降13%，降幅几乎为市场预期和四季度降幅的一半。公司AI芯片所在数据中心业务收入达42.8亿美元（环比+18%，同比+14%），创历史新高；游戏及可视化业务均实现环比增长。

二季度营收预期约110亿美元，同比增近33%；毛利率超64%。业绩会上英伟达称，众多云公司竞相部署AI芯片，其锁定了数据中心芯片的大幅增长，计划下半年大幅增加供应。

DGX“共享”AI能力，底层算力龙头平台化布局

在此前的GTC大会上，英伟达重磅推出AI服务平台，将AI能力提供给所有客户。平台为模型和算力之间的“桥梁”，是AIGC或大模型生成的必备要素，各项应用和底层技术有赖于平台实现整合调配。我们认为大模型带来的GPU需求目前仍保持强劲，从中短期来看，英伟达的业绩将持续受益。而从未来长周期来看，借助加速计算硬件及成熟架构、基础软件和垂直软件等领域的整合，公司有望构建竞争壁垒，打造AI生态，实现从GPU供应商到平台型综合AI解决方案领导者的转变。

大模型加速开启算力“军备赛”，硬件市场持续升温

万亿美元数据中心市场加速推进，AI算力需求已超越摩尔定律速度增长。各地都在积极投建数据中心，加速计算和生成式AI的发展势不可挡。算力是训练大模型的底层动力源泉，在ChatGPT引发AI浪潮后，国内大模型自研AI大模型进入“百花齐放”阶段，大模型的涌现开启了算力“军备赛”。同时，大模型参数呈现指数规模，进一步引爆海量算力需求。AI硬件市场也持续升温，此前英伟达AI旗舰芯片H100售价在多个商铺炒至4万美元，也进一步印证了AI算力产业链是高确定性和高景气度的方向。

投资建议：

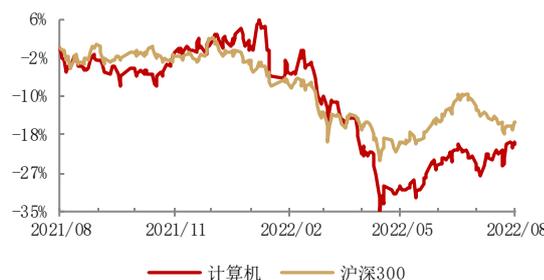
我们认为在加速计算和生成式AI如火如荼推进的背景下，重点推荐具备算力基础的厂商，受益标的为：

- 1、算力芯片厂商：**寒武纪、海光信息、景嘉微、龙芯中科**等；
- 2、算力服务厂商：**浪潮信息、中科曙光、神州数码、拓维信息**等；

评级及分析师信息

行业评级：推荐

行业走势图



分析师：刘泽晶

邮箱：liuzj1@hx168.com.cn

SAC NO: S1120520020002

3、边缘算力厂商：首都在线、青云科技、优刻得、光环新网、新炬网络、网宿科技等。

风险提示

核心技术水平升级不及预期的风险，AI 伦理风险，政策推进不及预期的风险，中美贸易摩擦升级的风险。

正文目录

| | |
|--------------------------------------|---|
| 1. 英伟达：AI 计算领导者，锚定算力主线..... | 4 |
| 1.1. DGX “共享” AI 能力，底层算力龙头平台化布局..... | 4 |
| 1.2. 大模型加速开启算力“军备赛”，硬件市场持续升温..... | 6 |
| 2. 投资建议：梳理 AI 算力相关受益厂商..... | 9 |
| 3. 风险提示..... | 9 |

图目录

| | |
|-------------------------------------|---|
| 图表 1 发布一季报，英伟达盘后股价大涨..... | 4 |
| 图表 2 英伟达数据中心业务收入（亿美元）..... | 5 |
| 图表 3 英伟达 DRX 云示意图..... | 5 |
| 图表 4 英伟达 AI 加速计算上云架构示意图..... | 5 |
| 图表 5 英伟达 AI Foundation 能力示意图..... | 6 |
| 图表 6 英伟达 AI Foundation 合作伙伴示意图..... | 6 |
| 图表 7 平台提供整合作用..... | 6 |
| 图表 8 全球算力规模及增速..... | 7 |
| 图表 9 大模型爆发带来更大算力需求..... | 7 |
| 图表 10 大模型参数数量和训练数据规模快速增长..... | 7 |
| 图表 11 近年大模型的参数规模增长趋势..... | 7 |
| 图表 12 国内大模型梳理..... | 8 |
| 图表 13 超级计算机架构..... | 9 |
| 图表 14 英伟达数据中心 GPU 类别..... | 9 |

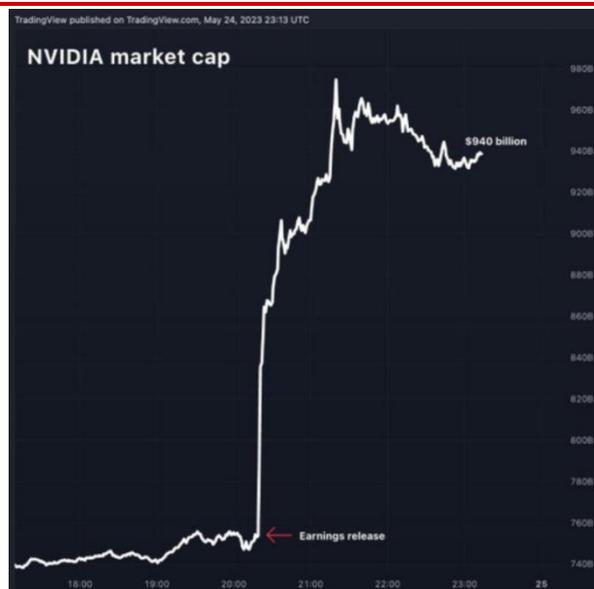
1. 英伟达：AI 计算领导者，锚定算力主线

1.1. DGX “共享” AI 能力，底层算力龙头平台化布局

2023 年 5 月 25 日，AI 底层算力龙头、芯片巨头英伟达召开 2024 财年一季度业绩会，公司过去一季度得到 AI 芯片相关业务支持，实现营收 71.9 亿美元（高于公司指引区间 63.7 亿-66.3 亿美元），环比增长 19%，同比下降 13%，降幅几乎为市场预期和四季度降幅的一半。公司二季度营收指引约 110 亿美元，同比增近 33%，给出了较华尔街预期高出 50% 的预测。

财报后及业绩会期间，英伟达股价迅速拉升，一度涨至 390 美元上方的历史最高水平，盘后涨幅接近 30%。市值增加 2000 亿美元至 9600 亿美元，距离万亿市值仅一步之遥。

图表 1 发布一季报，英伟达盘后股价大涨



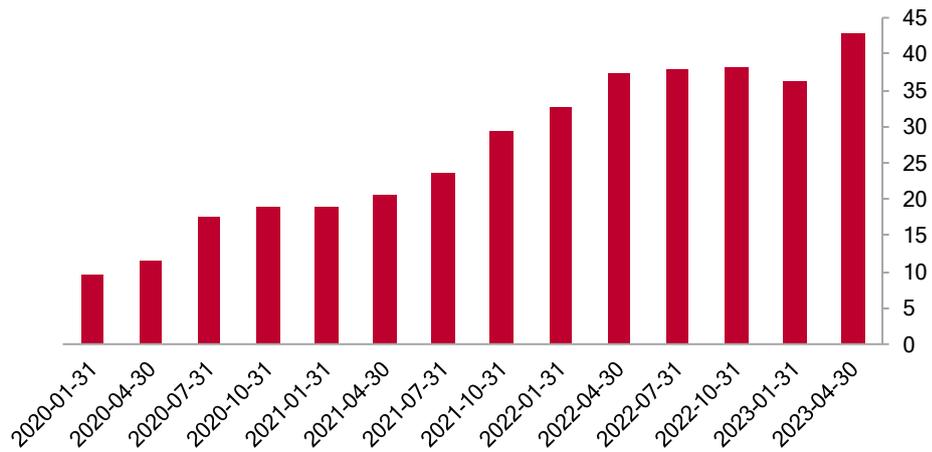
资料来源：华尔街见闻，华西证券研究所

英伟达核心业务数据中心收入创历史新高。一季度数据中心营收为 42.8 亿美元，同比增长 14%，环比增长 18%，较分析师预期的 39.1 亿美元高约 9.5%。英伟达称，数据中心收入激增主要是由于对使用基于 Hopper 和 Ampere 架构 GPU 的生成式 AI 和大语言模型的需求不断增长。

一季度游戏业务营收 22.4 亿美元，同比下降 38%，环比增长 22%，较分析师预期收入 19.8 亿美元高 13.1%，其四季度同比下降 46%，降速有所放缓。游戏业务下滑主要由于宏观经济增长放缓造成的需求疲软，以及渠道库存的正常化导致出货下降。

一季度的专业可视化业务营收 2.95 亿美元，同比下降 53%，环比增长 31%，同样高于分析师预期的 2.47 亿美元。英伟达认为可视化业务的同比下滑反映出渠道去库存影响，环比增长受到 PC 端和移动端 GPU 的需求增长推动。

图表 2 英伟达数据中心业务收入（亿美元）



资料来源：iFind，华西证券研究所

GPU 供应商到 AI 平台，定义人工智能发展方向。在此前的 GTC 大会上，英伟达在基础软件方面推出全新加速库；在硬件架构方面，推出数据中心 CPU Grace；在服务器方面，推出 DGX(超级计算机)；推出 DGX 云与 AI foundations 云服务。其中，DGX 云在一季度业绩会中被反复提及，英伟达认为通过 DGX 云可以将他们的 AI 能力提供给所有客户，帮助企业部署自己的模型。

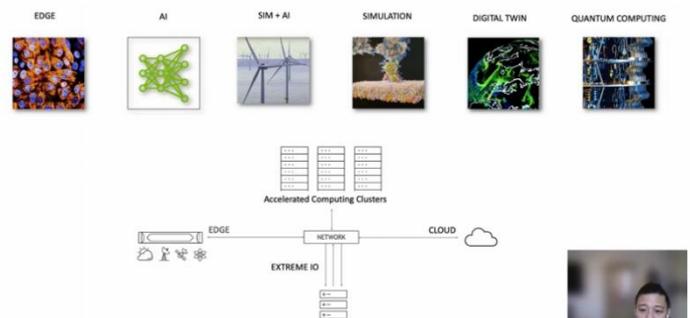
我们认为大模型带来的 GPU 需求目前仍保持强劲，从中短期来看，英伟达的业绩将持续受益。而从未来长周期来看，借助加速计算硬件及成熟架构、基础软件和垂直软件等领域的整合，公司有望构建竞争壁垒，打造 AI 生态。

图表 3 英伟达 DRX 云示意图



资料来源：英伟达官网，华西证券研究所

图表 4 英伟达 AI 加速计算上云架构示意图



资料来源：英伟达官网，华西证券研究所

平台为模型和算力之间的“桥梁”，是 AIGC 或大模型生成的必备要素。正如我们在《ChatGPT，百度文心一言畅想》中所论证，不论是数据库还是编译器，

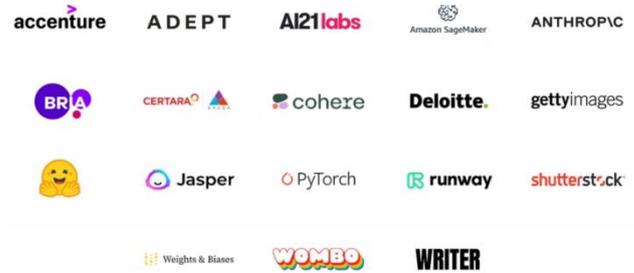
都需要通过平台来实现资源的合理分配以达到软硬件的最优组合，从而大幅提升模型效率。平台通过调用数据包来适配软硬件之间的结构，来达到模型的最优组合，从而提升模型乃至整个虚拟机的效率。英伟达的布局方向正是旨在成为平台型综合 AI 解决方案领导者。

图表 5 英伟达 AI Foundation 能力示意图



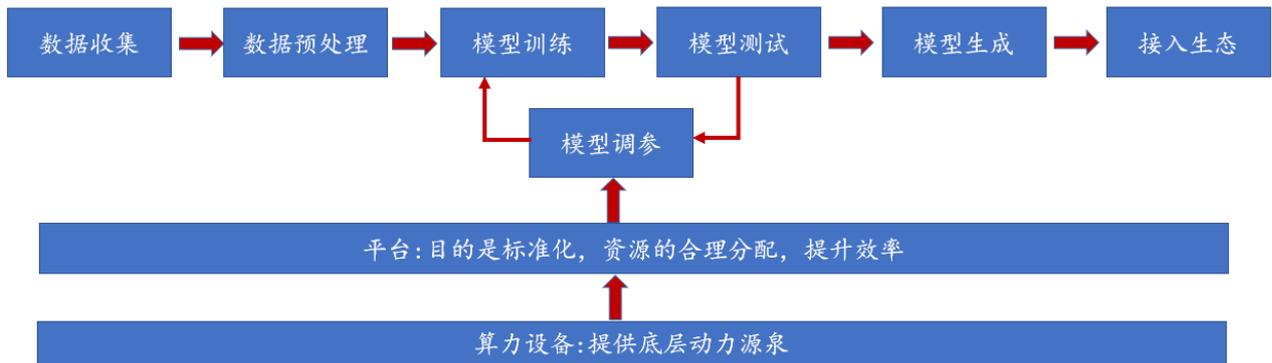
资料来源：英伟达官网，华西证券研究所

图表 6 英伟达 AI Foundation 合作伙伴示意图



资料来源：英伟达官网，华西证券研究所

图表 7 平台提供整合作用

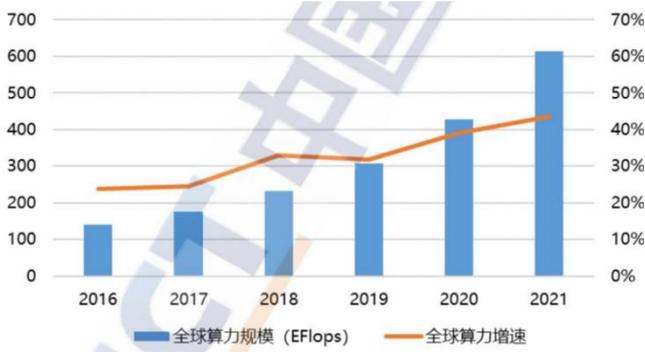


资料来源：IDC，华西证券研究所

1.2. 大模型加速开启算力“军备赛”，硬件市场持续升温

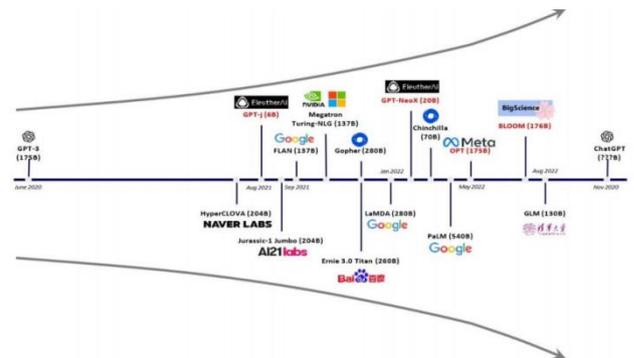
万亿美元数据中心市场加速推进，AI 算力需求超越摩尔定律速度增长。根据英伟达业绩会，数据中心规模已超万亿美元，各地都在积极投建数据中心，加速计算和生成式 AI 的发展势不可挡。由于大模型需要 GPU 进行训练，AI 算力将维持高确定性和高景气度方向。

图表 8 全球算力规模及增速



资料来源：中国信通院，华西证券研究所

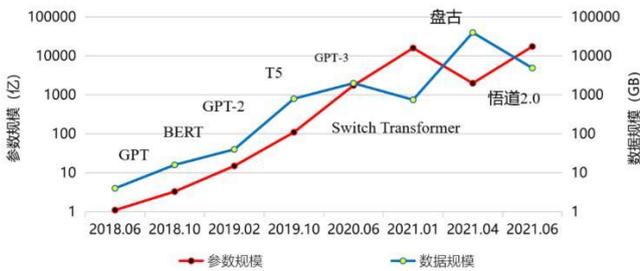
图表 9 大模型爆发带来更大算力需求



资料来源：陈巍谈芯，华西证券研究所

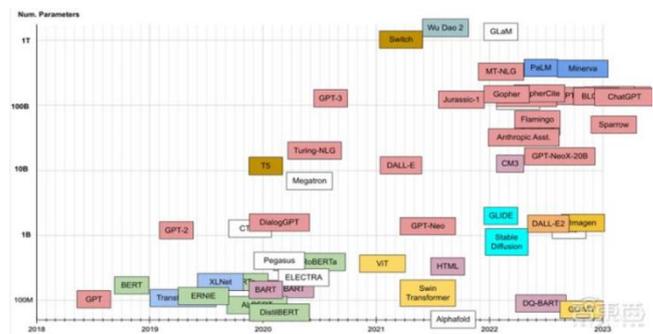
在 ChatGPT 引发 AI 浪潮后，国内大模型也开始于 3、4 月份密集发布，自研 AI 大模型进入“百花齐放”阶段，而大模型的涌现开启了算力“军备赛”。算力是大模型的底层动力源泉，一个优秀的算力底座在大模型(AI 算法)的训练和推理具备效率优势。其中 AI 服务器、AI 芯片、AI 云是核心产品。同时，大模型参数呈现指数规模，进一步引爆海量算力需求。根据 OpenAI 数据，模型计算量增长速度远超人工智能硬件算力增长速度，存在万倍差距，因此带动了对 AI 训练芯片单点算力提升的需求。根据智东西数据，过去五年，大模型发展呈现指数级别，因此对算力需求也随之攀升。

图表 10 大模型参数数量和训练数据规模快速增长



资料来源：智东西，华西证券研究所

图表 11 近年大模型的参数规模增长趋势



资料来源：智东西，华西证券研究所

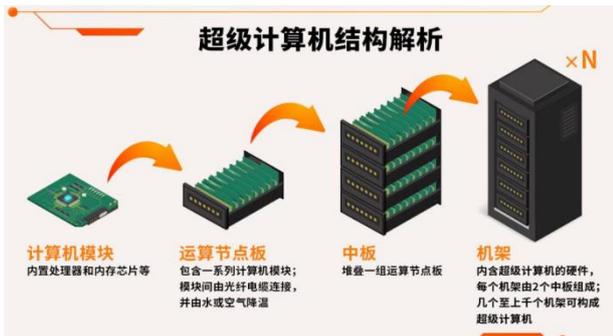
图表 12 国内大模型梳理

| 公司 | 最新发布时间 | 大模型 | 参数规模 | 模型特点 |
|------|-----------|---------------|-------------|-----------------------------|
| 网易 | - | 伏羲 | 110亿 | NLP大模型、多模态大模型 |
| 云从科技 | 研发 | 行业精灵 | 百亿-千亿 | 行业大模型 |
| 京东 | 待发布 | ChatJD | 千亿级 | 新一代产业大模型 |
| 字节跳动 | 预计今年 | 自研大模型 | | 在语言和图像两种模态上发力 |
| 科大讯飞 | 2023.5.6 | 1+N认知智能大模型 | | |
| 知乎 | 2023.4.13 | 知海图AI | 10亿级 | NLP大模型 |
| 阿里巴巴 | 2023.4.11 | 通义千问 | 超10万亿 | 大语言模型 |
| | 2021.3 | M6 | 10万亿 | 多模态大模型 |
| 昆仑万维 | 2023.4.10 | 天工3.5 | 百亿级 | 多模态大模型、代码大模型 |
| 商汤科技 | 2023.4.10 | 日日新 | 1300亿 | 商量：NLP大模型、代码大模型 |
| | | | 超10亿 | AIGC文生图、数字人物生成等 |
| 360 | 2023.4.9 | 360智脑 | | |
| 华为 | 2023.4.8 | 盘古CV 盘古NLP | 超30亿 千亿级 | NLP大模型、CV大模型、科学计算大模型（气象） |
| 百度 | 2023.3.16 | 文心大模型 | 2600亿 | NLP大模型、CV大模型、跨模态大模型、生物计算大模型 |
| 澜舟科技 | 2023.3.14 | 孟子Mchat | 百亿-千亿 | NLP大模型 |
| 复旦大学 | 2023.2.20 | MOSS | 175亿 | NLP大模型 |
| 腾讯 | 2022.4 | 混元AI大模型 | 万亿级 | NLP大模型、CV大模型、多模态大模型 |
| 浪潮信息 | 2021.9 | 源1.0 | 2457亿 | NLP大模型 |

资料来源：公开资料整理, 华西证券研究所

随着算力需求居高不下，AI 硬件市场也持续升温。根据界面新闻，英伟达 AI 旗舰芯片 H100 售价在多个商铺炒至 4 万美元，相比此前零售商报价 3.6 万美元，已明显提价。1 万枚英伟达 A100 芯片是做好 AI 大模型的算力门槛。上游芯片带动了服务器价格同步上行。一台 GPU 服务器的成本是普通服务器的 10 倍以上，GPU 价格高涨直接带动服务器价格显著上修。以国产浪潮 AI 智能服务器为例，根据 AI 市场报价，其 R4900G3 规格产品含税价已高达 55 万元。

图表 13 超级计算机架构



资料来源：CDSN，华西证券研究所

图表 14 英伟达数据中心 GPU 类别

| VideoCardz.com | NVIDIA H100 | NVIDIA A100 | NVIDIA Tesla V100 | NVIDIA Tesla P100 |
|------------------|---------------------|---------------------|---------------------|---------------------|
| Picture | | | | |
| GPU | GH100 | GA100 | GV100 | GP100 |
| Transistors | 80B | 54.2B | 21.1B | 15.3B |
| Die Size | 814 mm ² | 828 mm ² | 815 mm ² | 610 mm ² |
| Architecture | Hopper | Ampere | Volta | Pascal |
| Fabrication Node | TSMC N4 | TSMC N7 | 12nm FFN | 16nm FinFET+ |
| GPU Clusters | 132/114* | 108 | 80 | 56 |
| CUDA Cores | 16896/14592* | 6912 | 5120 | 3584 |
| L2 Cache | 50MB | 40MB | 6MB | 4MB |
| Tensor Cores | 528/456* | 432 | 320 | - |
| Memory Bus | 5120-bit | 5120-bit | 4096-bit | 4096-bit |
| Memory Size | 80 GB HBM3/HBM2e* | 40/80GB HBM2e | 16/32 HBM2 | 16GB HBM2 |
| TDP | 700W/350W* | 250W/300W/400W | 250W/300W/450W | 250W/300W |
| Interface | SXM5*/PCIe Gen5 | SXM4/PCIe Gen4 | SXM2/PCIe Gen3 | SXM/PCIe Gen3 |
| Launch Year | 2022 | 2020 | 2017 | 2016 |

资料来源：embeddedcomputing，华西证券研究所

2. 投资建议：梳理 AI 算力相关受益厂商

我们认为在加速计算和生成式 AI 如火如荼推进的背景下，重点推荐具备算力基础的厂商，受益标的为：

- 1、算力芯片厂商：**寒武纪、海光信息、景嘉微、龙芯中科等；**
- 2、算力服务厂商：**浪潮信息、中科曙光、神州数码、拓维信息等；**
- 3、边缘算力厂商：**首都在线、青云科技、优刻得、光环新网、新炬网络、网宿科技等。**

3. 风险提示

核心技术水平升级不及预期的风险，AI 伦理风险，政策推进不及预期的风险，中美贸易摩擦升级的风险。

分析师与研究助理简介

刘泽晶（首席分析师）：2014-2015年新财富计算机行业团队第三、第五名，水晶球第三名，10年证券从业经验。

分析师承诺

作者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，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

| 公司评级标准 | 投资评级 | 说明 |
|--------------------------------|------|--------------------------------|
| 以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。 | 买入 | 分析师预测在此期间股价相对强于上证指数达到或超过15% |
| | 增持 | 分析师预测在此期间股价相对强于上证指数在5%—15%之间 |
| | 中性 | 分析师预测在此期间股价相对上证指数在-5%—5%之间 |
| | 减持 | 分析师预测在此期间股价相对弱于上证指数5%—15%之间 |
| | 卖出 | 分析师预测在此期间股价相对弱于上证指数达到或超过15% |
| 行业评级标准 | | |
| 以报告发布日后的6个月内行业指数的涨跌幅为基准。 | 推荐 | 分析师预测在此期间行业指数相对强于上证指数达到或超过10% |
| | 中性 | 分析师预测在此期间行业指数相对上证指数在-10%—10%之间 |
| | 回避 | 分析师预测在此期间行业指数相对弱于上证指数达到或超过10% |

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层

网址：<http://www.hx168.com.cn/hxzq/hxindex.html>

华西证券免责声明

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本报告仅供本公司签约客户使用。本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持最新状态。同时，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，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。

所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，如需引用、刊发或转载本报告，需注明出处为华西证券研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。