

## 产品迭代显著增强，渠道调整基本完毕，在预算波动环境下预计保持韧性

### 核心观点

短期来看，美国宏观经济前景不确定性较高，客户在 IT 预算展望方面可能保持灵活性，导致短期面临下修风险。但 Snowflake 过去 1 年成本优化进展顺利，边际压力减轻&AI 敞口扩大，成为客户数字治理的首选平台之一，因此新增 AI 预算受益确定性较高，预计韧性较强。产品迭代上近期 Snowpark 下载量相对 PySpark 大幅提升，反馈接受度在不断追进 Databricks 产品。渠道调整上，24 年规模以上 SaaS 公司中 Databricks、Snowflake 销售人效提升幅度位列第一梯队，后续招聘专注于能带来收入的销售&研发岗位，预计人效延续平稳趋势。Snowflake 中长期发展态势较 1 年前有较大幅度改善。维持“买入”评级。

过去 1 年，Snowflake 在竞争格局及行业 IT 预算压力方面边际改善，其中

- 竞争格局上，Snowflake 拥抱 Iceberg+Polaris Catalog 后没有看到存储收入的压力，反而吸引更多客户将负载迁移至 Snowflake 执行。成本优化方面，Snowflake 2022 年 8 月以来持续推动 SQL 引擎优化、自动暂停长时间未活动负载等，年化成本节约达 20%（2023/10-2024/10），高于同行此前的水平（Mid to High single digits），我们认为大多数客户在 IT 预算优化方面实现绝大多数目标。后续的部分压力主要来自 ETL 向外部迁移，如采用云厂商或 Spark 生态的工具可以节约部分成本，但边际压力有所趋缓。
- 行业 IT 预算方面，Snowflake AI 敞口扩大，客户预算粘性增强。随着 Snowflake 过去 1 年大幅提升产品&工程迭代速度，AI 产品线日趋丰富，工程环节的 Snowpark Container Service、Iceberg 等已经开始产生收入，而应用环节的 Native App、Streamlit、Cortex AI 尚未明显贡献收入，预计 CY2H25/26 可能开始贡献一定体量的收入（受益于 DeepSeek 降低成本）。

## Snowflake(SNOW.N)

维持

买入

崔世峰

cuishifeng@csc.com.cn

SAC 编号:S1440521100004

SFC 编号:BUI663

许悦

xuyue@csc.com.cn

SAC 编号:S1440523030001

发布日期： 2025 年 04 月 21 日

当前股价： 143.43 美元

目标价格 6 个月： 175.00 美元

### 主要数据

#### 股票价格绝对/相对市场表现 (%)

	1 个月	3 个月	12 个月
	-13.53/-13.78	-16.25/-28.23	-1.01/-35.00
12 月最高/最低价 (美元)			236.00/135.48
总股本 (万股)			33,420.00
流通股本 (万股)			33,420.00
总市值 (亿美元)			514.20
流通市值 (亿美元)			514.20
近 3 月日均成交量 (万)			712.53
主要股东			
The Vanguard Group			5.40%

### 股价表现



### 相关研究报告

行业成长逻辑，上云仍然是数仓核心驱动力，2023 年上云率达 43.3%，相比于整体 50-60%的工作负载上云率，仍有一定的提升空间。据 IDC，Snowflake 所属的云关系型数据库市场 2022-27 年复合增长率预计达 20.6%，占数据管理领域的份额从 2022 年的 48.5% 下降至 46.9%，略低于行业整体 21.4% 的增速。

边际上，Snowpark 的流行度正在快速追进 Spark。2025 年 4 月 Snowpark 下载量大幅提升，4 月 1-11 日的下载量大约为 PySpark 65.0%。这里的下载量是边际值，因此从存量的角度差距仍然较大，但边际提升趋势非常积极，对于年内 Snowpark 商业化趋势具有一定前瞻性。Snowpark 客户基本上是客户转向 AI/ML 的第一步，后续是 Dynamics Tables/Cortex AI/Streamlit 等，其他 AI/ML 产品主要处于产品打磨阶段，商业化预计仍需一定时间。

**估值与建议：**短期来看，美国宏观经济前景不确定性较高，客户在 IT 预算展望方面可能保持灵活性，导致短期面临下修风险。但 Snowflake 过去 1 年成本优化进展顺利，边际压力减轻&AI 敞口扩大，成为客户数字治理的首选平台之一，因此新增 AI 预算受益确定性较高，预计韧性较强。产品迭代上近期 Snowpark 下载量相对 PySpark 大幅提升，反馈接受度在不断追进 Databricks 产品。渠道调整上，24 年规模以上 SaaS 公司中 Databricks、Snowflake 销售人效提升幅度位列第一梯队，后续招聘专注于能带来收入的销售&研发岗位，预计人效延续平稳趋势。Snowflake 中长期发展态势较 1 年前有较大幅度改善。按照 FY26E，12 x EV/Sales 给予估值，目标价 175 美元/股，同行业可比公司均值为 9.0x，但考虑数仓在数据管理技术栈的核心地位，市场空间及增速较高，确定性较强，给予一定估值溢价。FY25 公司回购价格平均在 130.87 美元/股，但在 FY4Q25 没有进行任何回购，FY4Q25 回购窗口期公司股价基本在 150-190 美元/区间。因此，我们认为公司内部考虑一定安全边际后合理估值在 130 美元/股以上，当前股价略高于内部估值，这一位置具备一定短线支撑。维持“买入”评级。

## 目录

1. 投资亮点：AI 应用技术栈转型，非结构化数据管理需求提升 .....	1
2. 数据管理架构：从三层架构转向哑铃形分布，中间层面临持续压力 .....	3
存算分离带来效率提升，成本优化空间较大 .....	3
统一的数据表格式、数据治理工具也进一步标准化数据连接层 .....	5
从数据管理延伸至 MLOps，构建 AI 应用技术栈 .....	9
Snowflake / Databricks 在架构迁移、成本优化及 AI 应用方面的进展 .....	11
成本优化：过去 1 年稳定负载成本节约 20% .....	11
AI 技术栈成熟度：MLOps 方面 Snowflake 加速追赶 Databricks，应用组件领先 Databricks .....	12
运营效率：人员规模基本与 Databricks 同步增长，但在产品/工程方面仍然落后于 Databricks，过去 1 年销售效率有所提升，领先同行 .....	16
盈利预测 .....	19
估值：目前交易于 FY26E 9.8x EV/Rev，略高于同行业可比公司 .....	20
投资评价和建议 .....	20
风险分析 .....	21
报表预测 .....	22

## 图目录

图 1:传统上数据管理的三层架构 .....	3
图 2:Elastic 不同版本存储成本与 Snowflake 存储+传输成本对比（美元/TB/月） .....	4
图 3:Polaris/Catalog 开源版本在数据爬取方面自动化程度弱于 Databricks 托管的 Catalog .....	7
图 4:Polaris 仅支持 Iceberg 格式的连接，而 Unity Catalog 支持 Delta/Iceberg/Hudi .....	7
图 5:Polaris 在权限控制方面整体优于 Unity Catalog .....	8
图 6:Unity Catalog/Polaris 在数据治理方面表现均较弱 .....	8
图 7:血缘追踪方面，Databricks 托管的 Unity Catalog 相对可用，开源版本的 Unity Catalog/Polaris 几乎不可用 .....	8
图 8:数据质量管理方面，Unity Catalog/Polaris 处于类似水平，可用性不高 .....	8
图 9:数据目录方面，Polaris 成熟度不高，仍有待提升以改善用户体验 .....	9
图 10:Snowflake 过去 12 个月对优化策略改进下企业实际负载成本下降约 20% .....	12
图 11:2023 年 8 月-2025 年 4 月 Snowpark/PySpark 下载量 .....	12
图 12:2025 年 4 月以来 Snowpark 份额大幅提升 .....	12
图 13:Polaris Catalog Github Stars 大约为 Unity Catalog 的 50.8% .....	15
图 14:Stack Overflow 显示 Streamlit 相较于 Tableau 的流行度更高 .....	15
图 15:Snowflake/Databricks 总员工数量基本同步增长 .....	17
图 16:但 Databricks 工程师数量高于 Snowflake .....	17
图 17:Snowflake 销售团队规模高于 Databricks .....	17
图 18:Databricks AI/ML 团队规模几乎是 Snowflake 的 2x .....	17
图 19:Databricks 从 Snowflake 挖角员工高于 Snowflake 反向招聘的规模 .....	17

图 20:Snowflake/Databricks 员工来源分布..... 17  
图 21:Snowflake / Databricks 过去 1 年销售额增长的拆分：1) 销售人员的增长；2) 销售人效的提升 ..... 18

## 表目录

表 1:Snowflake 发展态势评估总结..... 1  
表 2:2024 年 2 月 Elastic 与 Snowflake 存储成本、传输对比（单位：美元/TB/月） ..... 4  
表 3:三层架构与哑铃架构的优劣势对比 ..... 5  
表 4:Iceberg 对于数据管理架构整体成本的影响分析..... 6  
表 5:Unity Catalog 与 Polaris Catalog 产品功能对比 ..... 7  
表 6:在 MLOps 及 AI 应用技术栈方面，Unity Catalog 与 Polaris 的对比 ..... 10  
表 7:Snowflake vs Databricks 在 AI 应用开发方面的对比 ..... 11  
表 8:Snowflake AI 应用组件的成熟度及关键限瓶颈 ..... 11  
表 9:Snowpark 下载量异动时对应的版本更新..... 13  
表 10:Snowflake AI/ML 相关产品的指标 ..... 14  
表 11:Snowflake Streamlit 服务更新内容 ..... 15  
表 12:FY22-26 Snowflake 盈利预测（百万美元） ..... 19  
表 13:可比公司估值（截止 2025/4/19） ..... 20

# 1. 投资亮点：AI 应用技术栈转型，非结构化数据管理需求提升

过去 1 年，Snowflake 在竞争格局及行业 IT 预算压力方面边际改善，其中

- 竞争格局上，Snowflake 拥抱 Iceberg+Polaris Catalog 后没有看到存储收入的压力，反而吸引更多客户将负载迁移至 Snowflake 执行。成本优化方面，Snowflake 2022 年 8 月以来持续推动 SQL 引擎优化、自动暂停长时间未活动负载等，年化成本节约达 20%（2023/10-2024/10），高于同行此前的水平（Mid to High single digits），我们认为大多数客户在 IT 预算优化方面实现绝大多数目标。后续的部分压力主要来自 ETL 向外部迁移，如采用云厂商或 Spark 生态的工具可以节约部分成本，但边际压力有所趋缓。
- 行业 IT 预算方面，Snowflake AI 敞口扩大，客户预算粘性增强。随着 Snowflake 过去 1 年大幅提升产品&工程迭代速度，AI 产品线日趋丰富，工程环节的 Snowpark Container Service、Iceberg 等已经开始产生收入，而应用环节的 Native App、Streamlit、Cortex AI 尚未明显贡献收入，预计 CY2H25/26 可能开始贡献一定体量的收入（受益于 DeepSeek 降低成本）。

行业成长逻辑，上云仍然是数仓核心驱动力，2023 年上云率达 43.3%，相比于整体 50-60% 的工作负载上云率，仍有一定的提升空间。据 IDC，Snowflake 所属的云关系型数据库市场 2022-27 年复合增长率预计达 20.6%，占数据管理领域的份额从 2022 年的 48.5% 下降至 46.9%，略低于行业整体 21.4% 的增速。

**表 1: Snowflake 发展态势评估总结**

存在的问题/积极趋势	具体内容	评估
1.1 成本优化不透明，自动化不足	- 为实现易用性，Snowflake 将计算、存储等资源分离，降低配置/部署门槛，但导致资源计费的不透明。 - Databricks/Redshift Serverless 模式提供更精细的资源控制，达到一定阈值后自动停止运行，Snowflake 需要手动配置，可能导致 IT 支出浪费。 - 为实现 IT 支出控制，客户需要引入外部工具，导致额外运维成本，与易用性初衷相悖。	- Snowflake 2023 Summit 开始披露 Snowflake Performance Index/周度更新， <b>SPI 24/10 相比 228 性价比提升 40%，年化-20%的节约，进展积极。</b>
1.2 销售诉求与客户诉求不一致，即成本优化与承诺消费的冲突。	- 销售策略上鼓励预购/承诺付费，且 <u>年度预购积分未消耗后需要 100% 续约以换取积分使用权</u> ，这导致客户支出的浪费。 - 这种合同设置意味着如果初始设定额度高于实际需求，将持续导致资源浪费。	- 2022-23 年客户集中优化，大部分优化目标在 20-30%， <b>目前看基本优化结束，负面影响基本消化完毕。2024 年销售人效提升幅度高于同行。</b>
1.3 生态不够开放	- 如果持续聚焦成本优化（如 AWS），低成本&闭环生态更容易令人接受，但高成本&闭环生态对于构建长期技术栈的客户而言是一个较疑虑的选择。	- Iceberg 并未带来负面影响，Snowpark 客户接受度高且 AI/ML 收入贡献~3%
2.1 发布 Unistore 切入事务处理市场	- 2024 年 2 月 Snowflake 宣布 Unistore 在部分 AWS 区域推出公开预览版，试图统一数据治理。	可从分析市场切入事务处理市场，强化核心数据处理供应商的技术地位，但该产品仍处于早期预览阶段，有待进一步评估。
2.2 Data Sharing 构筑网络效应	- 持续推动数据共享，尤其是大型客户具备较大规模的数据集，在严格数据治理条件下，Snowflake 提供企业级交换平台，嵌入企业核心业务 workflow。	- 截止 2023 年 4 月，大型客户 Data Sharing 采用率 70%，总体平均值为 25%。数据飞轮效应确立意味着对大企业客户的 BI 分析预算粘性提

请务必阅读正文之后的免责条款和声明。

升。

- 其局限性在于 1) 可扩展性低导致用例不足。目前难以支持 TB/PB 级的交互，使用场景限于小数据集、离线查询或传统数据库的交换。

2.3 Document AI / Copilot / Snowpark 等	- 增强对非结构化数据的处理能力，并降低交互门槛，可通过 Python/Java/Scala 等语言进行操作。	<b>AI/ML CY3Q24 产生-5.6%收入，后续有望进一步拉动增速。</b>
3.1 IT 预算优化	- IT 预算优化仍未结束，管理层对 FY24 展望谨慎	<b>优化周期基本见底，2-3Q24 无进一步恶化趋势。</b>
3.2 数据安全/合规治理	- 欧盟 AI 法案落地，美国 AI 相关监管法案也在起草，行业严监管趋势明显，对于数据管理供应商的合规要求提升。	Azure/AWS 等优势明显。

资料来源：Snowflake 官网，Microsoft 官网，AWS 官网，PyPI，EU Commission，中信建投

**边际上，Snowpark 的流行度正在快速追进 Spark。**2025 年 4 月 Snowpark 下载量大幅提升，4 月 1-11 日的下载量大约为 PySpark 65.0%。这里的下载量是边际值，因此从存量的角度差距仍然较大，但边际提升趋势非常积极，对于年内 Snowpark 商业化趋势具有一定前瞻性。Snowpark 客户基本上是客户转向 AI/ML 的第一步，后续是 Dynamics Tables/Cortex AI/Streamlit 等，其他 AI/ML 产品主要处于产品打磨阶段，商业化预计仍需一定时间。

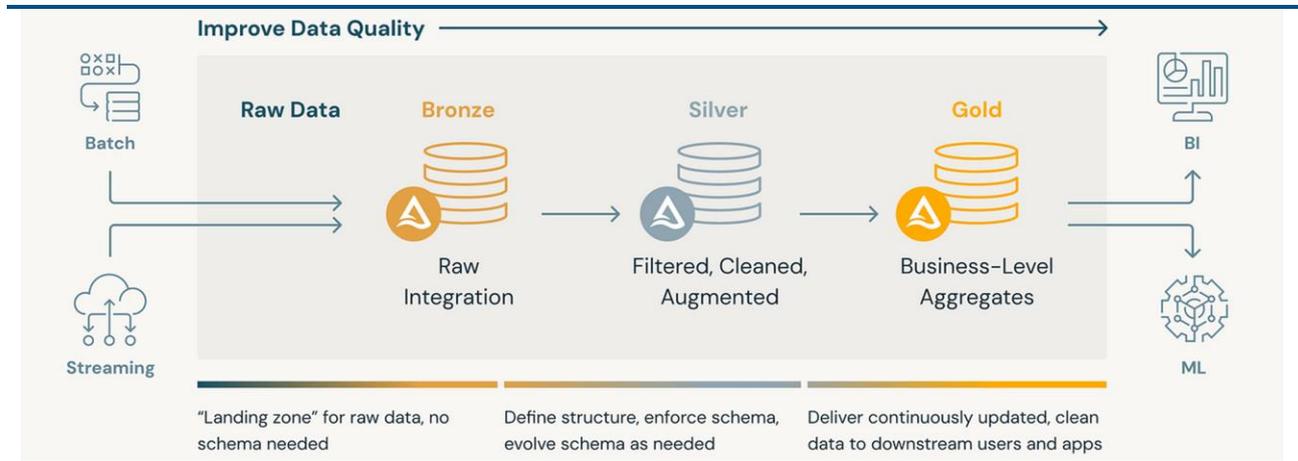
**估值与建议：**1) AI 方面，Snowflake 利用 AI 增强处理非结构化数据能力，并将下游场景从 BI 拓展至 AI/ML 等，后续关注 MLOps 等环节的变化；2) 数仓性能方面，Redshift/Databricks 在较大计算资源投入下具有较好的性能/成本优势，而 Snowflake/Big Query 在中小型计算资源下具备较好的性能/成本优势。但据 BigQuery 工程师，99% 的数仓查询都小于 10GB，绝大多数企业的数据库都小于 1TB，因此聚焦大规模查询场景的优化本质是面向 1% 的头部客户，而忽略剩余 99% 需求，在数仓领域 Snowflake 仍然具备较强的性能和成本表现；3) 大数据平台方面，Databricks 长期致力于构建开放、高性能的 Spark 生态，在开放性和工具/服务全面性上领先 Snowflake，Snowflake 通过 Snowpark / Unistore 等增强对非结构化数据的处理能力，但总体仍然落后于 Databricks。维持“买入”评级。

**风险提示：**①**市场竞争风险**-云数据仓库市场竞争激烈，竞争可能导致价格竞争、市场份额下降和利润率压力。②**依赖于云服务提供商**-Snowflake 构建在云平台之上，例如 AWS、Azure、GCP 等。它依赖于这些云服务提供商的基础设施和服务，任何与云服务提供商之间的合作关系变化、竞争、定价调整或服务中断都可能对 Snowflake 的业务和财务状况产生不利影响。③**法律和合规风险**：Snowflake 在全球范围内经营，因此需要应对各个国家和地区的法律、监管和合规要求。④**安全和数据隐私**：任何数据泄露、安全漏洞或违反数据隐私的事件都可能对 Snowflake 的声誉和客户信任产生负面影响。⑤**技术风险**：Snowflake 的成功依赖于其技术平台和解决方案的稳定性、可靠性和创新性。然而，技术创新和发展也可能带来技术风险，如软件缺陷、系统故障、数据一致性和性能问题等。

## 2. 数据管理架构：从三层架构转向哑铃形分布，中间层面临持续压力

数据管理从三层架构转向哑铃形（强化两端，弱化中间）。传统上，数据湖采用铜、银、金三层结构，其中铜层主要存储未经处理或仅轻度处理的数据，银层将数据直接转换为高度优化的分析形态，金层则是所有消费发生的地方，例如下游的 BI/ML 场景。这一架构存在的问题包括 1) 数据处理延迟：传统架构需经过原始数据到中间层（银层）再到消费层（金层）的多级转换，导致数据可用性时间较长；2) 复杂性高：需维护复杂的中间层逻辑，涉及定制化处理和非通用数据模型，增加开发与维护成本；3) 灵活性不足：未能同时满足交互式工作负载（如仪表盘）与非交互式工作负载（如批处理分析）的需求。

图 1:传统上数据管理的三层架构



数据来源：《What goes into bronze, silver, and gold layers of a medallion data architecture?》<sup>1</sup>，中信建投

针对这些问题的改进包括①分段优化（哑铃架构），具体来讲 1) 跳过中间层：直接从原始数据区转换到优化分析存储，缩短数据到洞察的时间；2) 按需定制：针对特定工作负载（如威胁建模、实时分析）创建定制化语义分析数据模型，提升效率<sup>2</sup>；3) 成本优化：降低冗余存储需求，结合廉价云存储（如 Delta Lake）节省原始存储层成本。②湖仓一体架构，涵盖 1) 支持多模态数据：通过 Delta Lake 等开放格式，统一管理结构化与非结构化数据，减少数据孤岛；2) 实时处理能力：通过流式处理（如 Kafka + Iceberg）直接导入数据湖，支持 AI/ML 场景（如自动驾驶开发）。

### 存算分离带来效率提升，成本优化空间较大

这一趋势背后仍然是存算分离带来的效率提升。以可观测性产品为例，如 Datadog/Elastic 往往主要采取存算一体架构，即数据存储和处理都在同一个节点上进行，但存算分离则强化架构的可扩展性，增强资源利用率，是经济效益驱动的选择。这种趋势下，底层的存储架构尤为重要，一些客户引入 Doris/Snowflake 等提供存储，基于外部存储方案进行可观测分析，这对应整体成本的大幅下降。

Elastic/Datadog 等厂商目前在存算分离方面也有布局，但主要集中于开源方案和一站式方案。Elastic 提供 ES-Hadoop/Elastic 与 Snowflake 等集成/Elastic Cloud，分别对应完全自建/外部集成/一体化解决方案，其中 ES-

<sup>1</sup> <https://lakshmanok.medium.com/what-goes-into-bronze-silver-and-gold-layers-of-a-medallion-data-architecture-4b6dfb405fc>

<sup>2</sup> 相应地，把通用负载卸载至存储层，并结合流传输等直接传送至终端消费环节，缩减中间层的运维/优化成本。

Hadoop 需要客户自行构建和维护 Hadoop 集群和 Elasticsearch 集群，Elastic Cloud 则完全由 Elastic 构建和维护基础设施，Elastic 与 Snowflake 集成则介于两者之间。Datadog 也有类似的方案，包括 Datadog Log Management/Archive to S3，即 Datadog 提供一站式解决方案或与 AWS 的 S3 存储集成。

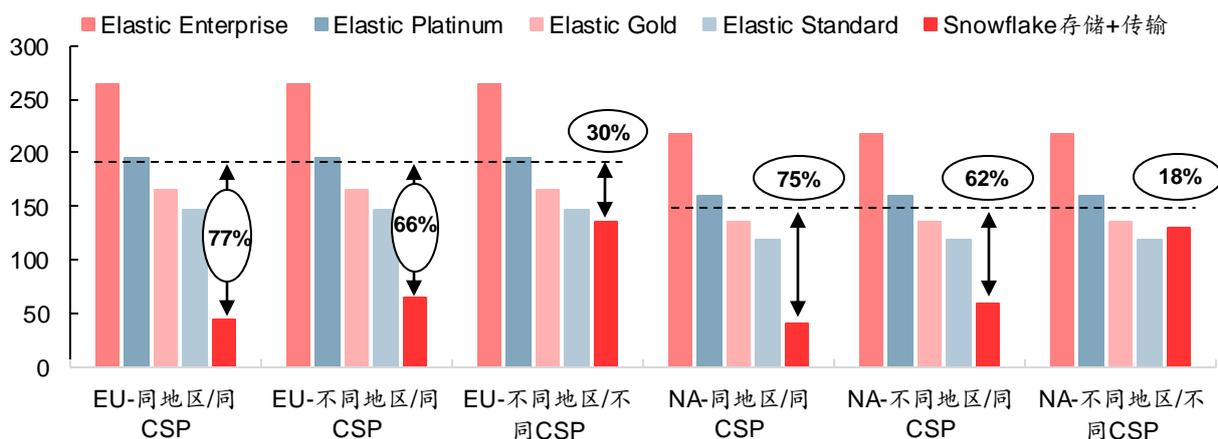
**对比 Snowflake + Elasticsearch 方案与 Elastic Cloud 方案的成本：**定性来看，Snowflake 在存储成本上低于 Elastic Cloud，主要由于列式存储架构的性能优势，但这会带来额外的网络传输成本，即将 Snowflake 的数据传输至 Elastic 集群，这可能涉及数据跨云/跨地区迁移。定量来看，2024 年初 Snowflake 在美国北弗吉尼亚州/欧洲地区的存储成本约为\$40、\$45/TB/月<sup>3</sup>，而 Elastic Enterprise 版北弗吉尼亚州冷存储价格为\$217/TB/月，爱尔兰（欧洲区数据中心）冷存储价格为\$265/TB/月<sup>4</sup>。根据测算，Snowflake 存储+传输后平均比 Elastic 存储成本低 55%，如果考虑均存储于 AWS，仅在不同地区传输则成本平均低 64%。因此，Elastic 等可观测性厂商在存算分离领域的布局尚不完善。但需要注意，将数据分布式存储可能带来响应时长提升，尽管运维分析并不需要毫秒级别的响应，但也是需要分钟级的响应，大量数据的分布式存储可能会对响应速度构成挑战，因此客户并非完全根据成本进行部署决策，而是首要考虑可靠性和稳定性，其次才是成本。

**表 2:2024 年 2 月 Elastic 与 Snowflake 存储成本、传输对比（单位：美元/TB/月）**

Elastic	Snowflake		Snowflake 传输成本-爱尔兰			Snowflake 传输成本-北弗吉尼亚				
	EU	NA	EU	NA	同地区/同 CSP	不同地区 / 同 CSP	不同地区 / 不同 CSP	同地区/同 CSP	不同地区 / 同 CSP	不同地区 / 不同 CSP
Enterprise	265	217	45	40	0	20	90	0	20	90
Platinum	195	160	45	40	0	20	90	0	20	90
Gold	165	135	45	40	0	20	90	0	20	90
Standard	146	119	45	40	0	20	90	0	20	90

资料来源: Elastic, Snowflake, 中信建投

**图 2:Elastic 不同版本存储成本与 Snowflake 存储+传输成本对比（美元/TB/月）**



数据来源: Elastic, Snowflake, 中信建投

<sup>3</sup> <https://www.chaosgenius.io/blog/ultimate-snowflake-cost-optimization-guide-reduce-snowflake-costs-pay-as-you-go-pricing-in-snowflake/>

<sup>4</sup> 均选取 AWS 为对照，根据 Elastic 官网定价计算器(<https://cloud.elastic.co/pricing>)，Platinum 版北弗吉尼亚州/爱尔兰的冷存储价格分别为 \$160/TB/月、\$195/TB/月；Gold 版北弗吉尼亚州/爱尔兰的冷存储价格分别为 \$135/TB/月、\$165/TB/月；Standard 版北弗吉尼亚州/爱尔兰的冷存储价格分别为 \$119/TB/月、\$146/TB/月。

**表 3:三层架构与哑铃架构的优劣势对比**

响应速度	高延迟（逐层处理）	低延迟（直接生成分析模型）
成本	高（维护中间层存储与计算）	低（跳过中间层，按需计算）
灵活性	通用模型难以适配垂直场景	定制模型精准匹配业务需求
维护复杂度	中等（标准化流程）	较高（多模型管理与工具集成）
适用场景	需长期稳定分析的通用场景	高时效性、垂直场景（如故障排查）
响应速度	高延迟（逐层处理）	低延迟（直接生成分析模型）

资料来源：Snowflake，《What goes into bronze, silver, and gold layers of a medallion data architecture?》，《计算存储和元数据三层分离架构下元数据和事务系统的挑战和关键技术》，中信建投

总结来看，尽管哑铃架构下增加了对于按需定制分析模型的需求，带来额外的开发及运维复杂度，导致①模型碎片化：每个定制化分析模型仅适配特定业务需求（如 Atlassian 的故障定位、供应链预测等），导致大量独立模型并存，需单独维护；②技术依赖增强：需集成多源数据（如 Splunk 日志、SignalFx 指标与自有应用数据），依赖 Dremio、Starburst 等工具直接连接原始数据和优化端，增加了技术栈复杂度；③频繁调整需求：模型需随数据特征变化迭代（如自动驾驶车辆传感器数据时效性要求），维护团队需持续监控与更新。但出于对实时性要求的提升，企业仍然倾向于将部分数据切换至哑铃架构以应对业务需求，典型用例包括①超大规模数据（如 PB 级车联网日志）下，模型需实时响应以支持高 SLA 的场景（如降级定位、自动驾驶决策）；②医疗公司的实时患者数据分析，使用 Kafka + Iceberg 实时摄取心电图数据，直接写入 Databricks 分析层，将数据生成到预警缩短至秒级，传统架构延迟则在分钟级。

新架构相应也带来一些问题，例如哑铃架构的维护负担可能随着数据规模增长而加剧，但企业可以选择利用第三方工具优化和统一管理定制化负载，例如通过工具链标准化（如 Iceberg 统一数据格式）和自动化（如 Airflow 调度模型训练）缓解重复开发问题。

按需定制的哑铃式架构本质是企业速度、成本、灵活性与复杂性、资源消耗间的战略性取舍。其核心逻辑是：牺牲部分可维护性以换取业务敏捷性。随着数据规模扩大，开发负担可能上升，但通过优化工具链和采用分层管理策略（如保留原始数据+部分中间缓存），企业可部分对冲负面影响。这一趋势反映了数据驱动时代对实时性与垂直场景深耕的迫切需求。

## 统一的数据表格式、数据治理工具也进一步标准化数据连接层

Snowflake 于 2024 加速支持 Iceberg 开放表格式。Snowflake 在 2023 年 7 月即开始通过 Iceberg Tables 更新支持数据湖工作负载，但当时主要面向早期用户。2024 年 6 月后，随着 Polaris Catalog 的开源和功能完善，其支持范围显著扩大。Snowflake 于 2024 年 6 月 3 日正式发布 Polaris Catalog，这是一个支持跨引擎访问 Iceberg 数据的开源工具，标志着其对 Iceberg 技术的深度整合。截至 2025 年 3 月，Snowflake 已有约 500 个企业账户采用 Iceberg 格式，表明其支持已进入规模化应用阶段。

Iceberg 支持 ACID 事务同时降低锁定风险，但相应牺牲专有引擎的性能优势。Iceberg 支持 ACID 事务、模式演变（Schema Evolution）及时间旅行（Time Travel），解决了传统数据湖中原子性更新和一致性难题。另外，Iceberg 采用 Parquet 文件存储数据，优化列式读取性能，并通过元数据抽象层（Table Metadata）实现数据分区和文件粒度的索引管理。Iceberg 成为事实标准后（如 HTLF 选择 Polaris 与 Iceberg 结合），用户可脱离专有存储（如 Delta Lake），降低迁移锁定的风险。相应地，采用通用格式后 Snowflake/Databricks 过去针对

SQL/Photon 引擎的优化则影响降低，客户面临牺牲性能换取开放性的权衡。据腾讯 2023 年的分享<sup>5</sup>，Iceberg 依赖对象存储（如 S3），在一些用例上存算分离导致本地计算性能损失约 30%，需更多计算资源弥补延迟。而 AWS 则分享<sup>6</sup>，对于实时摄入的场景，由于 Iceberg 元数据和版本管理的机制，会导致比较多的小文件，过多的小文件会导致查询变慢，也会带来更多的 S3 请求数量，导致成本的增加，因此需要定时对 Iceberg 表已经维护。

**Databricks 针对 Parquet 文件有针对性优化。**Databricks 的 Delta Lake 通过优化 Parquet 文件（如 Z-Ordering）提升查询效率，但传统上依赖自身生态，Iceberg 普及后逐步开放兼容（如收购 Tabular）。

**数据格式通用性也意味着 ETL 的需求相应降低，节约成本。**传统 ETL 需在数据写入后进行修正，而 Iceberg 通过 ACID 事务直接保障数据一致性，减少额外 ETL 步骤<sup>7</sup>。在预处理方面，Iceberg 的元数据版本控制允许直接查询原始数据，无需预先转换。例如，业务可直接分析 Iceberg 原始表，省去 ETL 中数据标准化的中间步骤。元数据还支持快速分区过滤，避免全表扫描，降低 ETL 对数据预处理的需求。在跨系统查询/修改时，Iceberg 作为开放表格式，支持多引擎（如 Trino、Spark）直接读写，避免传统 ETL 中数据在不同系统间迁移的开销。例如，数据可直接从 Iceberg 表供分析引擎消费，无需通过 ETL 导出到专用仓库。据小红书团队分享<sup>8</sup>，引入 Iceberg 并结合一系列数据同步策略/架构调整后存储/带宽成本优化 80%+。

**但 ETL 在复杂场景/强监管的场景下仍具备不可替代性。**ETL 可嵌入数据质量校验规则（如去重、空值填充、异常值过滤）。例如，金融业务需通过 ETL 移除敏感信息以满足 GDPR 合规要求，而实时流处理可能无法同步完成此类复杂清洗。另外，面向多种异构数据源（尤其是遗留系统）时，由于传统 ERP 系统接口封闭，需 ETL 适配器完成数据抽取，无法直接对接 Iceberg 等现代格式。

总结来看，**对于绝大多数业务场景**，例如 1) 存储成本主导型业务，例如大量社交、电商平台的用户行为数据、交易数据占据大量存储资源，对于这类场景引入 Iceberg 结合其他优化策略，小红书团队实现存储/带宽成本优化 80%+，高于此前腾讯团队测试下计算性能损失 30% 的水平，也就是总系统运行成本预计仍然是下降的；2) 中等实时分析场景，例如日志分析（运维/网络安全），结合 StarRocks 优化查询性能后，查询时长缩短 80%，也好于计算性能的损失。**但对于高并发且强实时性的场景，转变架构可能带来成本提升**，例如金融防欺诈对于延迟非常敏感，而 Iceberg 的小文件问题可能因频繁合并操作推高计算成本和延迟，相比于原有架构性能改善不明显。

**表 4:Iceberg 对于数据管理架构整体成本的影响分析**

成本维度	Iceberg 影响（个案数据）	结果
存储成本	减少数据冗余与重复存储，存储成本降低 90%	显著下降
计算成本	存算分离损失 30% 性能，小文件增加 S3 请求	可能上升（需具体分析）
ETL 开发与维护	简化管道，减少人工干预	显著下降

资料来源：腾讯云，小红书，AWS，中信建投

注：我们引用个案数据来大体反映影响程度，不具有普遍适用性，且百分比变动取决于此前各团队的数据架构/配置策略。

除统一表格式外，Snowflake 于 2024 年 6 月 3 日首次发布 Polaris Catalog，并于 2024 年 10 月 18 日全面可用（Generally Available），与 Unity Catalog 相比，Polaris Catalog 定位更加开放且中立<sup>9</sup>。**但在产品功能方面**，

<sup>5</sup> <https://cloud.tencent.com/developer/article/2205073>

<sup>6</sup> <https://aws.amazon.com/cn/blogs/china/use-serverless-architecture-to-quickly-build-a-transactional-real-time-data-lake-based-on-iceberg/>

<sup>7</sup> <https://www.infoq.cn/article/59lbbuvcrzlusmdwjbb>

<sup>8</sup> <https://juejin.cn/post/7342329071575334922>

<sup>9</sup> 据 <https://www.databricks.com/blog/open-sourcing-unity-catalog>，Unity Catalog 于 2024 年 6 月 13 日开源。

**Polaris Catalog 尚处于追赶 Unity Catalog 的状态**，例如在元数据管理方面，Unity Catalog 提供更全面的治理功能（如行/列级权限、数据血缘），覆盖数据、模型、特征全生命周期，而 Polaris 仅专注数据层；在安全合规方面，Unity Catalog 内置细粒度访问控制（如动态数据脱敏），更适合高监管行业，而 Polaris 依赖开源社区的安全策略（如 Gravitino）及 Snowflake 原生安全策略。因此 **Polaris Catalog 尚处于丰富工具箱的阶段，还没有达到成熟的端到端解决方案，因此主要吸引 Snowflake 生态内的客户，而非竞争新客户。**

**表 5:Unity Catalog 与 Polaris Catalog 产品功能对比**

功能/维度	Snowflake	Polaris Catalog	Databricks	Unity Catalog
发布时间	2024 年 6 月 3 日发布 <sup>10</sup> ，2024 年 10 月 18 日 GA <sup>11</sup>		2021 年 5 月 26 日发布 <sup>12</sup> ，2022 年 4 月于 AWS/Azure GA <sup>13</sup> ，2023 年 5 月于 GCP GA <sup>14</sup>	
技术架构	开源，基于 Apache Iceberg REST API，支持跨云/多引擎（Trino、Spark）		部分开源，深度集成 Databricks 生态（Delta Lake、MLflow），核心功能不开源	
元数据管理	统一管理 Iceberg 表，支持多云存储（如 AWS、Azure）		统一管理 Delta/Iceberg 表、模型及数据血缘，支持联邦查询外部数据源（如 Hive、CRM）	
安全与合规	依赖开源社区策略，支持行/列级权限（依赖 Iceberg），支持 GDPR 等合规审计		内置细粒度访问控制（行/列级）、动态脱敏，支持 GDPR 等合规审计	
生态兼容性	开源中立，兼容 Snowflake、Databricks、AWS 等多平台		核心功能绑定 Databricks 生态，可能引发厂商锁定	
AI/ML 集成能力	通过 Snowflake Document AI 支持模型微调与反馈循环，但功能较基础		与 MLflow 深度集成，支持模型版本追踪、自然语言查询（NLQ），提升 AI 治理效率	
客户反馈	HTLF 案例：简化 Snowflake 与 Iceberg 集成，但功能尚未成熟（需完善跨平台支持）		客户认可其开放性与 AI 能力，但未摄取数据无法用于建模，需额外开发资源	

资料来源: Snowflake, Databricks, 中信建投

**图 3:Polaris/Catalog 开源版本在数据爬取方面自动化程度弱于 Databricks 托管的 Catalog**

	Crawling (How easy is it to register all my data in the catalog?)	Data details (How much do I learn about the data in the catalog or glossary?)	Search / Discovery (How easy is it to find what I'm looking for?)
Unity Catalog OSS	F <a href="#">Manual table creation w/ schema definition</a>	C <a href="#">Schema and basic metadata</a>	C <a href="#">Basic catalog, schema, table tree with basic keyword search</a>
Unity Catalog Databricks	C <a href="#">Only tables created in Databricks automatically added, but not external data</a>	A <a href="#">Classifications, schema, stats, lineage, tags, docs, usage insights, etc</a>	A <a href="#">Navigational and intelligent search interprets schema, docs, stats</a>
Apache Polaris	F <a href="#">Manual table creation</a>	C <a href="#">Schema and basic metadata</a>	C <a href="#">Basic catalog, namespace, table tree with basic keyword</a>

 数据来源: 《unitycatalog vs apache polaris》<sup>15</sup>，中信建投

**图 4:Polaris 仅支持 Iceberg 格式的连接器，而 Unity Catalog 支持 Delta/Iceberg/Hudi**

	Open Table Formats (which are supported?)	File types and unstructured data (which are supported?)	Connectors (breadth of ecosystem)	Notable connector gaps (anything key missing?)
Unity Catalog OSS	A <a href="#">Delta, Hudi, Iceberg</a>	A <a href="#">Parquet, ORC, Avro, CSV, JSON, or TEXT</a>  <a href="#">Volumes for unstructured and semi-structured</a>	C Sources = Data lakes Engines = 5	Data lakes only No Flink or Trino
Unity Catalog Databricks	C <a href="#">Delta, Iceberg</a>  OSS Uniform and Unity can read Hudi, but not Databricks	A <a href="#">Parquet, Avro, CSV, JSON, TSV, XML</a>  <a href="#">Volumes</a>	C Sources = Data lakes External engines = 5  Can access as Unity APIs, HDFS APIs, Iceberg APIs	Data lakes only No Flink, no Hudi
Apache Polaris	D <a href="#">Iceberg Only</a>	F <a href="#">Iceberg tables only</a>	B Sources = Data lakes Engines = 20+	Nothing besides Iceberg

 数据来源: 《unitycatalog vs apache polaris》<sup>15</sup>，中信建投

<sup>10</sup> <https://www.snowflake.com/en/blog/introducing-polaris-catalog/>
<sup>11</sup> <https://other-docs.snowflake.com/en/opencatalog/release-notes#:~:text=October%2018%2C%202024&text=With%20this%20release%2C%20we%20are,available%20as%20a%20preview%20feature.>
<sup>12</sup> <https://www.databricks.com/company/newsroom/press-releases/databricks-enhances-data-management-capabilities-with-launch-of-delta-live-tables-and-unity-catalog>
<sup>13</sup> <https://azure.microsoft.com/en-us/updates?id=generally-available-unity-catalog-for-azure-databricks#:~:text=for%20Azure%20Databricks-,Azure%20Databricks,with%20the%20following%20key%20features:>
<sup>14</sup> <https://www.databricks.com/blog/announcing-general-availability-databricks-unity-catalog-google-cloud-platform>
<sup>15</sup> <https://medium.com/@kywe665/unity-catalog-vs-apache-polaris-522b69a4d7df>

**图 5: Polaris 在权限控制方面整体优于 Unity Catalog**

	Control Policy Authoring (how can user express who has access?)	Identity Authentication + Authorization (how is identity and credentials managed?)	Data lake access enforcement (is access to the root data enforced by the catalog?)	Query eng access enforcement (if access to root data enforced, which engines are compatible?)
Unity Catalog OSS	C Table and other asset level access controls Column lvl ACLs possible with shallow scope	A Authentication with 3P IDPs: Google, Okta, etc Authorization with IDP provided tokens	A Vends temp storage credential to query engine	D Spark, Delt, DuckDB, Pyspark, Graph, SpiceAI
Unity Catalog Databricks	B RBAC with privileges for tables and other assets	A Authentication with 3P IDPs SCIM for groups and identity federation	A Vends temp storage credential to query engine	D Spark, Fabric, DuckDB, Trino
Apache Polaris	A RBAC with granular privileges for tables and other asset level access controls	C Client ID and secret use OAuth2 for access token	A Vends temp storage credential to query engine	A All compatible with Iceberg REST Catalog API

数据来源: 《unitycatalog vs apache polaris》, 中信建投

**图 6: Unity Catalog/Polaris 在数据治理方面表现均较弱**

	Classification (can I annotate and classify certain tables, columns?)	Retention policies (Can data retention policies be set and enforced?)	Auditing (Can I audit all actions taken?)
Unity Catalog OSS	D Key/value properties for each asset	F No metadata or data lifecycle governance features	F No audit logs
Unity Catalog Databricks	B Tag based classification and auto detection of sensitive data	F No metadata or data lifecycle governance features	A Audit logs
Apache Polaris	F No tagging or classification features	F No metadata or data lifecycle governance features	F No audit logs

数据来源: 《unitycatalog vs apache polaris》, 中信建投

**图 7: 血缘追踪方面, Databricks 托管的 Unity Catalog 相对可用, 开源版本的 Unity Catalog/Polaris 几乎不可用**

	Data Lineage (How well can the source of data be tracked?)	Data Ownership Accountability (Is data ownership clear? How is accountability maintained?)	Metric Definitions (Can metrics be clearly defined, traced, and monitored?)
Unity Catalog OSS	F No data lineage	F	A Create, govern and leverage "functions"
Unity Catalog Databricks	B Rich visualizations, lineage, only from Spark DF or DBSQL	A Ownership tracked and visible	A User defined functions
Apache Polaris	F No data lineage	F	F No metric definitions, but can create a view

数据来源: 《unitycatalog vs apache polaris》, 中信建投

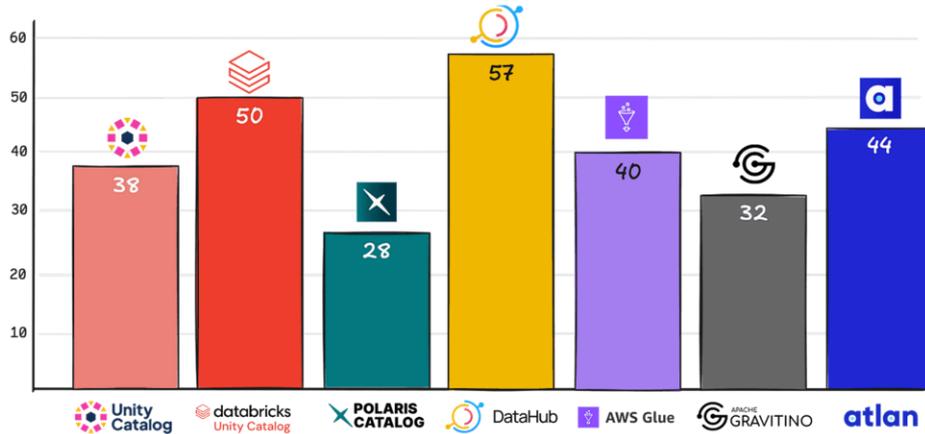
**图 8: 数据质量管理方面, Unity Catalog/Polaris 处于类似水平, 可用性不高**

	Schema Management (can schema be registered, monitored, or enforced?)	Data Quality policies (Can data quality expectations be set and monitored?)	Monitoring + Alerting (How well can you monitor and be made aware when there is a problem?)	Data Freshness (Can you monitor how recently your data has been updated?)
Unity Catalog OSS	C Basic schema evolution	F	F	F
Unity Catalog Databricks	C Basic schema evolution	F Data Constraints, Expectations, etc can be set in Databricks pipelines, not a feature of UC alone	F Slack, webhooks, etc, for Lakehouse Monitoring, not a feature of UC	F
Apache Polaris	C Basic schema evolution	F	F	F

数据来源: 《unitycatalog vs apache polaris》, 中信建投

图 9:数据目录方面, Polaris 成熟度不高, 仍有待提升以改善用户体验

### Data Catalog Comparison Metascore



数据来源: 《unity catalog vs apache polaris》, 中信建投

后续关注 Polaris/Unity Catalog 在如下方面的改进: 1) 自动化爬虫遍历数据并注册至数据目录中, 降低管理/维护成本; 2) Polaris 是否增加支持第三方身份验证产品, 如 Okta、Google Auth 等; 3) 增强对于非 Iceberg 格式的支持; 4) 强化自动化数据治理工具, 例如数据保留策略以符合外部合规要求, 提供审计日志, 自动化检测并进行权限分类; 5) 强化与后端 MLOps 的集成。

## 从数据管理延伸至 MLOps, 构建 AI 应用技术栈

由于对非结构化数据支持度不足, 且下游模型部署依赖外部工具, 全生命周期管理能力弱于 Databricks。在 AI/ML 功能集成方面, Databricks 通过统一的数据湖仓 (Delta Lake) 整合结构化与非结构化数据, 直接支持机器学习全流程 (数据准备→特征工程→模型训练→部署监控), 同时集成 MLflow (实验跟踪、模型注册)、AutoML、Feature Store (特征管理) 和向量索引服务, 减少对外部工具的依赖; 而 Snowflake 以 Snowpark 为核心, 通过 Python/Java API 支持数据转换和机器学习, 但依赖第三方工具 (如 Nvidia NeMo、Dataiku) 实现模型部署和监控, 由于对非结构化数据支持度不足, 且数据目录 Polaris 聚焦于 Iceberg 格式管理, 无法进行跨系统的联邦查询和血缘追踪。

Databricks 在 MLOps 环节具备优势, 但 DeepSeek、Qwen 等团队在 MLOps 方面的开源推动 SaaS 团队缩小差距。2024 年 2 月 Databricks 宣布以 13 亿美元收购 MosaicML, 主要考虑是纳入其 ML 团队 (此前发布正交微调框架, 优化模型微调效率)。此后, Snowflake 于 2024 年 5 月考虑以 10 亿美元收购 Reka, 强化自身 ML 团队能力, 提供自研模型训练/推理框架, 但后续交易终止。目前 Snowflake 暂无原生的分布式训练框架<sup>16</sup>。但考虑到 DeepSeek、Qwen、Google 等团队在 MLOps 方面持续的开源工作, 大量中小 SaaS 厂商溢价收购模型团队的意义正在缩小, 跟随业界开源工作并做好与生态的集成适配就能够满足大多数客户的需求。

Databricks 和 Snowflake 在 SQL 引擎方面取向不同, 但在复杂场景下 Databricks 路线具备优势。Databricks 的 Genie 工具适应于预定义语义层后进行自然语言转换 SQL, 准确率较高, 误报/后期验证成本较低; 而 Snowflake Cortex AI 无需预定义语义层, 但通过 LLM 解析转换 SQL 效果弱于 Genie, 容易导致后续审核/验证的额外成本。换言之, 在复杂场景下预定义语义层可以更精确地进行转换, Genie 的效果更佳, 而标准化场景下双方差异不大。

<sup>16</sup> <https://www.snowflake.com/en/blog/up-to-75-lower-inference-cost-llama-meta-llm/>, snowflake 推出 SwiftKV 优化推理框架。

典型的复杂用例包括自动驾驶数据标注管道、实时视频流特征提取等，而简单用例包括财务部门报表自动化提取。

在血缘追踪方面，Snowflake 覆盖度不如 Databricks 全面。Databricks 通过 Unity Catalog（统一元数据管理层）实现了跨数据湖、数据库和实时流数据的端到端血缘追踪能力，其中 1）包括非结构化数据处理（如 JSON、图像、文本）的元数据关联；2）Unity Catalog 基于 Delta Lake 的 ACID 事务特性，通过事务日志（Transaction Log）记录所有数据操作（如 INSERT/UPDATE/MERGE）的上下游依赖关系，并实时更新血缘图谱。Snowflake 血缘追踪主要通过 Snowflake Account Usage Schema 提供，但其设计更侧重于表级和查询级的统计信息（如查询历史、访问日志），颗粒度弱于 Databricks，因此无法追溯 Python/Scala/Java 等代码中的动态数据处理逻辑，且对 Airflow、dbt 等外部工具操作的元数据捕获能力较弱。在 AI/ML 方面，Databricks 针对生成式 AI 场景（如微调、模型部署）设计 MLflow Tracking<sup>17</sup>，可自动记录数据输入、模型版本、参数和输出结果，形成完整的实验血缘；而 Snowflake 的 AI/ML 功能（如 Snowpark ML）更依赖外部工具（如 AWS SageMaker），血缘信息需手动维护，难以自动化扩展。

**表 6:在 MLOps 及 AI 应用技术栈方面，Unity Catalog 与 Polaris 的对比**

功能对比	Unity Catalog	Polaris Catalog
AI/ML 功能集成	通过 MLflow 实现模型注册与生命周期管理，支持端到端审计（如 Experian 追踪模型输入输出）	通过 MLflow 实现模型注册与生命周期管理，但不支持模型层面的血缘追踪/日志审计，无原生模型训练/监控能力
	收购 MosaicML，内置正交微调框架，提升微调效率	收购 Reka 失败，依赖外部开源工具集（如 Hugging Face 等）
	Genie 工具支持自然语言转 SQL，依赖语义层理解数据模式，在定义清晰情况下性能更优，但仅限于 Databricks 生态	Cortex AI 支持 LLM 进行自然语言和 SQL 转换，无需预定义语义层（表/列的注释），可跨系统交互
	新增 AI 代理开发能力，管理 API 函数与工具链（如 Workday/NetSuite 接口）	核心场景围绕数据湖仓迁移，未涉及复杂 AI
数据治理与血缘追踪	血缘覆盖至模型级别（如 Experian 追踪模型输入输出，符合金融业监管）	血缘能力限于数据层（仅管理 Iceberg 表目录，未提模型/特征血缘）
	支持联邦查询外部 SQL 源（如未摄取的 CRM 数据）	无外部数据联邦治理能力（仅支持 Iceberg 格式），聚焦 Snowflake 生态内数据
生态系统与工具链	闭环生态：DeltaLive Tables 优化训练数据准备，与 MLflow 结合	需搭配多工具链（如 Informatica、OpenAI）
	向量搜索可通过 Weaviate 扩展	实际部署依赖多平台（如 Snowflake、Databricks、Azure SQL）
应用场景侧重	金融业复杂场景：DataRobot 通过 Unity 构建财务风控代理，自动化合同审计	核心场景为数据湖仓迁移（HTLF 选择主因兼容 Snowflake）
	实时 AI 处理：米其林用 DeltaLive Tables 优化供应链预测	未涉及复杂 AI 场景

资料来源：Snowflake, Databricks, 中信建投

在 AI 应用构建上，Snowflake 推出 Snowpark Container Services、Native App Framework、Streamlit 等工具，而 Databricks 则依靠既有工具组合，例如 MLflow、Notebooks 等，成熟度低于 Snowflake（工具箱 vs 解决方案）。

<sup>17</sup> <https://docs.databricks.com/aws/en/mlflow/tracking>

具体到 Snowflake AI 组件，根据一些早期反馈<sup>18</sup>，客户寻求将部分运维的工作负载迁移至 Snowpark 之上，在不考虑折扣的情况下，Snowpark Container Services 的价格较 EKS 标价低~20%，但 Snowflake 不具备 EKS 的所有功能，且面临低/中等吞吐量和延迟（每笔交易 10-50 毫秒）的限制。而复杂工作负载下相比于 Databricks 基于 Photon 引擎，Snowpark 的计算成本较高，目前更适应于简单负载。另外，更长远来看，在 Snowflake、Snowpark 运行分析负载后无需支付 CSP 的数据传输费用，长期来看计算引擎的优化速度快于带宽<sup>19</sup>，就地处理负载的方案性价比提升。

**表 7: Snowflake vs Databricks 在 AI 应用开发方面的对比**

Snowflake AI 应用相关功能	Databricks AI 应用相关功能	对比
Snowpark Container Services	MLflow+ 云容器服务	Databricks 依赖云厂商容器，Snowflake 提供全托管服务
Native App Framework	Notebooks + dbutils	Databricks 缺乏模块化应用框架，功能较分散
Streamlit	Tableau/SQL 仪表盘	Databricks 无原生低代码工具，需依赖第三方

资料来源: Snowflake, Databricks, 中信建投

**表 8: Snowflake AI 应用组件的成熟度及关键瓶颈**

组件	成熟度	适用场景	关键限制
Snowpark Container Services	中等	模型部署与轻量级 AI 服务	生态案例少，依赖外部工具链
Snowflake Native App	低	数据仓库内简单应用开发	功能基础，无法支持复杂 AI
Streamlit	高	快速原型开发、数据科学展示	安全风险、扩展性不足

资料来源: Snowflake, Reddit, 中信建投

关于 Native App/Streamlit，二者均用于应用开发<sup>20</sup>，但 Native App 可将应用以包体形式组装便于分发和管理版本，而 Streamlit 是一种更松散的形式，主要可用于内部应用。Snowflake 战略上倾向于引导合作伙伴/客户构建 Native App 后在生态内销售，从而产生规模效应，降低中小客户的应用门槛。目前 Native App 的典型用例<sup>21</sup>包括 1) 财务报表实时分析；2) 营销自动化；3) 销售业绩洞察；4) 实时整合商品库存及用户行为数据；5) 供应链报告自动化。Streamlit 此前一直是开源生态内 python app 开发的流行框架，成熟度相对较高，但企业级安全等尚待改善。

## Snowflake / Databricks 在架构迁移、成本优化及 AI 应用方面的进展

### 成本优化：过去 1 年稳定负载成本节约 20%

与 Databricks 引入 Catalyst 优化器对应，Snowflake 通过研发力量引入重大改进（成为平台默认配置）大部分都是自动发生的，无需任何配置或额外的努力来修改代码。1) **查询执行改进**：缩短执行时间并更有效地处理复杂的查询模式。示例包括优化连接查询、自动处理偏差和扩展对 Top-K 修剪的支持，以提高具有特定聚合和过滤模式的查询的性能。2) **数据提取和复制**：减少元数据复制所花费的时间，加快克隆速度，并优化大型数据集的提取，以更快、更可靠地将数据带入 Snowflake，从而简化工作流程和管道。3) **自适应优化**：推出一系列自适应优化，使 Snowflake 能够更智能地选择最佳查询执行策略。例如，扩展 Top-K 修剪以包含更广泛的查询。4) **平台效率**：Snowflake 继续提升平台的整体可靠性和速度。例如，团队缩短克隆操作所需的时间，提高

<sup>18</sup> [https://www.reddit.com/r/snowflake/comments/1eg2iso/eli5\\_snowpark\\_container\\_services Seeking overview/](https://www.reddit.com/r/snowflake/comments/1eg2iso/eli5_snowpark_container_services Seeking overview/)

<sup>19</sup> [https://www.reddit.com/r/dataengineering/comments/1dl52cu/any\\_frequent\\_snowpark\\_users\\_here/](https://www.reddit.com/r/dataengineering/comments/1dl52cu/any_frequent_snowpark_users_here/)

<sup>20</sup> [https://www.reddit.com/r/snowflake/comments/1h117m1/native\\_apps\\_vs\\_streamlit/](https://www.reddit.com/r/snowflake/comments/1h117m1/native_apps_vs_streamlit/)

<sup>21</sup> <https://www.infometry.net/blog/snowflake-native-apps/unlocking-innovation-real-world-use-cases-of-snowflake-native-apps-in-2025/>



**表 9: Snowpark 下载量异动时对应的版本更新**

日期	重大异动时更新	具体功能
2024/6/25	Snowpark Python 1.19.0 发布	<p><b>兼容性提升:</b> ①支持更多 Pandas 原生功能 (如 <code>df.index</code> 索引操作), 降低学习成本; ②优化数据上传逻辑, 减少内存占用; ③更直观的错误提示, 当使用未实现的功能时, 直接提示“暂不支持”而非复杂报错。</p> <p><b>问题修复:</b> ①修复了处理超大表格 (500 列以上) 时可能崩溃的缺陷; ②修正了部分函数因数据索引不连续导致的报错问题; ③解决存储过程在任务中执行失败、时区转换错误等兼容性问题。</p> <p><b>增强数据处理能力:</b> ①支持将数据直接转为“真/假”格式 (<code>to_boolean</code>)。新增对日期、时间字段的更多操作 (如星期几、年份中的第几天); ②支持计算百分比变化 (<code>pct_change</code>)、字符串截取 (<code>str.get</code>)、去空格 (<code>lstrip/rstrip</code>)。优化分组统计 (如 <code>size</code>)、滑动窗口计算 (如 <code>rolling</code> 求和/均值); ③通过优化技术, CSV 文件上传速度提升约 2 倍, 支持更多数据格式 (如自动识别日期、数字)。</p>
2024/7/17	Snowpark Python 1.20.0 发布	<p><b>兼容性继续提升:</b> ①支持 <code>df.assign</code> (快速新增列)、<code>df.stack</code> (堆叠数据)、<code>df.equals</code> (对比数据一致性); ②优化 <code>plot()</code> 绘图功能 (数据自动下载到本地处理); ③支持直接修改索引顺序 (<code>sort_index</code> 原地操作); ④空数据透视表也能正常生成, 无需额外处理。</p> <p><b>易用性提升:</b> ①新增后台任务追踪功能 (如存储过程、UDF), 方便开发者排查问题; ②提升数值计算速度, 减少不必要的数据类型转换; ③修复处理大整数 (如 <code>uint64</code>) 时精度丢失; ④修复透视表 (<code>pivot_table</code>) 生成时列名混乱问题。</p> <p><b>增强数据处理能力:</b> ①新增 <code>arrays_zip</code> 函数, 可合并多列数组数据; ②支持直接导出数据到 CSV 文件 (<code>to_csv</code>); ③新增 <code>nlargest/nsmallest</code> 快速筛选最大/最小值, <code>pivot</code> 透视表功能; ④支持自定义模拟函数参数 (如指定唯一值、传递行列索引); ⑤保存表时可控制列顺序, 避免数据错位。</p>
2024/10/9	Snowpark Python 1.23.0 发布	<p><b>兼容性提升:</b> ①新增读取 SAS 文件功能, 支持更多数据源。处理空数据或特殊索引时更稳定 (如索引名与列名重复)。</p> <p><b>增强数据处理能力:</b> ①新增生成时间间隔的工具 (<code>make_interval</code>), 支持更灵活的时间窗口分析。支持时区转换 (<code>tz_localize/tz_convert</code>) 和本地化时间处理, 避免时区混乱。时间序列重采样支持更多频率 (如周、月末、年末) 和统计方法 (如 <code>nunique</code>、<code>quantile</code>); ②支持直接导出数据到文件 (如 CSV/SAS)。新增快速筛选最大/最小值 (<code>nlargest/nsmallest</code>)、透视表 (<code>pivot</code>) 功能。优化 <code>map/apply</code> 方法, 支持更多自定义函数操作; ③新增存储过程性能分析工具, 方便定位运行慢的问题。自动清理临时数据, 减少内存占用。</p>
2025/2/5	Snowpark Python 1.27 发布	<p><b>增强函数支持:</b> 数学与统计新增反双曲函数 (<code>acosh/asinh</code>)、立方根 (<code>cbrt</code>)、百分位近似 (<code>percentile_approx</code>) 等。字符串处理支持 Base64 编码/解码 (<code>base64/unbase64</code>)、十六进制转换 (<code>hex</code>)、编辑距离 (<code>edit_distance</code>)。空值处理新增 <code>divnull</code> (除零安全)、<code>nullifzero</code> (零转空值)、<code>nvl</code> (空值替换)。JSON 与结构化数据支持 <code>from_json</code> 解析、<code>map_cat</code> 合并键值对、<code>map_contains_key</code> 检查键是否存在。</p> <p><b>增强数据处理能力:</b> 支持直接插入数据到现有表 (<code>insert_into</code>), 并兼容本地测试模式。新增临时视图创建 (<code>create_temp_view</code>), 避免命名冲突。允许在 <code>pivot</code> 分组后使用多</p>

请务必阅读正文之后的免责条款和声明。

个聚合函数（如同时统计均值和总和）。支持自定义列顺序（keep\_column\_order）、结构化类型字段重命名。数组和字典类型支持标记是否包含空值（contains\_null）。

**易用性提升：**①新增对齐填充（ljust/rjust/center）、时间格式化（strftime 支持 %Y-%m-%d 等常用格式）。支持快速判断值范围（Series.between）。②map/apply 方法提速，自动映射 NumPy 函数到 Snowpark 原生函数。减少冗余警告（如 SettingWithCopyWarning）。③支持从字典/记录快速创建 DataFrame（from\_dict/from\_records）。分组聚合新增 unique、first/last 等统计方式。

**涉及复杂应用：**①对象管理，通过 Session.catalog 直接管理数据库对象（如表、视图）。②JSON 文件读取，支持从文件路径直接推断复杂结构（无需手动定义 Schema）。

**增强数据处理能力：**①数据转换更灵活，新增 relaxed\_ordering 参数，允许在读取数据时放宽顺序要求，提升处理超大数据集的效率。支持从 Snowflake 表或查询结果直接读取数据时，自动适配索引标签（不再强制要求手动指定）；②优化随机抽样（sample\_by/random\_split），生成更简洁的 SQL 查询，速度提升明显（尤其适合复杂抽样）。Parquet 文件导入提速，启用矢量化处理减少内存占用。

**兼容性提升：**fillna 和 replace 支持自动将整数/浮点数适配到 Decimal 类型，避免类型报错。新增对 pd.Grouper 时间分组的支持（如按月/年聚合）。

2025/3/27 Snowpark Python 1.30.0 发布

资料来源：Snowflake<sup>22</sup>，中信建投

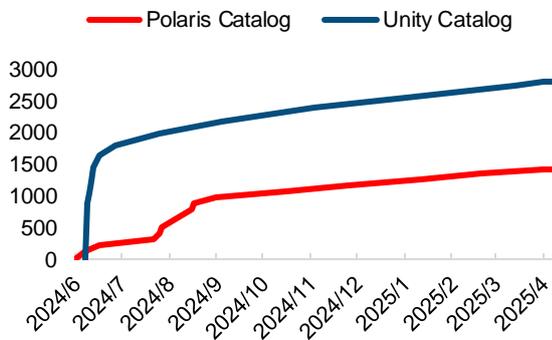
**表 10: Snowflake AI/ML 相关产品的指标**

AI 指标	CY24Q1	CY24Q2	CY24Q3	CY24Q4
Snowpark 客户占比	50%	/	/	/
AI/ML 客户数	/	2500	3200	4000
Iceberg 客户数	300	400	500	/
Dynamic tables 客户数	/	/	/	/
Notebooks 客户数	/	1600	/	/
Cortex AI 客户数	750	/	/	/
客户采用 Snowflake 处理非结构化数据的比例	40%	/	/	/

资料来源：公司公告，中信建投

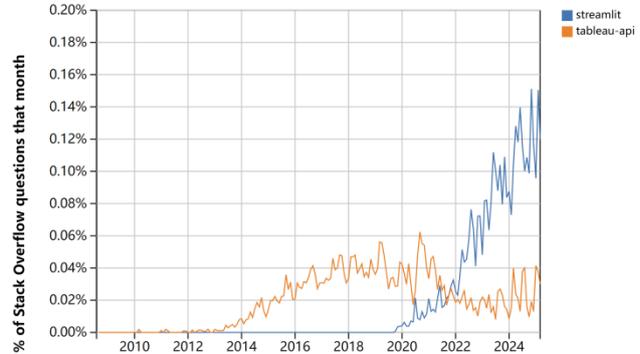
<sup>22</sup> <https://docs.snowflake.com/en/release-notes/clients-drivers/snowpark-python-2025>

图 13:Polaris Catalog Github Stars 大约为 Unity Catalog 的 50.8%



数据来源: Github, 中信建投

图 14:Stack Overflow 显示 Streamlit 相较于 Tableau 的流行度更高



数据来源: StackOverflow<sup>23</sup>, 中信建投

Streamlit 方面, 目前仍聚焦于支持私有云环境、强化数据传输安全等, 对 AWS/Azure/GCP 环境的适配。在产品功能方面于 2025Q1 开始迭代, 支持文件上传、多种文件类型 (如音视频), 支持复杂页面管理和 Git 同步 (以便于版本控制), 但仍然较为基础。我们预计 Streamlit 正在不断强化应用开发能力, 有助于为后续 Snowflake 生态降低开发和部署门槛, 但 1-2 个季度内可能不会看到商业化的明显信号。

表 11:Snowflake Streamlit 服务更新内容

日期	更新内容	状态	详细描述
2025/4/7	Google Cloud Private Service Connect 支持	Preview	支持通过 Google Cloud Private Service Connect 建立私有连接, 避免数据通过公共互联网传输, 增强安全性。
2025/3/27	Git 集成与多页应用支持	Preview	1.多页应用: 通过 pages/目录管理多页面, 支持导航栏自动生成。2.Git 同步: 在 Snow sight 中直接关联 Git 仓库, 支持版本控制。
2025/3/24	st.experimental_audio_input 和 st.camera_input	GA	1.音频输入: 通过麦克风录制音频并上传。2.摄像头输入: 捕获图像或视频流, 支持实时处理。
2025/3/14	st.file_uploader 支持	GA	支持上传多种文件类型 (如 CSV、JSON), 并可限制文件大小和类型。
2025/2/14	st.file_uploader 预览	Preview	初步支持文件上传功能, 但限制文件大小不超过 200MB。
2025/2/7	Material 图标支持	GA	支持在按钮、标签等组件中使用 Material Design 图标
2025/1/31	Future Grants 支持	GA	允许为未来创建的对象 (如表、视图) 预定义权限, 简化权限管理。
2024/12/20	Streamlit 1.39.0 支持	Preview	支持 Streamlit 1.39.0 新特性, 如 st.pydeck_chart 的性能优化和 st.data_editor 的只读模式。
2024/12/16	Azure Private Link 支持	GA	通过 Azure Private Link 实现私有 VNet 连接, 确保数据传输不经过公共网络。
2024/12/4	Azure Private Link 预览	Preview	预览版 Azure Private Link 支持, 需手动配置 DNS 和端点。
2024/10/31	自定义主题支持	Preview	允许通过 theme.base 和 theme.primaryColor 等参数自定义应用主题颜色和字体。
2024/10/31	AWS PrivateLink 正式支持	GA	通过 AWS PrivateLink 实现跨 VPC 安全连接, 支持 Snowflake 与用户 AWS 账户的私有通信。

<sup>23</sup> <https://trends.stackoverflow.com/?tags=streamlit,tableau-api>

2024/10/7	AWS PrivateLink 预览	Preview	预览版 AWS PrivateLink 支持，需配置终端节点服务。
2024/8/9	AWS GovCloud 支持	GA	支持在 AWS GovCloud 部署 Streamlit 应用，满足 FedRAMP 合规要求。
2024/8/2	自定义 UI 正式支持	GA	支持通过以下方式自定义 UI: 1.st.markdown(unsafe_allow_html=True)插入 HTML/CSS。 2.st.components.v1.html 嵌入 iframe 或 JavaScript。
2024/8/1	Streamlit 1.35.0 支持	GA	支持 Streamlit 1.35.0 新特性，如 st.query_params 和 st.connection 的优化。
2024/7/31	上下文函数与行级访问策略	GA	1.上下文函数：通过 st.connection 直接调用 Snowflake 数据。2.行级策略：基于用户角色过滤数据行。
2024/7/3	外部网络访问	GA	允许通过 EXTERNAL_ACCESS_INTEGRATION 安全访问外部 API，需在 Snowflake 中预配置集成。
2024/6/28	自定义 UI 预览	Preview	允许通过 HTML/CSS 自定义应用布局，但需启用 unsafe_allow_html 或使用 st.components。
2024/5/14	GCP 环境正式支持	GA	支持在 Google Cloud Platform 部署 Streamlit 应用，与 BigQuery 等服务集成更便捷。
2024/5/8	自定义休眠计时器	Preview	通过 config.toml 设置应用空闲后自动休眠时间（5-240 分钟），降低计算资源消耗。
2024/3/15	Streamlit 1.26.0 支持	Preview	支持以下新组件：构建聊天界面、显示异步任务状态、开关按钮。
2024/1/25	Azure 环境正式支持	GA	支持在 Azure 部署 Streamlit 应用，与 Azure Synapse 等服务无缝集成。

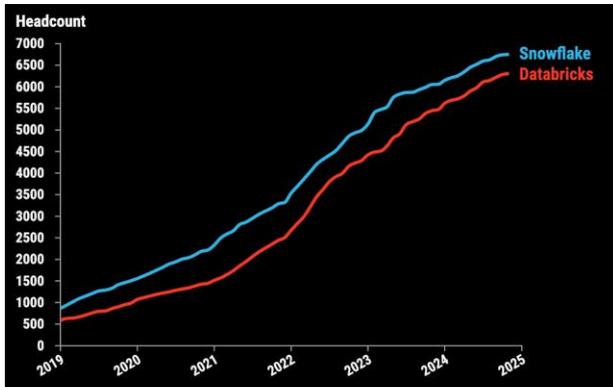
资料来源：Snowflake<sup>24</sup>，中信建投

## 运营效率：人员规模基本与 Databricks 同步增长，但在产品/工程方面仍然落后于 Databricks，过去 1 年销售效率有所提升，领先同行

FY24-25 Snowflake 员工新增规模在 1000 人左右，且聚焦于 R&D 及 S&M，尤其是 AI/ML 部门的团队扩张。FY26 管理层指引倾向于增加收入直接相关的部门员工（工程及销售团队）。产品侧过去几个季度管理层分享的指标主要涉及 Iceberg、Notebooks、Cortex AI、Snowpark 等 Data Engineering 产品，其他诸如应用层 Native App/Streamlit 短期内无法产生实质性收入的产品优先级有所下降。因此，我们认为 AI 产品方面 Snowflake 目前更多处于产品打磨阶段，而前述分析总结来看大多数产品处于“基本可用”（及格水平）、“难用”（不及格）阶段，没有达到“易用”水平，缺乏足够多的标杆案例吸引客户投入预算进行 AI 流程改造。

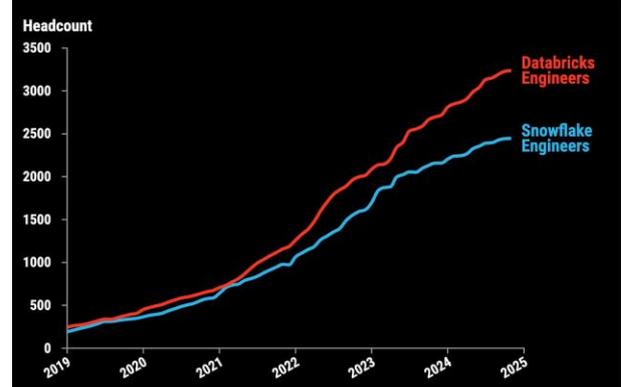
<sup>24</sup> <https://docs.snowflake.com/en/release-notes/streamlit-in-snowflake>

图 15: Snowflake/Databricks 总员工数量基本同步增长



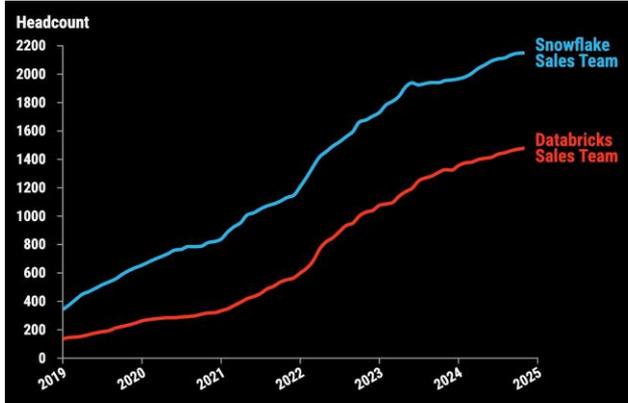
数据来源: Live Data, 中信建投

图 16: 但 Databricks 工程师数量高于 Snowflake



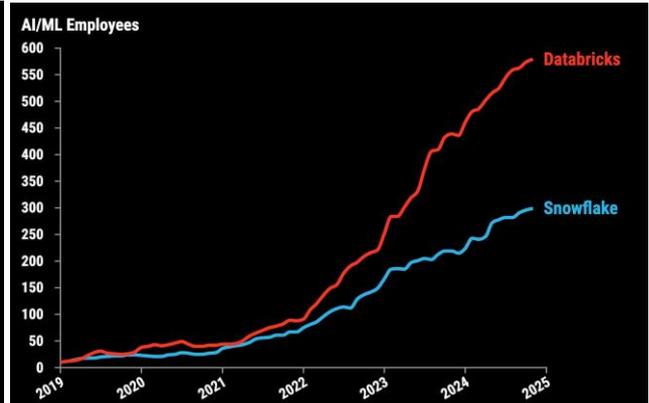
数据来源: Live Data, 中信建投

图 17: Snowflake 销售团队规模高于 Databricks



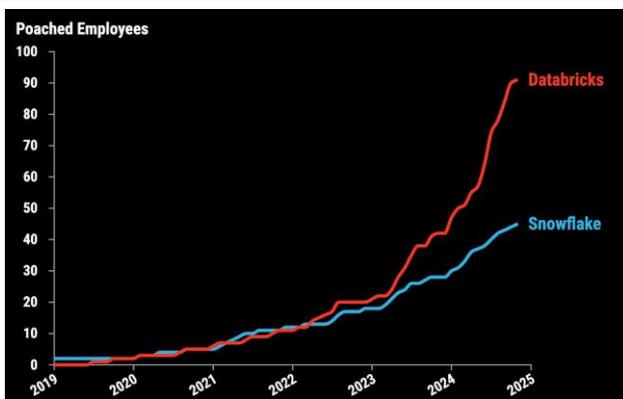
数据来源: Live Data, 中信建投

图 18: Databricks AI/ML 团队规模几乎是 Snowflake 的 2x



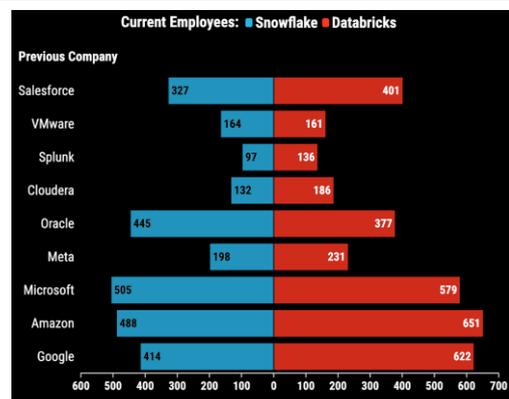
数据来源: Live Data, 中信建投

图 19: Databricks 从 Snowflake 挖角员工高于 Snowflake 反向招聘的规模



数据来源: Live Data, 中信建投

图 20: Snowflake/Databricks 员工来源分布



数据来源: Live Data, 中信建投

**图 21: Snowflake/Databricks 过去 1 年销售额增长的拆分: 1) 销售人员的增长; 2) 销售人效的提升**

Company Name	RepVue S	Month of	Sales Output Change	Sales Headcount Ch	Quota Attainment Char
Rippling	86.8775379	1/1/25	122%	110%	10%
Atlassian	86.8528528	1/1/25	46%	58%	-15%
Databricks	87.0184999	1/1/25	44%	41%	4%
CrowdStrike	87.8487029	1/1/25	38%	46%	-10%
Verkada	80.527758	1/1/25	38%	48%	-11%
Grafana Labs	85.9989919	1/1/25	37%	40%	-4%
Snowflake	86.1145431	1/1/25	36%	31%	6%
Samsara	87.740302	1/1/25	31%	46%	-19%
Paylocity	80.8225939	1/1/25	30%	31%	-2%
SailPoint	86.7141721	1/1/25	29%	32%	-3%
Klaviyo	84.8568419	1/1/25	28%	42%	-18%
Palo Alto Networks	86.3235395	1/1/25	28%	30%	-3%
Paycor	77.2832517	1/1/25	26%	31%	-7%
Stripe	86.0429229	1/1/25	25%	12%	20%
Workiva	88.5370088	1/1/25	25%	28%	-4%
ServiceNow	90.4561422	1/1/25	23%	28%	-6%
Adobe	85.7641669	1/1/25	23%	22%	2%
UiPath	84.0366629	1/1/25	23%	24%	-2%
Birdeye	76.5913002	1/1/25	22%	22%	-1%
Gong.io	86.0637739	1/1/25	21%	6%	21%
Shopify	81.6191368	1/1/25	21%	31%	-14%
Toast	87.7035994	1/1/25	20%	34%	-19%
OpenText	74.6393351	1/1/25	20%	19%	1%
Datadog	82.4200987	1/1/25	19%	25%	-7%
OneTrust	76.7014203	1/1/25	19%	21%	-3%
MongoDB	85.6014775	1/1/25	18%	24%	-9%
CloudFlare	78.5617384	1/1/25	17%	20%	-4%
ZoomInfo	86.5396889	1/1/25	16%	16%	0%

数据来源: RepVue, 中信建投

## 盈利预测

关键假设：1) 中性假设 FY25 增速与公司指引匹配，对应 IT 预算大规模优化结束，但 IT 支出仍然保持谨慎，而没有大幅反弹；2) 产品收入占总收入比重维持在 94%-96% 的高水平，即 Snowflake 主要收入来自其核心产品线；3) 随着 Snowflake 在扩大规模和市场份额时，短期内支出增加导致利润率承压，但长期将获得规模效应。

我们预计公司 FY2026-28 年营业收入分别达 44.8/55.0/66.7 亿美元，同比增速分别为 23.6%/22.6%/21.3%。毛利润分别为 30.2/37.6/47.8 亿美元，毛利率分别为 67.1%/68.1%/69.6%。Non-GAAP 净利润 FY2026-28 预计达 4.3/6.1/8.6 亿美元，对应 Non-GAAP 净利率为 9.7%/11.0%/12.9%。自由现金流分别为 11.2/14.4/17.3 亿美元，对应自由现金流利润率 24.6%/25.9%/26.9%。

**表 12: FY24-28 Snowflake 盈利预测 (百万美元)**

经营数据	FY2024	FY2025	FY2026E	FY2027E	FY2028E
营业收入	2,806.5	3,626.4	4,482.9	5,497.7	6,670.9
%yoy	35.86%	29.21%	23.62%	22.64%	21.34%
毛利润	1,907.9	2,411.7	3,015.7	3,762.5	4,776.1
毛利率	67.98%	66.50%	67.05%	68.10%	69.63%
调整后净利润	353.4	301.3	432.5	612.4	862.4
调整后利润率	12.53%	8.31%	9.72%	11.00%	12.93%
FCF	810.2	941.5	1,122.1	1,444.9	1,725.7
FCF 利润率	27.75%	24.38%	24.63%	25.90%	26.86%

资料来源：公司公告，中信建投

## 估值：目前交易于 FY26E 9.8x EV/Rev，略高于同行业可比公司

按照 FY26E，12 x EV/Sales 给予估值，目标价 175 美元/股，同行业可比公司均值为 9.0x，但考虑数仓在数据管理技术栈的核心地位，市场空间及增速较高，确定性较强，给予一定估值溢价。

公司回购价格平均在 130.87 美元/股，但在 FY4Q25 没有进行任何回购。在 FY25 回购了 1480 万股流通股，总回购价为 19 亿美元（不包括与回购相关的交易成本），加权平均价格为每股 130.87 美元<sup>25</sup>。FY4Q25 回购窗口期公司股价基本在 150-190 美元/区间。因此，我们认为公司内部考虑一定安全边际后合理估值在 130 美元/股以上，当前股价略高于内部估值，这一位置具备一定短线支撑。

表 13:可比公司估值（截止 2025/4/19）

公司	股票代码	市值 (USD)	BF P/E	BF EV/EBITDA	BF EV/EBIT	BF EV/Rev	LF P/BV
Snowflake	SNOW US	479 亿	114.51	78.11	112.20	9.79	15.96
当前相对可比公司溢价均值			109.3%	104.4%	145.5%	9.2%	-10.7%
均值 (不含 Snowflake)			54.70	38.22	45.71	8.96	17.88
MongoDB	MDB US	129 亿	57.76	41.10	42.14	4.54	4.61
Datadog	DDOG US	314 亿	48.46	35.60	39.43	8.61	11.49
Cloudflare	NET US	372 亿	120.98	75.78	119.03	16.29	35.57
Zscaler	ZS US	311 亿	57.96	37.84	43.62	9.71	19.36
CrowdStrike	CRWD US	931 亿	102.56	74.71	85.80	17.92	28.39
Rubrik	RBRK US	117 亿	--	--	--	9.34	--
Palo Alto Networks	PANW US	1110 亿	47.29	32.29	35.68	10.26	17.37
Nutanix	NTNX US	164 亿	34.34	26.27	30.09	5.81	--
C3.ai	AI US	26 亿	--	--	--	4.10	2.98
Gen Digital	GEN US	151 亿	10.07	9.43	9.39	5.59	7.00
Fortinet	FTNT US	739 亿	37.64	29.20	31.06	10.10	49.36
Okta Inc	OKTA US	170 亿	29.95	19.92	20.82	5.30	2.65

资料来源: Bloomberg, 中信建投

## 投资评价和建议

总体来看，1) AI 方面，Snowflake 的 AI 主要用于增强非结构化数据的处理能力，从而增强 BI 洞察能力；2) 数仓性能方面，Redshift/Databricks 在较大计算资源投入下具有较好的性能/成本优势，而 Snowflake/BigQuery 在中小型计算资源下具备较好的性能/成本优势。但据 BigQuery 工程师，99% 的数仓查询都小于 10GB，绝大多数企业的数据仓库都小于 1TB，因此聚焦大规模查询场景的优化本质是面向 1% 的头部客户，而忽略剩余 99% 需求，在数仓领域 Snowflake 仍然具备较强的性能和成本表现。3) 大数据平台方面，Databricks 长期致力于构建开放、高性能的 Spark 生态，在开放性和工具/服务全面性上领先 Snowflake，Snowflake 通过 Snowpark /Unistore

<sup>25</sup> 据公司公告，所有回购均在公开市场交易中进行，但不包括 360 万股已发行普通股，这些普通股以 3.996 亿美元的价格从本次发行的票据购买者手中回购，这些票据的购买价格为每股 112.50 美元，是通过与本次票据发行相关的私下协商交易达成的。

等增强对非结构化数据的处理能力，但总体仍然落后于 Databricks。总体来看，我们认为 Snowflake 在数仓领域仍然地位稳固，大数据平台的竞争仍有待观察。按照 FY26E，12x EV/Sales 给予估值，目标价 175 美元/股，维持“买入”评级。

## 风险分析

**市场竞争风险：**云数据仓库市场竞争激烈，Snowflake 面临来自其他云服务提供商和数据仓库解决方案提供商的竞争，包括 AWS、Azure、Google Cloud、Oracle Cloud。竞争可能导致价格竞争、市场份额下降和利润率压力。

**依赖于云服务提供商：**Snowflake 构建在云平台之上，例如 AWS、Azure、GCP。它依赖于这些云服务提供商的基础设施和服务，任何与云服务提供商之间的合作关系变化、竞争、定价调整或服务中断都可能对 Snowflake 的业务和财务状况产生不利影响。

**法律和合规风险：**Snowflake 在全球范围内经营，因此需要应对各个国家和地区的法律、监管和合规要求。

**安全和数据隐私：**作为一家数据仓库提供商，安全和数据隐私是 Snowflake 的重要关注点。任何数据泄露、安全漏洞或违反数据隐私的事件都可能对 Snowflake 的声誉和客户信任产生负面影响。

**技术风险：**Snowflake 的成功依赖于其技术平台和解决方案的稳定性、可靠性和创新性。然而，技术创新和发展也可能带来技术风险，如软件缺陷、系统故障、数据一致性和性能问题。

## 报表预测

### 资产负债表 (百万美元)

会计年度	FY2024A	FY2025A	FY2026E	FY2027E	FY2028E
<b>流动资产</b>	5,039.26	5,869.37	5,539.25	5,546.73	5,647.43
现金	3,846.25	4,637.67	3,689.85	3,278.69	2,895.42
应收票据及应收账款合	926.90	922.81	1,391.61	1,706.62	2,070.80
其他应收款	0.00	0.00	0.00	0.00	0.00
预付账款	180.02	211.23	322.58	395.60	480.02
存货	0.00	0.00	0.00	0.00	0.00
其他流动资产	86.10	97.66	135.20	165.81	201.19
<b>非流动资产</b>	3,184.12	3,164.57	3,071.79	2,981.42	2,892.80
长期投资	916.31	656.48	656.48	656.48	656.48
固定资产	247.46	296.39	249.96	203.52	157.09
无形资产	331.41	278.03	231.69	185.63	139.93
其他非流动资产	1,688.94	1,933.67	1,933.67	1,935.79	1,939.30
<b>资产总计</b>	8,223.38	9,033.94	8,611.05	8,528.15	8,540.23
<b>流动负债</b>	2,731.23	3,301.18	3,703.82	4,366.96	5,204.87
短期借款	0.00	0.00	0.00	0.00	0.00
应付票据及应付账款合	51.72	169.77	113.40	134.64	155.53
其他流动负债	2,679.51	3,131.42	3,590.42	4,232.32	5,049.34
<b>非流动负债</b>	301.56	2,726.11	2,726.11	2,726.11	2,726.11
长期借款	0.00	2,271.53	2,271.53	2,271.53	2,271.53
其他非流动负债	301.56	454.58	454.58	454.58	454.58
<b>负债合计</b>	3,032.79	6,027.30	6,429.93	7,093.07	7,930.98
少数股东权益	10.29	6.71	5.29	4.01	2.58
股本	0.03	0.03	0.03	0.03	0.03
资本公积	9,264.10	10,295.71	10,295.71	10,296.71	10,298.71
留存收益	-4,083.82	-7,295.81	-8,119.92	-8,865.67	-9,692.08
归属母公司股东权益	5,180.31	2,999.93	2,175.82	1,431.07	606.66
<b>负债和股东权益</b>	8,223.38	9,033.94	8,611.05	8,528.15	8,540.23

### 现金流量表 (百万美元)

会计年度	FY2024A	FY2025A	FY2026E	FY2027E	FY2028E
<b>经营活动现金流</b>	848.12	959.76	-936.37	-398.71	-369.83
净利润	-1,033.34	-1,496.78	-825.53	-747.04	-827.83
折旧摊销	194.69	234.96	92.77	93.37	94.62
财务费用	0.00	0.00	0.00	0.00	0.00
其他经营现金流	1,686.78	2,221.59	-203.61	254.96	363.38
<b>投资活动现金流</b>	832.26	190.65	-11.45	-12.45	-13.45
资本支出	-139.33	-228.74	0.00	3.00	6.00
其他投资现金流	971.59	419.38	-11.45	-15.45	-19.45
<b>筹资活动现金流</b>	-854.10	-226.52	0.00	0.00	0.00
短期借款	0.00	0.00	0.00	0.00	0.00
长期借款	0.00	2,271.53	0.00	0.00	0.00
其他筹资现金流	-854.10	-2,498.05	0.00	0.00	0.00
<b>现金净增加额</b>	826.28	923.89	-947.82	-411.16	-383.28

资料来源: 公司公告, iFinD, 中信建投

### 利润表 (百万美元)

会计年度	FY2024A	FY2025A	FY2026E	FY2027E	FY2028E
<b>营业收入</b>	2,806.49	3,626.40	4,482.93	5,497.70	6,670.86
营业成本	898.56	1,214.67	1,477.13	1,753.77	2,025.94
其他营业费用	0.00	0.00	0.00	0.00	0.00
销售和管理费用	1,714.76	2,084.35	2,389.05	2,805.13	3,255.06
研发费用	1,287.95	1,783.38	1,430.83	1,675.40	2,208.23
财务费用	0.00	0.00	0.00	0.00	0.00
其他经营损益	0.00	0.00	0.00	0.00	0.00
投资收益	0.00	0.00	0.00	0.00	0.00
公允价值变动收益	0.00	0.00	0.00	1.00	2.00
<b>营业利润</b>	-1,094.77	-1,456.01	-814.08	-735.59	-816.37
其他非经营损益	44.89	-35.34	-12.67	-12.67	-12.67
<b>利润总额</b>	-1,049.89	-1,491.35	-826.75	-748.26	-829.05
所得税	-16.54	5.43	-1.22	-1.22	-1.22
<b>净利润</b>	-1,033.34	-1,496.78	-825.53	-747.04	-827.83
少数股东损益	-1.89	-3.57	-1.42	-1.29	-1.43
<b>归属母公司净利润</b>	-1,031.45	-1,493.21	-824.11	-745.76	-826.40
<b>EBITDA</b>	-855.20	-1,256.39	-733.98	-654.89	-734.42
<b>EPS (元)</b>	-3.09	-4.47	-2.47	-2.23	-2.47

### 主要财务比率

会计年度	FY2024A	FY2025A	FY2026E	FY2027E	FY2028E
<b>成长能力</b>					
营业收入(%)	35.86	29.21	23.62	22.64	21.34
归属于母公司净利润	-17.99	-42.05	44.56	9.49	-10.80
<b>获利能力</b>					
毛利率(%)	67.98	66.50	67.05	68.10	69.63
净利率(%)	-36.82	-41.27	-18.41	-13.59	-12.41
ROE(%)	-19.91	-49.77	-37.88	-52.11	-136.22
ROIC(%)	-64.63	-89.94	-81.03	-60.72	-83.58
<b>偿债能力</b>					
资产负债率(%)	36.88	66.72	74.67	83.17	92.87
净负债比率(%)	-74.10	-78.70	-65.03	-70.18	-102.40
流动比率	1.85	1.78	1.50	1.27	1.09
速动比率	1.81	1.75	1.46	1.23	1.05
<b>营运能力</b>					
总资产周转率	0.34	0.40	0.52	0.64	0.78
应收账款周转率	3.03	3.93	3.22	3.22	3.22
<b>每股指标 (元)</b>					
每股收益(最新摊薄)	-3.09	-4.47	-2.47	-2.23	-2.47
每股经营现金流(最新)	2.54	2.87	-2.80	-1.19	-1.11
每股净资产(最新摊薄)	15.51	8.98	6.51	4.28	1.82
<b>估值比率</b>					
P/E	-46.46	-32.09	-58.15	-64.26	-57.99
P/B	9.25	15.97	22.02	33.49	78.99
EV/EBITDA	4.15	1.59	1.31	0.68	-0.07

## 分析师介绍

### 崔世峰

海外研究首席分析师，南京大学硕士，7 年买方及卖方复合从业经历，专注于互联网及海外 TMT 龙头公司研究，2021 年加入中信建投，2022-2023 年新财富海外研究最佳研究入围，2019-2020 年新财富传媒最佳研究团队第二名团队成员。

### 许悦

海外研究员，南洋理工大学硕士，专注于港股互联网及美股软件研究，2022 年加入中信建投海外前瞻组，2023 年新浪金麒麟港股及海外市场菁英分析师第二名，2023-24 第十七届、十八届水晶球最佳分析师海外行业入围，2024 年度新财富海外市场研究团队第五名。

## 评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后6个月内的相对市场表现,也即报告发布日后的6个月内公司股价(或行业指数)相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数作为基准;新三板市场以三板成指为基准;香港市场以恒生指数作为基准;美国市场以标普500指数为基准。	股票评级	买入	相对涨幅 15%以上
		增持	相对涨幅 5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅 5%—15%
		卖出	相对跌幅 15%以上
	行业评级	强于大市	相对涨幅 10%以上
		中性	相对涨幅-10-10%之间
		弱于大市	相对跌幅 10%以上

## 分析师声明

本报告署名分析师在此声明：(i) 以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，结论不受任何第三方的授意或影响。(ii) 本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

## 法律主体说明

本报告由中信建投证券股份有限公司及其附属机构(以下合称“中信建投”)制作，由中信建投证券股份有限公司在中华人民共和国(仅为本报告目的，不包括香港、澳门、台湾)提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格，本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

在遵守适用的法律法规情况下，本报告亦可能由中信建投(国际)证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

## 一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础，不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料，但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断，该等观点、评估和预测可能在不发出通知的情况下有所变更，亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件，而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策，中信建投都不对该等投资决策提供任何形式的担保，亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内，中信建投可能持有并交易本报告中所述公司的股份或其他财产权益，也可能在过去12个月、目前或者将来为本报中所提公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点，分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系，分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可，任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分的内容，亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分的内容。版权所有，违者必究。

## 中信建投证券研究发展部

北京  
 东城区朝内大街2号凯恒中心B座12层  
 电话：(8610) 8513-0588  
 联系人：李祉瑶  
 邮箱：lizhiyao@csc.com.cn

上海  
 上海浦东新区浦东南路528号南塔2103室  
 电话：(8621) 6882-1600  
 联系人：翁起帆  
 邮箱：wengqifan@csc.com.cn

深圳  
 福田区福中三路与鹏程一路交汇处广电金融中心35楼  
 电话：(86755) 8252-1369  
 联系人：曹莹  
 邮箱：caoying@csc.com.cn

## 中信建投(国际)

香港  
 中环交易广场2期18楼  
 电话：(852) 3465-5600  
 联系人：刘泓麟  
 邮箱：charleneliu@csci.hk