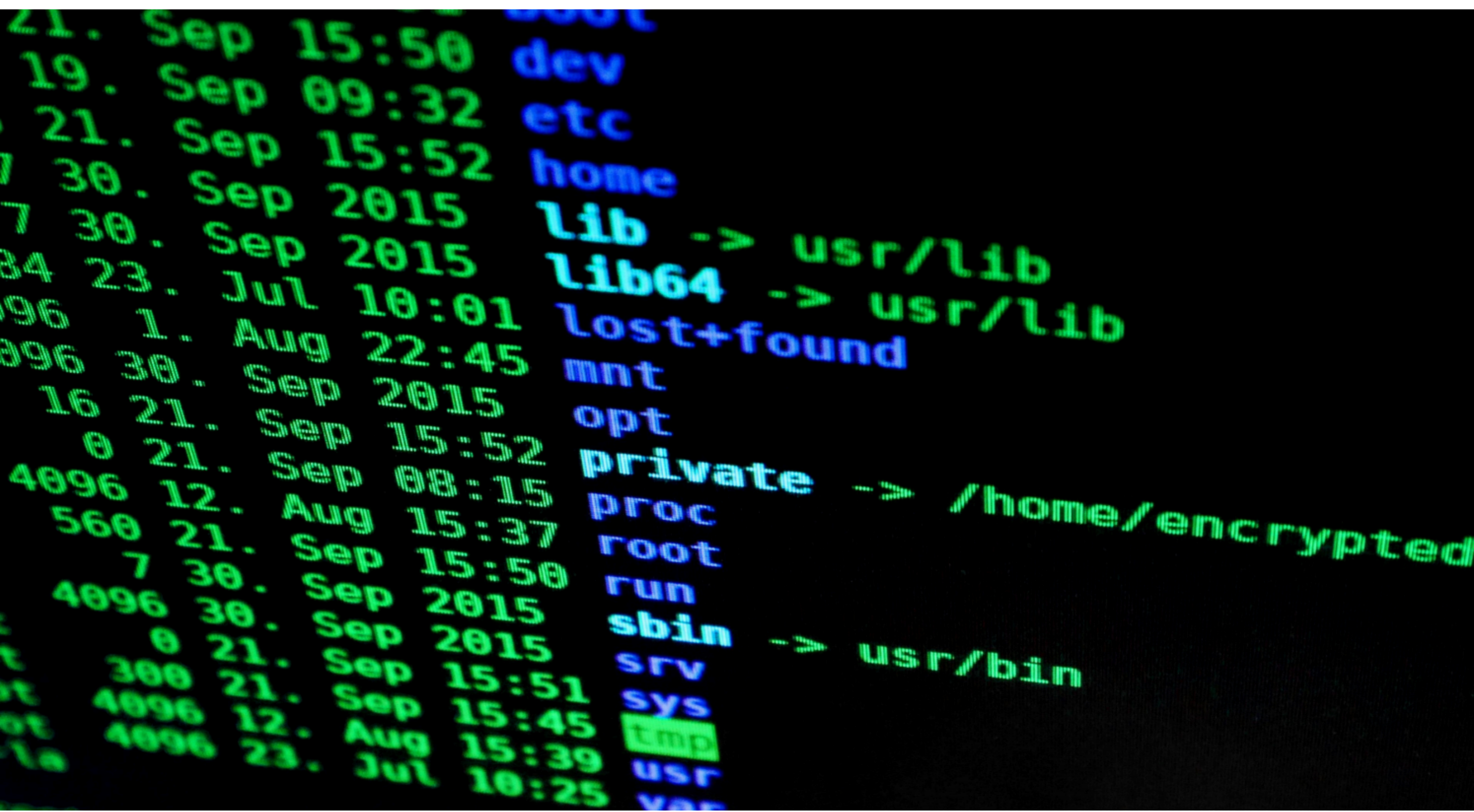


# 对云上分析SQL引擎 进行I/O优化的成本效益

The Cost Effect on I/O Optimization for Analytical SQL on Cloud



\* 本白皮书版权属于北京开元维度科技有限公司，未经授权，不得进行任何形式转载和分发。

# 摘要

---

本文探讨了将数据密集型分析应用从本地迁移到云原生环境这一普遍的行业趋势。我们发现，与云存储相关的独特成本模型要求对性能优化有更细致的了解。具体而言，根据从Uber Presto 生产环境中收集的数据，我们认为在云中简单地应用常见的 I/O 优化，比如table scan 和 filter（表扫描和过滤），以及 broadcast join（广播连接）可能会产生意想不到的成本。这是因为传统的 I/O 优化主要侧重于改善本地环境中的吞吐量或延迟，而没有考虑到与存储API调用相关的财务成本。在云环境中，这些成本可能会非常高昂，仅在Uber实际的使用规模下，Presto工作负载每天就可能涉及数十亿次的API调用。本文将以案例研究的形式呈现 I/O 优化逻辑和思路，可作为读者进一步研究的起点，从而设计出专门针对云环境中数据密集型应用的高效 I/O 策略。



# 目录

---

<b>1 引言</b>	3
<b>2 云成本模型</b>	5
<b>3 I/O 优化对工业分析负载的影响</b>	6
3.1 工业分析流量模式在云 I/O 成本方面的挑战	6
3.2 Table scan 和 filter 优化带来的成本挑战	8
3.3 Broadcast Join 优化的成本挑战	9
<b>4 讨论</b>	11
4.1 提高存储资源效率	11
4.2 创建带缓存的虚拟存储层	12
4.3 具备成本意识，重新设计数据密集型应用	12
<b>5 结论</b>	14
<b>参考文献</b>	15

---

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/876202240123010142>