

引言

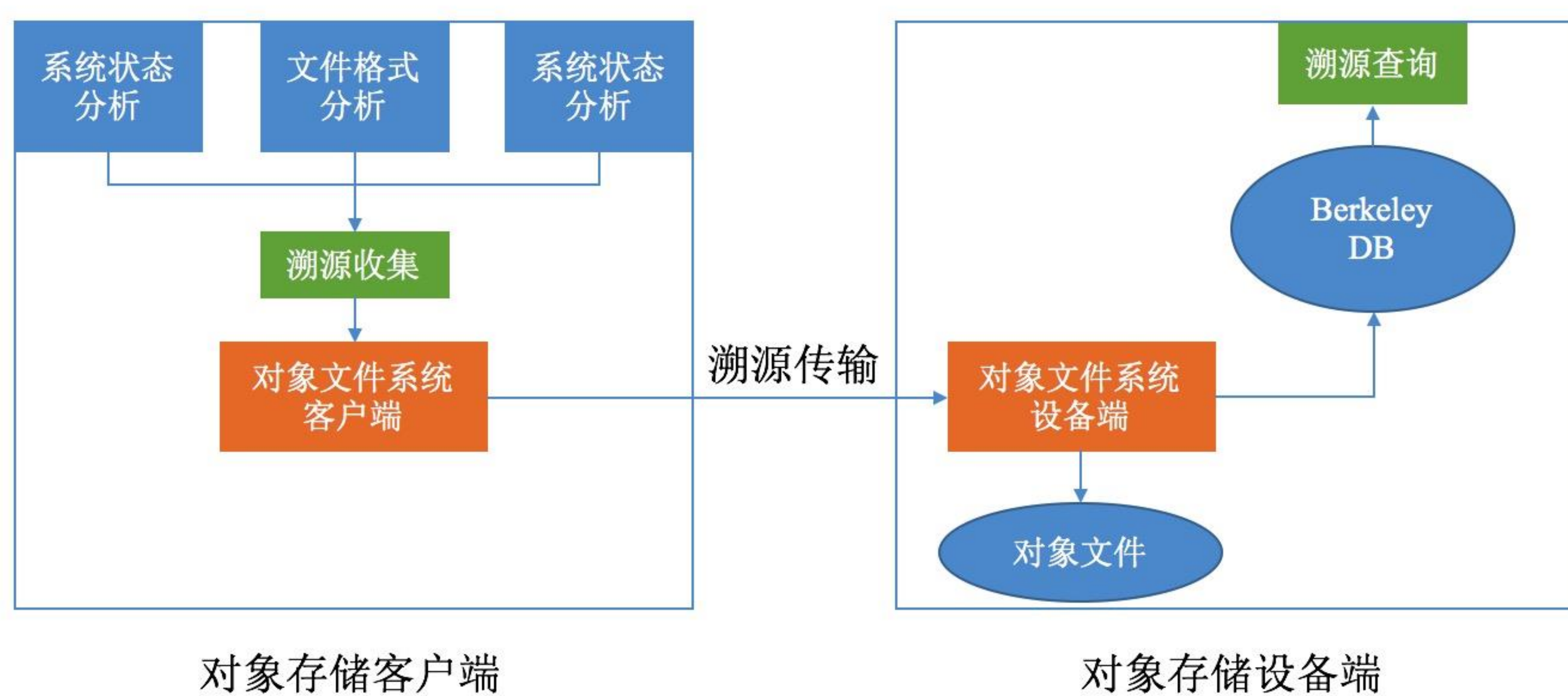
溯源的概念及应用:

- 在系统领域，一个对象的溯源是所有影响该对象最终状态的过程和数据。
- 溯源提高了数据本身所描述的价值。它给出了“对象是如何创建的？它依赖了哪些其他对象？这两个对象的历史操作有何不同”等问题的答案。
- 溯源不仅能用来处理数据，还能够用来管理存储系统。

利用对象存储系统来收集和存储溯源的优势:

- 不同粒度不同大小的溯源信息可以封装为对象信息进行存储。
- 对象存储设备可以对存储的溯源信息进行自我管理，例如，自动的溯源压缩，基于溯源进行数据重建等等。
- 分布式对象存储系统满足了对大容量溯源信息进行高速并发访问的需求。
- 能够利用对象的读写命令来方便地存储和访问溯源信息。

设计与分析



基于对象的溯源存储系统的设计框架

- 利用系统状态文件收集内核信息。
- 调用JHOVE应用程序来分析和封装文件格式。
- 使用Linux系统的审计功能对普通应用程序进行监听。
- 将收集到的溯源信息封装成对象，存储到对象存储设备端Berkeley DB数据库或日志文件中。

测试与评估

溯源收集功能:

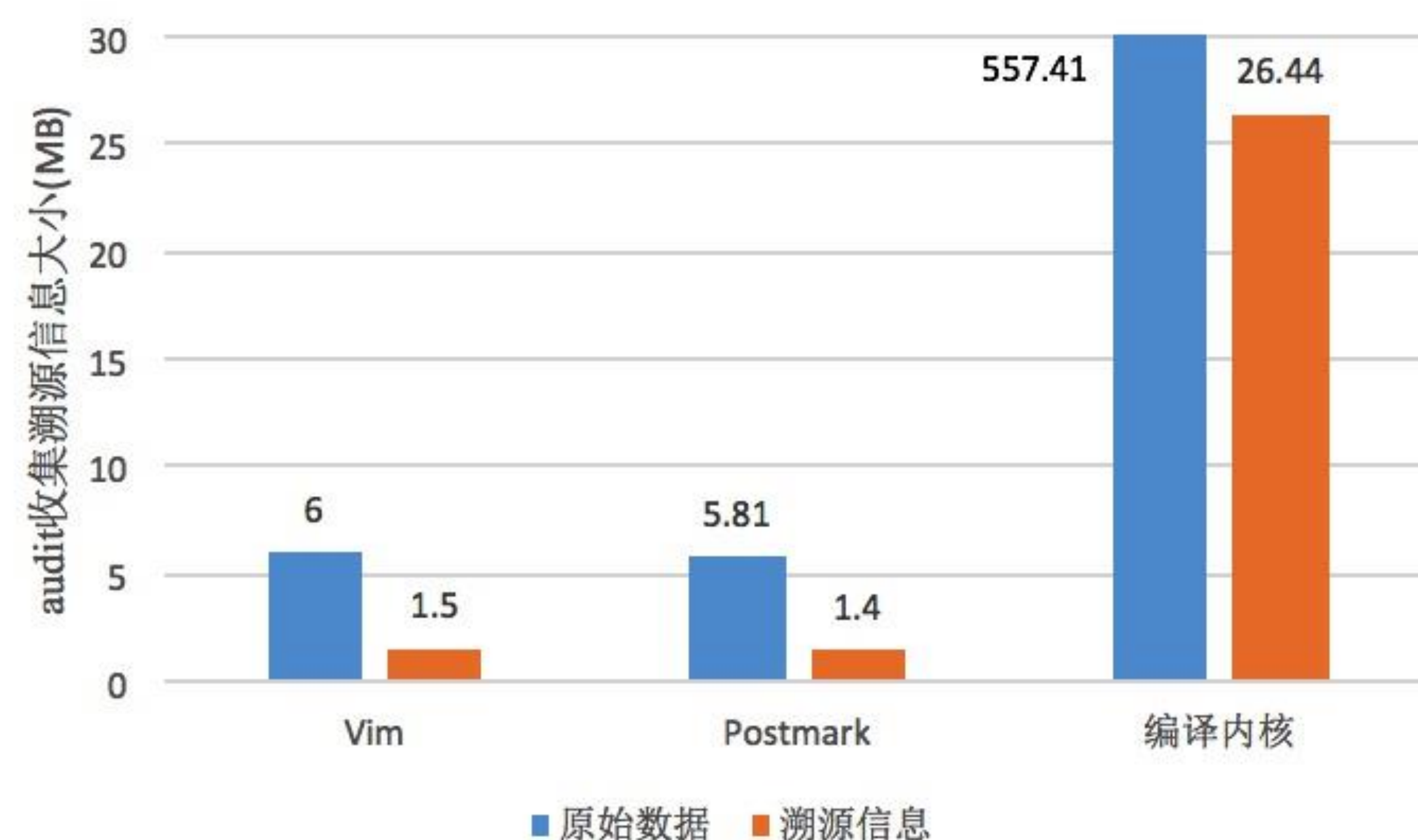
文件格式溯源信息

名称	内容
Format	BYTESTREAM
Release	1.3
Date	2007-4-10
URI	/readresult.txt
LastModified	Tue Apr 12 2016
Size	0
Status	Well-Formed

系统状态内核溯源信息

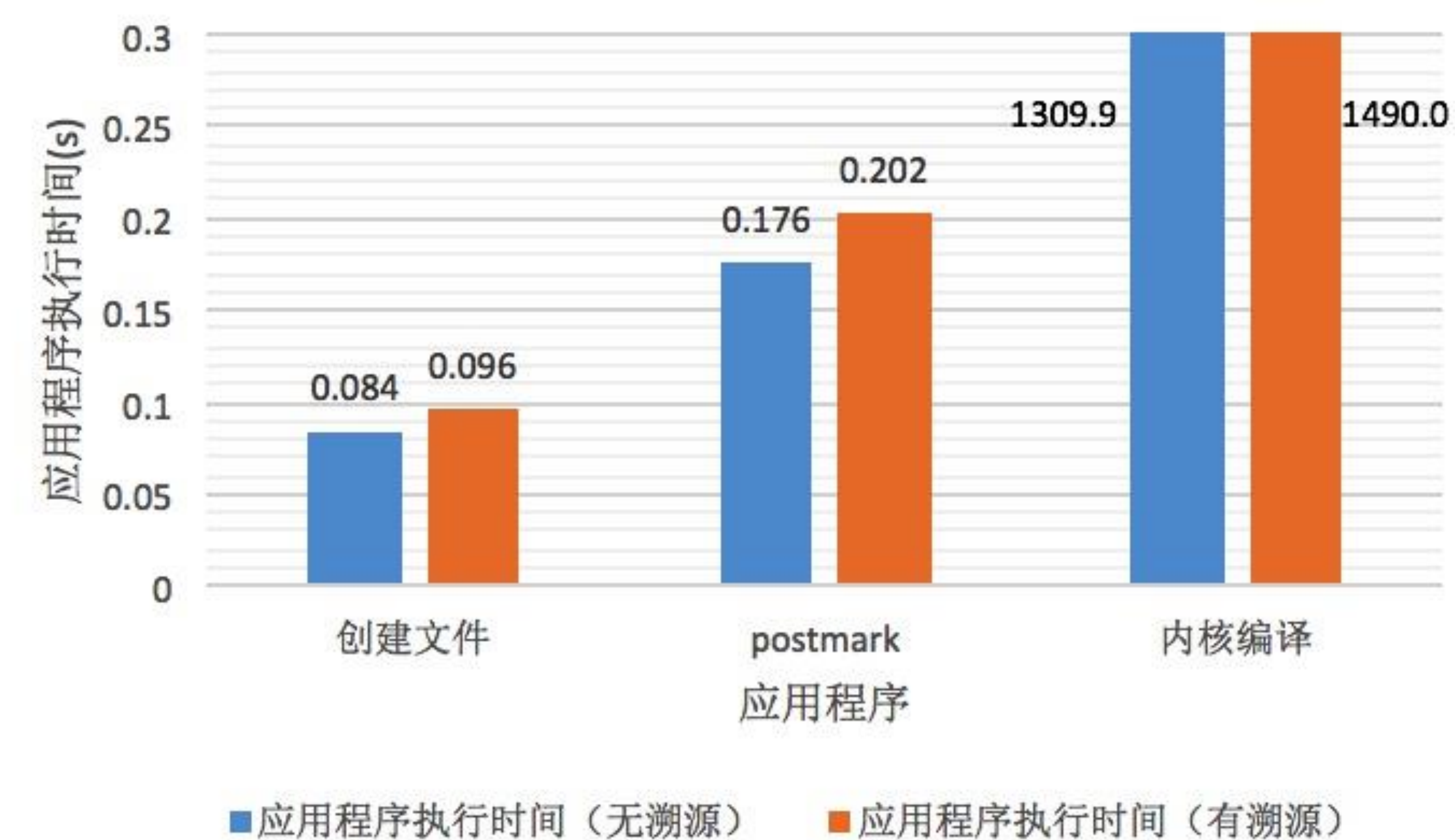
表项	内容
sysname	Linux
nodename	localhost.localdomain
release	2.4.20-8
version	#1 Thu Mar 2003
machine	i686

溯源存储开销:



audit溯源收集方式原始数据大小及生成的溯源信息大小

溯源时间开销:



不同应用程序的溯源时间开销

总结

研究如何在对象存储系统中收集和存储溯源，设计并实现了一个基于对象存储的溯源系统的基本框架。该系统充分利用对象存储体系结构，能够收集系统内核、文件格式及普通应用程序等溯源信息，具有较低的空间和时间开销。