

团体标准

T/CESA XXXX—202X

高性能计算 分布式存储文件性能测试方法

High performance computing-Performance testing methods for distributed storage
file

征求意见稿

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

已授权的专利证明材料为专利证书复印件或扉页，已公开但尚未授权的专利申请证明材料为专利公开通知书复印件或扉页，未公开的专利申请的证明材料为专利申请号和申请日期。

202X-XX- XX 发布

202X-XX- XX 实施

中国电子工业标准化技术协会 发布



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以任何形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前 言..... III

1 范围..... 1

2 规范性引用文件..... 1

3 术语、定义和缩略语..... 1

 3.1 术语和定义..... 1

 3.2 缩略语..... 2

4 测试指标..... 5

 4.1 概述..... 5

 4.2 指标描述..... 5

5 测试原则及测试准备..... 6

 5.1 测试原则要求..... 6

 5.2 测试环境..... 6

 5.3 测试硬件..... 7

 5.4 测试工具..... 7

6 测试方法..... 7

 6.1 元数据性能..... 7

 6.2 数据一致性..... 7

 6.3 带宽性能..... 8

 6.4 单流最大带宽..... 8

 6.5 IOPS 性能及时延..... 9

 6.6 存储集群扩展性..... 9

 6.7 存储集群可靠性..... 10

附录 A（资料性）..... 11

附录 B（资料性）..... 12

附录 C（资料性）..... 13

附录 D（资料性）..... 14

附录 E（资料性）..... 16

附录 F（资料性）..... 17

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由曙光信息产业（北京）有限公司提出。

本文件由中国电子工业标准化技术协会高性能计算机标准工作委员会归口。

本文件起草单位：

本文件主要起草人：



高性能计算 分布式存储文件性能测试方法

1 范围

本文件描述了高性能计算分布式存储文件性能测试方法，包括分布式存储元数据性能、数据一致性、带宽性能、单流最大带宽、IOPS性能及时延、集群扩展性能和可靠性的测试方法。

本文件适用于HPC应用中分布式存储文件的IO性能及集群扩容能力的测试，为用户选型提供可量化的参考依据。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 37737-2019 信息技术 云计算 分布式块存储系统总体技术要求

T/CESA 1312-2024 高性能计算 分布式存储系统技术要求

3 术语、定义和缩略语

3.1 术语和定义

GB/T 37737-2019 界定的以及下列术语和定义适用于本文件。

3.1.1

卷 volume

块存储系统中物理空间的逻辑分区，为虚拟机（VM）或物理机提供裸设备方式进行数据存取。可以进行创建、删除、扩展等操作。

[来源：GB/T 37737-2019，3.1.3]

3.1.2

客户端 client

发起请求并接收响应的设备或软件。

3.1.3

存储池 storage pool

逻辑卷中数据块的分布范围。

[来源：GB/T 37737-2019，3.1.7]

3.2 缩略语

下列缩略语适用于本文件。

- Cache: 缓存内存 (Cache Memory)
- CPU: 中央处理器 (Central Processing Unit)
- HPC: 高性能计算/超算 (High Performance Computing)
- I/O: 输入/输出 (Input/Output)
- IOPS: 每秒的输入输出量 (Input/Output Operations Per Second)
- IP: 网际协议 (Internet Protocol)

4 测试指标

4.1 概述

高性能计算分布式存储文件性能的测试，主要测试以下性能指标：

- a) 针对高性能计算应用中大规模分布式集群所用的元数据服务器，测试元数据性能；
- b) 针对高性能计算应用场景中的高并发和多线程的读写操作时，文件数据一致性的测试；
- c) 针对大文件（如大小1G），测试带宽性能；
- d) 针对高性能计算应用场景下的单客户端的单流最大带宽性能测试；
- e) 针对小文件（如大小8K）测试IOPS性能, 为了保证满负荷压力测试，压力机的数量为分布式存储节点数量的2倍；
- f) 针对分布式存储节点的可扩展性测试线性扩容性能，如10节点的性能理论上是5节点的性能的2倍左右；
- g) 针对分布式存储系统内各节点的关键部件或是节点级别的故障，整体系统的可靠性测试。

4.2 指标描述

高性能计算分布式存储文件性能测试指标描述见表1。

表 1 性能测试指标描述

指标名称	测试目的	测试方法简要描述
元数据性能	元数据节点在高并发情况下的性能表现	模拟大量用户同时访问元数据节点的情况，测试的文件基线为过亿小文件。观察元数据节点的响应时间和吞吐量。
数据一致性	测试高并发场景下数据一致性	在分布式文件系统中创建共享目录，由压力机进行多线程测试模拟高性能计算中多线程高并发情况下分布式文件系统如何保持读写数据的一致性。
带宽性能	大文件的并发吞吐量	生成适合高性能计算场景的测试数据, 设计测试用例来模拟大文件（1GB）的创建、读取、更新和删除等操作。

表 1 性能测试指标描述（续）

指标名称	测量目的	测试方法简要描述
单流最大带宽	测试高性能计算场景下要求的单流带宽测试	在一个节点上，创建一个客户端，创建测试用例，测试单个计算节点和单个客户端的带宽性能。测试这种场景下的最大带宽性能。
IOPS 性能及时延	测试高性能计算下小文件的并发 IO 及时延	生成适合高性能计算场景的测试数据，多个计算节点和存储节点，计算节点配备高速 CPU 和大容量内存，测试小文件的 IOPS 及时延数据，建议以 100 万 IOPS 为测试基线。
存储集群近线性扩展性能	评估分布式存储集群在不同节点数量下的性能表现	设置初始节点数量，并配置分布式存储集群，确保集群正常运行并满足测试需求。逐步增加节点数量，按照一定的规律或步长进行扩容，进行带宽和 IOPS 的性能近线性增长测试。
集群可靠性	测试 HPC 应用中分布式存储集群的可靠性能力，即硬盘的故障、节点的故障和网络故障等情况下存储集群的可靠性	考虑实际的故障情况，如硬盘故障、节点故障和网络故障等。设计测试场景来模拟这些故障情况，系统的数据重构时间测试，且时间越短越好。

5 测试原则及测试准备

5.1 测试原则要求

测试原则符合下列要求：

- a) 高性能计算分布式文件性能测试只针对文件系统做性能测试，测试的系统应满足 T/CESA 1312-2024 的技术要求。
- b) 针对分布式存储的性能测试，宜从 3 个节点起步做文件性能的输出测试。
- c) 读写性能指标符合下列检查方法：
 - 1) 应使用测试工具测试8KB小文件随机读、随机写，记录集群IOPS和平均响应时间；
 - 2) 应使用测试工具测试1GB大文件顺序写、顺序读，记录集群带宽；
 - 3) 测试数据量大于客户端和分布式存储节点内存总和；
 - 4) 宜关闭分布式节点硬盘Cache；
 - 5) 根据需求，宜采用直通模式（Direct IO）即数据直接写入硬盘。

5.2 测试环境

测试环境应在下述条件下进行：

- a) 温度：10℃～35℃；
- b) 相对湿度：35%～80%；
- c) 大气压：86 kPa～106 kPa。

5.3 测试硬件

测试硬件应符合下列要求：

- a) 不限定CPU线程、内存、全闪和混闪等配置要求，但在一个集群下各节点硬件配置一致；
- b) 存储设备集群的节点数为3个节点起步，数据保护采用多副本或纠删码模式；
- c) 测试压力机的物理节点与存储节点比例为2：1；
- d) 前端IO网络与后端存储带宽大于1：1；
- e) 保障测试环境一致性，即根据实际存储节点数量，动态调整，保证以上测试环境一致性；
- f) 进行参数优化，保证最佳测试效果。

5.4 测试工具

本文件中使用的测试工具可参见附录 A。

6 测试方法

6.1 元数据性能

元数据性能测试方法见表 2。

表 2 元数据性能测试方法

序号	条目	要求表述
1	测试目的	高性能计算分布式存储文件系统元数据性能
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本或纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据）； c) 若采用 N+M 纠删码冗余模式，M 应大于等于 2； d) 分布式存储系统（一体机/软件）正常运行。
3	测试过程	a) 评估系统在包并发，高负载下的响应时间和吞吐量，应针对 1 亿个空文件和 1 亿个 8K 文件进行 create、stat、remove 的完整测试； b) 应评估分布式存储集群在创建、查询和删除大量空文件时的性能； c) 应评估集群在创建、查询和删除带有实际数据（8K 大小）的文件时的性能； d) 应记录测试过程中的性能数据，包括响应时间、吞吐量等。
4	预期结果	获得关于集群在处理带有实际数据的大量文件时的元数据性能数据
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等。 b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图。
6	备注	测试脚本可参见附录 B。

6.2 数据一致性

数据一致性测试方法见表 3。

表 3 数据一致性测试方法

序号	条目	要求表述
1	测试目的	测试高性能计算分布式存储文件系统高并发场景下的数据一致性
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本或纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据）； c) 若采用 N+M 纠删码冗余模式，M 应大于等于 2； d) 分布式存储系统（一体机/软件）应正常运行。
3	测试过程	a) 在分布式文件系统中创建共享目录 test，挂载到目录/mnt/test 下，应按照 2: 1 比例，准备 6 台压力机，每台压力机起 10 个进程，同时对/mnt/test/test_file 文件进行写操作； b) 有进程写完后，应对文件进行数据校验。
4	预期结果	获取数据的结果，并和元数据对比看是否一致
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等； b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图； c) 更新测试步骤中的参数或已设定的配置。
6	备注	测试脚本可参见附录 C。

6.3 带宽性能

带宽性能测试方法见表 4。

表 4 带宽性能测试方法

序号	条目	要求表述
1	测试目的	测试带宽指标性能（建议使用 IOR 工具）
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据）； c) 若采用 N+M 纠删码冗余模式，M 应大于等于 2； d) 分布式存储系统（一体机/软件）正常运行。
3	测试过程	按等比例测试，client 数量设定，在测试客户端，保证物理节点机的带宽负载满负荷运行。部署测试工具(如 IOR)，设置负载参数，对挂载共享目录应分别进行以下测试： a) 使客户端缓存/存储端缓存写满，确保数据写入 SSD，并保证写入硬盘，SSD 数据预写满，占容量 80%，而后开始记录测试数据； b) 对若干个 1GB（预埋数据，文件类型）文件（文件总容量超过存储节点所有内存之和的 2 倍），1M 块大小，进行顺序写测试，记录集群写带宽； c) 若干个 1GB 文件（文件总容量超过存储节点所有内存之和的 2 倍），1M 块大小，进行顺序读测试，记录集群读带宽。
4	预期结果	测试结果按照不同的测试场景以及各项指标参数以表格形式呈现。
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等； b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图； c) 更新测试步骤中的参数或已设定的配置。
6	备注	测试脚本可参见附录 D。

6.4 单流最大带宽

单流最大带宽测试方法见表 5。

表 5 单流最大带宽测试方法

序号	条目	要求表述
1	测试目的	测试单流最大带宽
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据）； c) 若采用 N+M 纠删码冗余模式，M 应大于等于 2； d) 分布式存储系统（一体机/软件）正常运行。
3	测试过程	在测试计算节点机启动一个客户端，部署测试工具，设置负载参数，对挂载共享目录进行单流最大带宽测试。
4	预期结果	测试结果按照不同的测试场景以及各项指标参数以表格形式呈现。
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等； b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图； c) 更新测试步骤中的参数或已设定的配置。
6	备注	测试脚本可参见附录 E。

6.5 IOPS 性能及时延

IOPS 性能及时延测试方法见表 6。

表 6 IOPS 性能及时延测试方法

序号	条目	要求表述
1	测试目的	测试 IOPS 及时延指标性能
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本或纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据）； c) 若采用 N+M 纠删码冗余模式，M 应大于等于 2； d) 分布式存储系统（一体机/软件）正常运行。
3	测试过程	在测试客户端部署测试工具，设置负载参数，对挂载共享目录应分别进行以下测试： a) 6000 万 8K 文件，8K 块大小，进行随机 70%读 30%写测试； b) 预热时间 5 分钟，测试时间 30 分钟，采用 directIO 模式，不使用客户端缓存，记录集群真实 IOPS 及时延性能。
4	预期结果	在高性能计算应用场景下，预计 IOPS 性能在 100 万基线之上，时延在毫秒级别。 测试结果按照不同的测试场景以及各项指标参数以表格形式呈现。
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等； b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图； c) 更新测试步骤中的参数或已设定的配置。
6	备注	测试用例及测试脚本请参考附录 F。

6.6 存储集群扩展性

存储集群扩展性测试方法见表 7。

表 7 存储集群扩展性测试方法

序号	条目	表述
1	测试目的	测试存储集群扩容及集群性能拟线性性能
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本或纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据），若采用 N+M 纠删码冗余模式，M 应大于等于 2； c) 分布式存储系统（一体机/软件）集群环境搭建完成，处于正常运行状态； d) 采用第三方性能工具测试性能，监控 IOPS 和带宽性能，性能值保持稳定。
3	测试过程	a) 上架一台即将加入存储集群的存储物理节点，其物理状态及相应的 IP 地址可正常加入原集群； b) 点击存储集群管理页面，将上述存储节点加入集群，算节点数量同比例增加； c) 新存储节点加入集群后，集群性能短期内会有波动，待波动结束后记录稳定性能的 IOPS 和带宽性能； d) 重复以上的测试，比较不同节点数量下的性能数据表现。
4	预期结果	测试结果按照不同的测试场景以表格形式呈现性能线性增长情况。
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等； b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图； c) 更新测试步骤中的参数或已设定的配置。
6	备注	无

6.7 存储集群可靠性

存储集群可靠性测试方法见表 8。

表 8 存储集群可靠性测试方法

序号	条目	表述
1	测试目的	测试存储集群可靠性
2	预置条件	a) 分布式存储设备集群的节点数应大于等于 3，数据保护采用多副本或纠删码； b) 若采用多副本冗余模式，副本数应大于等于 2（含原始数据）； c) 若采用 N+M 纠删码冗余模式，M 应大于等于 2； d) 分布式存储系统（一体机/软件）正常运行。
3	测试过程	a) 硬盘故障测试：例如模拟硬盘损坏或失效。可以通过模拟工具或手动操作来模拟硬盘故障，观察集群的反应和容错能力。可以分别在各个节点停止 1 块盘，而后第 2 块盘轮询方式，记录集群的恢复时间和数据完整性等指标，测试过程中保持 IO 不停止。 b) 节点故障测试：模拟节点故障场景，例如模拟节点宕机或断电。可以通过停止节点的运行或模拟节点失效，观察集群的自动化故障转移和恢复能力。（多长时间稳定 IO 或性能） c) 网络故障测试：模拟网络故障场景，模拟网络中断或高延迟。可以通过断开网络连接或配置网络模拟工具来模拟网络故障，观察集群的容错和恢复能力。记录集群的恢复时间和数据完整性等指标。
4	预期结果	测试结果按照不同的测试场景以及各项指标参数以表格形式呈现。
5	测试结果	a) 环境配置应包括规格及详细配置、编译器类型及版本、CPU 型号、内存频率、操作系统版本、测试工具的来源及用途、测试工具的版本型号等； b) 应按照测试步骤测试验证，进行操作截图（形成完整的测试逻辑链），并辅以必要的文字说明，截图和截图之间应有简单的文字说明，不能出现连续的几张截图； c) 更新测试步骤中的参数或已设定的配置。
6	备注	无

附录A
(资料性)
测试工具参考

在进行高性能计算分布式存储文件性能测试的时候，建议的测试工具见表A. 1。

表 A. 1 测试工具表

工具名称	工具简介
MDtest	MDTest 是一个在并行计算机系统上进行性能测试的工具，主要用于评估并行计算机系统的性能和可靠性。MDTest 是由美国橡树岭国家实验室 (Oak Ridge National Laboratory) 开发的一种基准测试程序，是 HPC Challenge 基准测试的一部分。
IOR (IO Reference)	IOR 是一个用于评估并行计算机系统 I/O 性能的基准测试工具。它是由美国橡树岭国家实验室 (Oak Ridge National Laboratory) 开发，并且是 HPC Challenge 基准测试套件的一部分。IOR 专门设计用来测量大规模并行系统上的文件 I/O 性能，尤其是在高并行度下的性能表现。IOR 测试通常包括对不同大小文件的读写操作，以及不同的 I/O 模式（如顺序 I/O 和随机 I/O）。它可以模拟真实世界中的科学计算应用的 I/O 模式，并且支持多种并行文件系统，包括 POSIX 文件系统、的对象存储系统等。通过 IOR 测试，用户可以得到关于并行计算机系统在实际工作负载下的 I/O 性能指标，这些指标对于评估系统的能力、优化 I/O 配置以及提高大型计算应用的总体性能都是非常重要的。
VDbench	VDbench 是一款由 Sun Microsystems（现为 Oracle 公司）开发的通用存储性能测试工具。VDbench 提供了一个模拟环境，可以在不同的工作负载模式下测试存储系统的读写性能、延迟、吞吐量和其他关键性能指标。VDbench 支持多种工作负载模式，包括基本的顺序访问、随机访问以及混合模式，可以配置多个客户端并发执行测试，模拟大规模用户环境下的存储访问，并能提供详细的性能报告，包括响应时间、吞吐量、IOPS（每秒操作次数）等关键指标。

附录B
(资料性)
元数据性能测试脚本

元数据性能测试脚本见表B. 1。

表 B. 1 元数据性能测试脚本

测试脚本	元数据性能测试（测试工具推荐 MDtest）
测试代码	<pre>测试脚本 1. 对 1 亿个空文件进行 create、stat、remove 测试 #!/bin/bash # 设置测试参数 TEST_DIR="./testdir/" NUM_FILES=100000000 CLIENTS=100 HOSTFILE="./hostfile_list" # 运行 MDtest 测试 mpirun -np \$CLIENTS -hostfile \$HOSTFILE ./mdtest -C -d \$TEST_DIR -F -L -z 4 - b 10 -I \$NUM_FILES # 记录性能数据 # 可以在此添加性能数据收集的代码，比如将 MDtest 的输出重定向到日志文件进行分 析。 2. 对 1 亿个 8K 文件进行 create、stat、remove 测试 #!/bin/bash # 设置测试参数 TEST_DIR="./testdir/" NUM_FILES=100000000 CLIENTS=100 HOSTFILE="./hostfile_list" FILE_SIZE=8192 # 运行 MDtest 测试 mpirun -np \$CLIENTS -hostfile \$HOSTFILE ./mdtest -C -d \$TEST_DIR -F -L -w \$FILE_SIZE -z 4 -b 10 -I \$NUM_FILES # 记录性能数据 # 可以在此添加性能数据收集的代码，比如将 MDtest 的输出重定向到日志文件进行分 析。</pre>

附录C
(资料性)
数据一致性测试脚本

数据一致性测试脚本见表C. 1。

表C. 1 数据一致性测试脚本

测试脚本	数据一致性测试
测试代码	在分布式文件系统中创建共享目录 test，挂载到目录/mnt/test 下，按照 2: 1 比例，准备 6 台压力机，每台压力机起 10 个进程，同时对/mnt/test/test_file 文件进行写操作。这 60 个进程编号为 1~60，具体写的方式如下： a) 每个进程一次写 8KB 数据（块大小 8K） b) 每个进程写的区间不重叠，进程 1 写文件的 (0~8K-1)，进程 2 写 (8K, 16K-1)，进程 3 写 (16K, 24K-1)……，进程 60 写 (472K, 480K-1)，进程 1 写 (480K, 488K-1)……。每个进程写 100 个块，文件总大小为 48000K。 c) 每个块写入的数据不同。 d) 保证 60 个进程同时对文件进行写操作。 e) 以第一个进程的流程举例，其他进程的流程相同仅写的位置不同： i. open 文件 ii. seek(0, 0) iii. write(8K 数据) iv. seek(480K, 0) v. write(8K 数据) vi. …（写 100 次） vii. close 文件 所有进程写完后，对文件进行数据校验。 测试用例可参考如下测试工具进行测试，基于 mpirun 写的 python 脚本： mpirun -hostfile nd(配置 文件) -np 60(所有 节点 进程 数 之 和) --mca btl_tcp_if_include eno2(网卡) --allow-run-as-root python mpi_ntol_fileio.py -b 8k(块大小) -n 100(块数目) -d /mnt/test/(测试目录) -f 1(测试文件的个数) -l 1(循环次数)

附录 D
(资料性)
带宽性能测试脚本

带宽性能测试脚本见表D. 1。

表 D. 1 带宽性能测试脚本

测试脚本	带宽性能测试，推荐测试工具 IOR
测试代码	<pre>#!/bin/bash # 创建预埋数据的文件 FILE_SIZE=1G BLOCK_SIZE=1M TOTAL_FILES=\$(((\$ (free -m grep Mem awk '{print \$2}') * 2 / 1024)) # 假设 free 命令 输出的的是 MB，转换为 GB TEST_DIR="./testdir" mkdir -p \$TEST_DIR # 创建足够多的 1GB 文件 for ((i=1; i<=TOTAL_FILES; i++)); do dd if=/dev/zero of=\$TEST_DIR/testfile_\$i bs=\$BLOCK_SIZE count=\$((FILE_SIZE/\$BLOCK_SIZE)) done echo "预埋数据文件创建完成" 顺序写测试脚本。 #!/bin/bash # 设置预热时间，客户端缓存/存储端缓存，时间拉长内存写入硬盘等参数 WARMUP_TIME=600 # 预热时间 600 秒 CACHE_POLICY="" # 客户端和存储端缓存策略，根据具体存储系统来设置 # IOR 测试参数 # -w 表示写测试 # -k 表示使用块 I/O # -b 10g 表示块大小为 10GB（注意这里与文件块大小不同，是指每次 I/O 操作的大小） # -F 表示使用文件 I/O 而不是内存映射 # -t 1m 表示线程数为 100（这里应该是 -t 100，1m 可能是一个错误） # -o ./testdir 表示输出目录 # 执行预热 echo "开始预热..." mpirun -np 100 -hostfile ./hostfile_list ./IOR -w -k -b 10g -F -t 100 -o ./testdir &> /dev/null sleep \$WARMUP_TIME # 执行顺序写测试 echo "开始顺序写测试..." mpirun -np 100 -hostfile ./hostfile_list ./IOR -w -k -b 10g -F -t 100 -o ./testdir > write_bandwidth.log 2>&1</pre>

表 D.1 （续）

	<pre>echo "顺序写测试完成，结果保存在 write_bandwidth.log" 顺序读测试脚本： #!/bin/bash # 假设 SSD 已经预写满，这里不需要额外的预热步骤 # IOR 测试参数 # -r 表示读测试 # 执行顺序读测试 echo "开始顺序读测试..." mpirun -np 100 -hostfile ./hostfile_list ./IOR -r -k -b 10g -F -t 100 -o ./testdir > read_bandwidth.log 2>&1 echo "顺序读测试完成，结果保存在 read_bandwidth.log"</pre>
--	--



附录E
(资料性)

单流最大带宽性能测试脚本

单流最大带宽性能测试脚本见表 E.1。

表 E. 1 带宽性能测试脚本

测试脚本	单流最大带宽性能测试，推荐 IOR
测试代码	在测试计算节点机启一个客户端，部署测试工具，设置负载参数，对挂载共享目录分别进行以下测试， 测试单流最大带宽 <pre># 执行单流顺序写测试 echo "开始单流顺序写测试..." mpirun -np 1 -hostfile ./hostfile_list ./IOR -w -k -b 1024g -F -t 1 -o ./testdir > single_write_bandwidth.log 2>&1 echo "单流顺序写测试完成，结果保存在 single_write_bandwidth.log" # 执行单流顺序读测试 echo "开始单流顺序读测试..." mpirun -np 1 -hostfile ./hostfile_list ./IOR -r -k -b 1024g -F -t 1 -o ./testdir > single_read_bandwidth.log 2>&1 echo "单流顺序读测试完成，结果保存在 single_read_bandwidth.log"</pre>

附录F
(资料性)

IOPS及时延指标性能测试脚本

IOPS 及时延性能测试脚本见表 E.1。

表 E. 1 IOPS 及时延性能测试脚本

测试脚本	IOPS 及时延指标性能测试，推荐使用 VDbench 工具
测试代码	<pre># VDbench 测试脚本 # 禁用消息扫描 messagescan=no # 定义硬件资源 hd=default, vdbench=/path/to/vdbench, shell=ssh, user=your_username # 假设我们有 12 个存储节点，每个节点具有独立的 IP 地址 hd=h1, system=192.168.101.1 hd=h2, system=192.168.101.2 hd=h3, system=192.168.101.3 hd=h4, system=192.168.101.4 hd=h5, system=192.168.101.5 hd=h6, system=192.168.101.6 hd=h7, system=192.168.101.7 hd=h8, system=192.168.101.8 hd=h9, system=192.168.101.9 hd=h10, system=192.168.101.10 hd=h11, system=192.168.101.11 hd=h12, system=192.168.101.12 # 定义文件系统参数 fsd=fsd1, anchor=/path/to/shared/directory, depth=2, width=100, files=60000000, size=8k, shared=yes # 定义格式化工作负载 fwd=format, threads=64, xfersize=8k # 定义默认的随机 I/O 工作负载 fwd=default, xfersize=8k, fileio=random, fileselect=random, rdpct=70, wrpct=30, threads=64, openFlags=o_direct # 映射工作负载到文件系统和数据节点 fwd=fwd1, fsd=fsd1, hd=h1 fwd=fwd2, fsd=fsd1, hd=h2 fwd=fwd3, fsd=fsd1, hd=h3 fwd=fwd4, fsd=fsd1, hd=h4 fwd=fwd5, fsd=fsd1, hd=h5 fwd=fwd6, fsd=fsd1, hd=h6 fwd=fwd7, fsd=fsd1, hd=h7 fwd=fwd8, fsd=fsd1, hd=h8 fwd=fwd9, fsd=fsd1, hd=h9 fwd=fwd10, fsd=fsd1, hd=h10 fwd=fwd11, fsd=fsd1, hd=h11 fwd=fwd12, fsd=fsd1, hd=h12 # 定义读取工作负载，预热和正式测试 rd=rd1, fwd=fwd*, fwdrate=max, format=norestart, elapsed=1800, interval=5, warmup=300</pre>

参考文献

- [1] T/CCSA 263—2019 分布式块存储总体技术要求
- [2] T/CCSA 325—2021 数据中心存储能效测评规范

