

团体标准《高性能计算 分布式存储文件性能测试方法》 (征求意见稿) 编制说明

一、标准编制情况

1. 工作简况

2023年6月25日,中国电子工业标准化技术协会下达了2023年第八批团体标准制修订项目的通知(中电标通(2023)016号)。根据该通知,团体标准制订计划《高性能计算 分布式存储文件性能测试方法》正式下达,项目计划号为CESA-2023-064。该标准项目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,牵头单位为曙光信息产业(北京)有限公司。

2、主要工作过程

(一) 标准预研

近来高性能计算的概念在技术和应用方面有一些进展和演变。高性能计算正逐渐与大数据处理技术相结合,高性能计算中不仅关注计算能力,还要注重数据并发存储能力,在复杂的数据密集型应用和科学研究中,针对文件存储和调用的性能非常重要。

另外,人工智能和机器学习等领域的快速发展对计算资源提出了更高的要求,而大规模训练和推理任务所需的数据,都是以文件系统的方式来保存和调用。

2023年1月-3月,曙光信息产业(北京)有限公司研究分析了分布式存储文件性能在高性能计算中的重要因素,并初步确定了分布式存储文件性能测试方法的标准草案结构和主要内容。通过组织3次标准讨论会,对标准草案结构、内容,以及标准制定的必要性、可行性、拟解决的问题进行了充分讨论,形成一致意见。

2023年4月,曙光信息产业(北京)有限公司向分布式存储厂商、存储产品厂商、行业用户及科研院所等发起参编邀请,各意向参编单位对于该标准制定的必要性、可行性、拟解决的问题表示一致认同并共同推进标准立项。

(二) 标准立项

2023年5月,曙光信息产业(北京)有限公司作为发起方申请立项,并于2023年6月通过立项评审,成为协会正式标准制定项目。

（三）标准编制

2023年6月20日，标准工作组组织召开了标准启动会（现场会议方式），对标准制定的背景、参编单位及参编人员、标准推动计划进行了介绍，并就标准框架展开讨论，在此基础上确定了草案进一步编写完善的分工计划。

2023年7月14日，标准工作组组织召开第二次讨论会（现场会议方式），对标准结构进行完善，对标准技术要求范围进行明确，并确定了牵头汇总各章节内容的分工计划。

2023年8月9日，标准工作组组织召开第三次讨论会（现场会议方式），重点对高性能计算分布式存储文件性能测试的关注点进行了相关讨论。

2023年8月18日，标准工作组组织召开第四次讨论会（现场会议方式），重点对高性能计算分布式存储文件性能测试的主要测试项章节展开讨论，并进一步修改完善。

2023年9月8日，标准工作组组织召开第五次会议（现场会议方式），对标准逐条讨论、修改，就测试的大纲，并就编制说明内容进行介绍和修改完善。

2024年1月19日，标准工作组组织召开第六次会议（现场会议方式），对标准逐条讨论、修改，就测试的具体性能细项定稿，并就编制说明内容进行介绍和修改完善。

2024年3月5日，标准工作组组织召开第七次会议（线上会议方式），对标准逐条讨论、修改，就测试的细项及测试用例进行了深入讨论，并就编制说明内容进行介绍和修改完善。

2024年3月22日，标准工作组组织召开第八次会议（现场会议方式），对标准逐条讨论、修改，就测试的细项定稿，并重新做了调整，将测试用例及脚本文件放入参考附录。

3、主要起草人及其所做的工作

曙光信息产业（北京）有限公司、中移动信息技术有限公司，北京星辰天合科技有限公司，中电云计算技术有限公司，合肥大唐存储科技有限公司，北京得瑞领新科技有限公司，国家超级计算郑州中心，荣联科技集团股份有限公司，北京忆恒创源科技股份有限公司，中研益企（北京）信息技术研究院有限公司，北京焱融科技有限公司，中国科学院高能物理研究所，工业和信息化部电子第五研

究所、融科联创（天津）信息技术有限公司，中国电子科技集团公司第52研究所，北京尊冠科技有限公司、西安奥卡云数据科技有限公司、极道科技（北京）有限公司、北京青云科技股份有限公司、西北农林科技大学、泸州老窖股份有限公司、中移(苏州)软件技术有限公司等参与了技术细节讨论，提供了修改意见和建议。

二、标准编制原则和确定主要内容的论据及解决的主要问题

1、编制原则

在标准编制过程中，遵循了以下四方面的原则。

a) 规范性。本标准的制定完全遵守国家法律、法规等相关规定，制定过程严格按照程序执行，且格式符合标准编写要求。

b) 适用性。本标准的制定充分考虑了分布式存储技术在高性能计算场景中的实际应用需求，既保证分布式存储技术的先进性，又与高性能计算数据中心建设相关规范接轨，条款具备可操作性。

c) 中立性。本文件制定过程中编制组成员单位对标准文本进行了充分讨论，达成一致意见

d) 科学性。本文件技术指标科学合理并可验证，试验方法具有落地指导性，并提供实际的测试用例和脚本文件。

2、确定主要内容的论据

高性能计算中，数据的存储与调用必须与数据的计算能力相匹配，这就要求分布式存储系统的文件性能要与算力在性能上相匹配，并能随着高性能计算集群规模的提高，分布式存储的文件性能也需要动态的提高。

本标准给出了高性能计算分布式存储文件性能的测试方法，涵盖了针对高性能实际应用场景下，针对元数据性能测试，带宽性能，IOPS和时延测试，单流最大带宽测试，及动态扩容下拟线性性能测试及可靠性测试等具体测试方法和实际测试用例及脚本。

本文件可适用于高性能计算分布式存储文件性能测试的通用参考标准，也可以在特定应用中，对其中的关键项进行独立的测试参考。

3、编制过程中解决的主要问题

针对高性能计算分布式存储的文件性能国内还没有一个标准，所以在行业需要高密集，高并发的需求中没有明确的参考指标，有必要制定高性能计算分

布式存储系统文件性能测试的标准，填补高性能计算领域，存储文件性能测试的标准，促进高性能计算产业链上下游技术和产品的融合，推动建立成熟完善的存储生态系统。

三、主要试验[或验证]情况分析

为在标准起草过程中，标准工作组对标准规定的高性能计算分布式存储文件性能测试方法主要征集了存储厂商、服务器厂商、分布式存储行业用户、相关领域科研院所等的意见，并充分考虑科技发展及应用。

本文件规定的测试方法是合理的，并且能切实落地，给出符合产业实际需要的测试用例和脚本文件。

四、知识产权情况说明

本标准不涉及知识产权情况。

五、产业化情况、推广应用论证和预期达到的经济效果

标准发布后，牵头及参编单位将在国内外重量级会议进行推广。并协同国内权威测试机构开始进行实际的落地评测，对致力于高性能设计的分布式存储厂商，软件厂商可参考本标准进行文件性能测试。

六、转化国际标准和国外先进标准情况

本文件未采用国际标准和国外先进标准。

七、与现行相关法律、法规、规章及相关标准的协调性

本标准与国内行业相关标准没有冲突。

本标准项目属于高性能计算机标准体系中“存储”领域，存储是高性能计算机标准体系的重要组成部分，本标准将为高性能计算体系整体架构提供技术支撑。

八、重大分歧意见的处理经过和依据

无。

九、贯彻标准的要求和措施建议

建议列为推荐性标准，在标准发布后尽快组织标准宣贯、试验验证。

十、替代或废止现行相关标准的建议

无需要替代或废止的现行相关标准。

十一、其它应予说明的事项

无。

《高性能计算分布式存储文件性能测试方法》

团体标准编制起草组

2024-05-09