

Linux集群的存储和共享文件系统安装 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/253/2021\\_2022\\_Linux\\_E9\\_9B\\_86\\_E7\\_BE\\_c103\\_253028.htm](https://www.100test.com/kao_ti2020/253/2021_2022_Linux_E9_9B_86_E7_BE_c103_253028.htm) 简介 本文是系列文章中的第 3 部分，该系列涵盖了安装和设置一个大型 Linux 计算机集群的全部内容。有关如何从多个独立的硬件和软件创建一个可运行的 Linux(R) 集群的资源很多，本系列力求把这些公共领域里的资源中的最新信息集中起来以供您参考。本系列文章无意为您提供有关如何设计一个新的大型 Linux 集群的基础知识；要获得一般性的架构方面的知识，请参考本文中提到的相关参考资料和 Redbooks#8482。本系列是针对系统架构师和系统工程师使用 IBM eServer Cluster 1350 框架（关于该框架的更多信息，请参见参考资料）计划和实现 Linux 集群而设计的。其中的一部分内容可能会与集群管理员的日常集群操作有关。本文的每个章节都针对相同的示例安装。本系列的第 1 部分详细阐述了设置集群硬件的操作指导。第 2 部分带您亲历了硬件配置的后续步骤：安装软件，这其中会用到 IBM systems 管理软件、Cluster Systems Management (CSM) 和节点安装。第 3 部分是两篇有关集群后端存储的文章中的上部，这两篇文章合在一起，全面介绍了存储的硬件配置以及 IBM 共享文件系统 GPFS (General Parallel File System) 的安装和配置。本部分将侧重于讨论存储系统的架构、硬件准备以及存储区域网络 (Storage Area Network, SAN) 的设置。本系列的第 4 部分即最后一部分则提供了有关特定于此示例集群存储后端的 CSM 的详细信息，尤其是如何执行存储系统的节点安装以及 GPFS 的集群配置。存储架构在继续之前，先阅读

一下本系列第 1 部分中有关总体架构的章节将会很有益处。图 1 给出了本系列中描述的示例集群的存储配置概览。此配置在本文中将进一步加以解释。该设置基于 GPFS 版本 2.3。包括一个大型的 GPFS 集群，该集群分成两个逻辑部分，共用一个大型文件系统。示例设计还提供了应付灾难的恢复能力，如果存储后端的一半丢失，另一半仍可继续操作。图 1. 存储架构概览图 1 显示了四个存储服务器，可管理由两个磁盘子系统提供的存储。在右上角，可以看到一个 tie-breaker 服务器。网络连接和光纤连接也显示出来以供参考。所有这些内容会在本文的后续章节详细加以介绍。集群的其他部分显示为云状，本文不会涉及。更多关于该集群其他部分的信息，请参看本系列的第 1 部分和第 2 部分。节点 GPFS 集群的大部分节点都运行 Red Hat Enterprise Linux 3。本例使用服务器/客户机架构，其中，少数几个服务器通过光纤通道拥有对存储的可见性。它们充当集群其他部分的网络共享磁盘 (NSD) 服务器。这意味着大多数 GPFS 集群的成员都将使用 NSD 服务器经由 IP 访问存储。NSD 节点（即存储节点）一共有四个：GPFS 集群的逻辑分区各两个。它们均按对分组，每一对管理其中的一个存储子系统。Tiebreaker 集群的每一半都包含完全相同数目的节点，若其中的一半丢失，就会带来仲裁问题。通过 GPFS，若要使文件系统保持可用，就需要有一个有效的节点仲裁。Quorum 由  $Quorum = (\text{number of quorum nodes} / 2) + 1$  定义。在本例所示的这种配置中，集群由完全相同的两个部分组成，若其中任何一个丢失，GPFS 文件系统都会不可用。为了避免这种情况，系统采用了一个 tie-breaker 节点。此节点的物理位置与主集群不在一起。这意味着若任何

一部分不可用，另一部分都能继续访问 GPFS 文件系统。这是通过使用三个故障组得以实现的，我们将在 数据复制 一节详细对之进行介绍。这意味着数据的两个副本都可用：集群的每一半各一个。架构如图 1 所示，每个节点均连接至两个网络。其中的第一个网络用于计算流量和总的集群通信。第二个网络则专门针对 GPFS，为那些不具有 Storage Area Network (SAN) 存储系统直观视图的节点提供经由 IP 的存储访问。这个网络使用巨大帧来获得好的性能。有关存储网络的更详细的信息，请参阅本系列第 4 部分的 GPFS 网络调优一节。存储区域网络本解决方案的存储后端包括两个磁盘子系统，均为 IBM TotalStorage DS4500（之前为 FAStT 900）磁盘系统，每一个都附带完全填充过的 EXP710 扩展磁盘柜。每个 DS4500 均被配置成 RAID 5 4 P 阵列外加热后备磁盘。每个 DS4500 均由一对存储服务器所有。在这两个服务器间拆分 4 P 阵列，以便每个服务器都可充当阵列其中一半的主服务器、阵列另一半的备用服务器。这样一来，若任何一个服务器故障，另一个服务器都能立即作为主服务器从故障的服务器接管磁盘。100Test 下载频道开通，各类考试题目直接下载。详细请访问 [www.100test.com](http://www.100test.com)