

# 基于单磁盘的 RAID 系统可靠性分析<sup>①</sup>

张都乐, 祝怀杰, 何 淼

(中国石油天然气管道工程有限公司, 廊坊 065000)

**摘要:** RAID 系统具有高可靠性、高可扩展性和高性价比的特性, 是当前解决信息系统外部存储的一种主要技术. 随着该技术的普及以及系统规模的扩大, 系统失效事件日益频繁, 因此 RAID 系统的可靠性研究成为一个热点. 简单介绍了 RAID 系统工作原理及其硬盘组合方式, 从单个硬盘的可靠性出发, 推导出 RAID 系统的可靠性数学模型, 并进行了分析.

**关键词:** 独立磁盘冗余阵列; 冗余技术; 可靠性; 数学模型

## Reliability Analysis of RAID Storage System Based on Single Physical Hard Disk

ZHANG Du-Le, ZHU Huai-Jie, HE Miao

(China Petroleum Pipeline Engineering Corporation, Langfang 065000, China)

**Abstract:** As it possesses characteristics such as high concurrency, high scalability, high cost-effectiveness and so on, the RAID system is the current solution is one of the main technology of information system external storage. With the popularity of the applications of the technology as well as the expansion of the system size, its failure increases dramatically, so the research on the reliability of RAID system has become a hot spot. Based on the reliability analysis of RAID system, the reliability mathematical model was derived.

**Key words:** Redundant Array of Inexpensive Disk(RAID); redundant technology; reliability; mathematical model

存储设备在信息系统中处于非常重要的地位, 一方面, 存储设备是信息系统数据的仓库, 他的安全直接关系到整个信息系统的安全, 另一方面, 存储设备采用硬盘作为其主要部件, 由于其结构方面的原因, 相对于信息系统的其它元器件而言, 它是低速的、串行设备, 多年来其性能一直难以得到明显的提高, 是整个系统架构中的性能瓶颈. 要改善存储设备的性能, 目前可行的方法主要是组建磁盘阵列(RAID), 与单个磁盘驱动器及其简单的组合相比, RAID 可以在现有的磁盘技术上, 为用户提供更高的数据传输速率、可靠性和简便管理, 从而加强存储设备的性能: 首先, RAID 采用一定的技术手段如数据划分和映射技术, 将 I/O 请求并行分布到各个磁盘成员上, 通过并行处理, 减少了读写等待时间, 提高了系统的性能; 其次, 在 RAID 中存储了校验信息, 当某个或某些磁盘成员

发生故障时, 可以根据工作正常的磁盘信息来重新生成发生故障磁盘的信息, 从而极大提高了数据的可用性; 第三, RAID 将多个物理硬盘集成为一个大的逻辑盘, 为用户提供了一个统一的、与硬件无关的存储资源池, 简化了存储管理, 提高了存储的灵活性和可扩展性.

常见的 RAID 有很多种实现方案, 如 RAID0、RAID1、RAID01/10、RAID 5、RAID 6 等等, 磁盘阵列实现方案因控制器采用的工作模式和算法不同, 即使采用相同的硬盘组建, RAID 系统的可靠性、吞吐率、容量/磁盘利用率和相应时间等重要指标也不相同, 有时差距较大. 而 RAID 系统的可靠性成为 RAID 系统设计、采购、运维非常重要的参数, 其可靠性模型成为热点之一<sup>[1]</sup>. 文献[2-7]给出部分 RAID 系统的可靠性模型, 并进行了验证.

<sup>①</sup> 收稿时间:2013-06-22;收到修改稿时间:2013-07-15

本文在研究了各种 RAID 方案工作原理的基础上, 借助可靠性理论, 综合地介绍求解各种 RAID 技术的可靠性模型, 使用这些模型, 对不同的 RAID 类型的影响进行评估, 从安全视角, 给出了 RAID 系统采购及配置建议, 有助于建设投资少、可靠性高、满足需求的存储系统。

### 1 RAID技术简介<sup>[8, 9]</sup>

磁盘阵列是当前计算机系统中主要的外存储设备,

它将多个物理硬盘通过某种算法抽象成一个具有高可靠、高可用和便于管理的逻辑的存储资源池提交给主机使用。

在 RAID 系统中, 磁盘阵列通过 stripe 方式来组织数据, 将主机 I/O 请求数据以数据块(block)为单位进行分割和冗余校验的计算, 将数据块(包含冗余信息)以物理磁盘总数 n 为模进行交叉编址——地址映射, 利用并行技术存放在不同的独立物理数据磁盘上, n 个独立的数据盘可以并行操作, 其数据组织形式如图 1<sup>[9]</sup>。

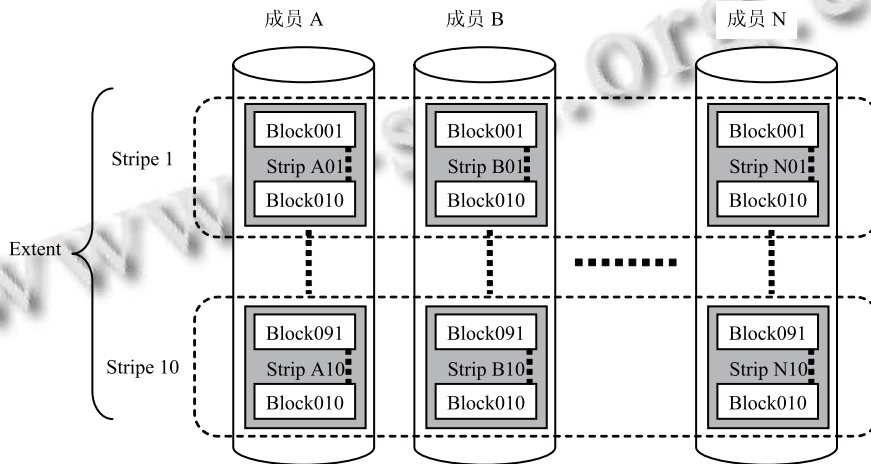


图 1 RAID 数据组织形式

根据数据块、地址映射、冗余信息等数据保护方法的不同, 形成了不同的 RAID 级别, 常见的有 RAID0、RAID1、RAID2、RAID3、RAID4、RAID5、

RAID6 等 7 个基本层级, 以及基本层级的组合层级, 如 RAID01/10 等. 其基本特征如表 1.

表 1 RAID 级别的基本特征

RAID 级别	特征	数据块	冗余信息编码	冗余信息		备注
				长度	存放	
0	数据条带/无冗余信息	数据块	无	无	无	
1	镜像阵列	数据块	复制	无	独立	
01/10	磁盘镜像与复制	数据块	复制	无	独立	
2	具有错误纠正功能	Bit	海明码	m	独立	
3	位交叉奇偶校验阵列	Bit	奇偶校验	1	独立	
4	块交叉奇偶校验阵列	Block	奇偶校验	1	独立	
5	块交叉分布式奇偶校验阵列	Block	奇偶校验	1	分布式	
6	块交叉分布式 P+Q 校验阵列	Block	Reed-Solomon 编码	2	分布式	

### 2 RAID系统可靠性数学模型

组建 RAID 系统重要目的是为了提升硬盘的读写性能以及系统的可靠性等. 一个磁盘阵列可以包含多

个不同 RAID 级别和多种不同的硬盘, 为简化模型, 本文作如下假定: a.一个磁盘阵列中只包含一个 RAID 级别; b.在一个 RAID 系统中, 所有硬盘性能相同, 包

括容量、RPM、IOPS、可靠度等参数。

2.1 符号说明

- n——RAID 系统中硬盘的数量;
- $R_{HDD}$ ——单个硬盘的可靠度;
- $R_{RAID}$ ——RAID 系统的可靠度;
- A,  $A_0, A_1, A_2$ ——分别表示事件“RAID 系统是可靠的”, “所有硬盘均正常”, “有且仅有一个硬盘出现故障”, “有且仅有两个硬盘同时出现故障”;
- $P(A)$ ——表示“事件 A 发生的概率”。

2.2 可靠性模型

一个存储系统要达到一定的可靠性, 则各个部分都需要达到一定的可靠度要求, 在 RAID 系统, 系统的可靠性可分配到每个硬盘中, 但是整个系统的可靠性并不是它们的总和。

1) RAID0 的可靠性模型

RAID0 系统中所有硬盘为串行, 其可靠性框图如图 2。

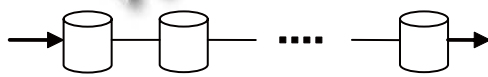


图 2 RAID0 系统的可靠性框图

其可靠性的数学模型为:

$$R_{RAID0} = \prod_{i=1}^n R_i = R_{HDD}^n \quad (1)$$

2) RAID1 的可靠性模型

从可靠性角度看, RAID1 系统是将两块硬盘进行并联, 其可靠性框图如图 3。

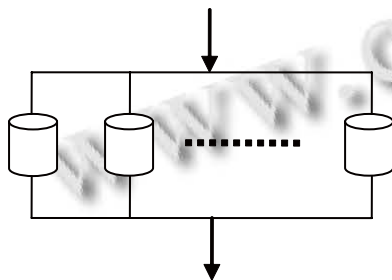


图 3 RAID1 系统的可靠性框图

其可靠性的数学模型为:

$$R_{RAID1} = 1 - (1 - R_1) \times (1 - R_2) = 1 - (1 - R_{HDD})^2 \quad (2)$$

3) RAID2、3、4、5 的可靠性模型

RAID2、3、4、5 系统采用海明编码、奇偶校验

编码等容错算法, 在由 n 个硬盘组成的 RAID2、3、4、5 磁盘阵列中, 在同一时刻, 最多只有一个硬盘出现故障, 数据不会丢失, 则称系统是可靠的。其可靠性框图如图 4。

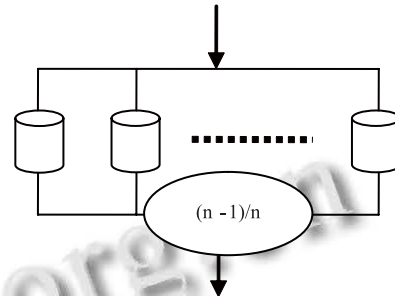


图 4 RAID2/3/4/5 系统的可靠性框图

设  $A_0$  为事件“n 个硬盘均正常”,  $A_1$  为事件“有且仅有一个硬盘出现故障”, A 为事件“最多只有一个硬盘出现故障”,  $A = A_0 \cup A_1$ ;

$$R_{RAID2/3/4/5} = P(A) = P(A_0 \cup A_1) = P(A_0) + P(A_1)$$

其中:

$$P(A_0) = C_n^n \times R_{HDD}^n = R_{HDD}^n$$

$$P(A_1) = C_n^{n-1} \times R_{HDD}^{n-1} \times (1 - R_{HDD}) = n \times R_{HDD}^{n-1} \times (1 - R_{HDD}) \quad (3)$$

$$R_{RAID2/3/4/5} = R_{HDD}^n + n \times R_{HDD}^{n-1} \times (1 - R_{HDD})$$

4) RAID6 的可靠性分析

RAID6 系统采用 Reed-Solomon 编码方式, 生成两个奇偶校验位, 根据系统的工作原理可知: 在由 n 个硬盘组成的 RAID6 磁盘阵列中, 最多只有两个硬盘同时出现故障, 数据不会丢失, 则称系统是可靠的。其可靠性框图如图 5。

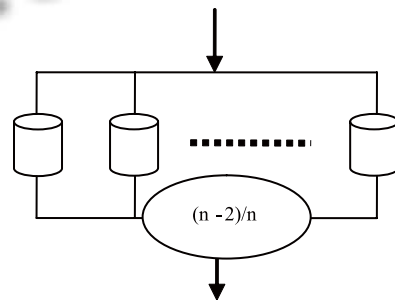


图 5 RAID6 系统的可靠性框图

设  $A_0$  为事件“n 个硬盘均正常”,  $A_1$  为事件“有且仅有一个硬盘出现故障”,  $A_2$  为事件“有且仅有两个硬盘同时出现故障”, A 为事件“最多只有两个硬盘同时出现故障”,  $A = A_0 \cup A_1 \cup A_2$ ; 则其可靠性的数学模型为:

$$R_{RAID6} = P(A) = P(A_0 \cup A_1 \cup A_2)$$

$$R_{RAID6} = P(A_0) + P(A_1) + P(A_2)$$

$$P(A_0) = C_n^n \times R_{HDD}^n = R_{HDD}^n$$

$$P(A_1) = C_n^{n-1} \times R_{HDD}^{n-1} \times (1 - R_{HDD}) = n \times R_{HDD}^{n-1} \times (1 - R_{HDD}) \quad (5)$$

$$P(A_2) = C_n^{n-2} \times R_{HDD}^{n-2} \times (1 - R_{HDD})^2 = \frac{n \times (n-1)}{2} \times R_{HDD}^{n-2} \times (1 - R_{HDD})^2$$

$$R_{RAID6} = R_{HDD}^n + n \times R_{HDD}^{n-1} \times (1 - R_{HDD}) + \frac{n \times (n-1)}{2} \times R_{HDD}^{n-2} \times (1 - R_{HDD})^2$$

### 6) RAID01 的可靠性分析

在 n 个硬盘组成的 RAID01 阵列中, 将硬盘分为两组, 每组 n/2 个硬盘, 其可靠性模型可视为 n/2 个硬盘串联, 然后进行并联, 其可靠性框图如图 6.

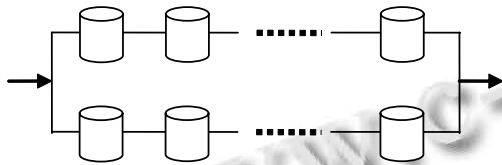


图 6 RAID01 系统的可靠性框图

其数学模型为:

$$R_{RAID01} = 1 - (1 - R') \times (1 - R') = 1 - (1 - R_{HDD}^{n/2}) \times (1 - R_{HDD}^{n/2}) \quad (6)$$

$$R_{RAID01} = 1 - (1 - R_{HDD}^{n/2})^2$$

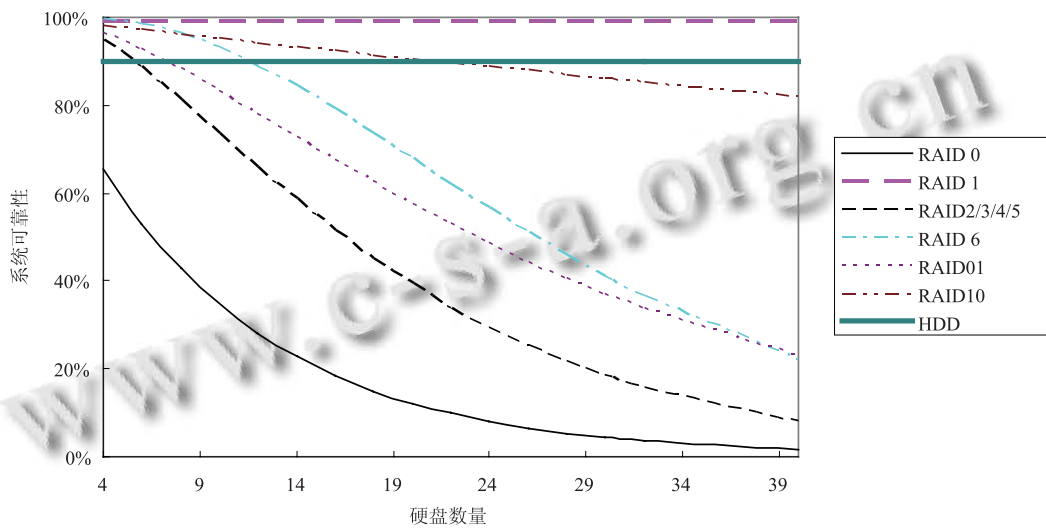


图 8 RAID 系统可靠性与硬盘数量的关系

从图 8 可以得出如下结论:

1) 随着硬盘数量的增加, RAID 系统的可靠性都在不同程度的降低, 其中 RAID0 降低速度最大, RAID10 最为平稳, 依次为: RAID10、RAID6、RAID01、RAID5、RAID2/3/4/5、RAID0。故在性能满

### 7) RAID10 的可靠性分析

在 n 个硬盘组成的 RAID10 阵列中, 2 个硬盘组成一个镜像对, 共有 n/2 组, 其可靠性模型可视为 2 个硬盘进行并联, 然后 n/2 组镜像对串联, 其可靠性框图如图 7.

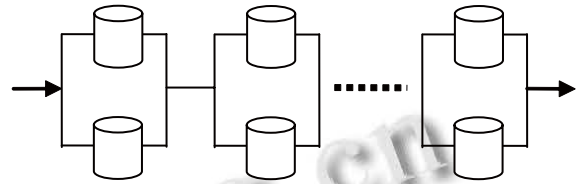


图 7 RAID10 系统的可靠性框图

其数学模型为:

$$R_{RAID10} = (1 - (1 - R_{HDD})^2)^{n/2} \quad (7)$$

### 3 RAID 可靠性分析

公式(1)~(7)组成了 RAID 系统的可靠性模型, 根据该模型给出 RAID 系统的可靠性分析.

假设硬盘在 t 时间内的可靠度为 90%<sup>[4]</sup>, 各 RAID 系统可靠性与硬盘数量的关系如图 8.

足的情况下, 可采用尽量少的硬盘配置阵列(选用大容量的硬盘——基于安全视点).

2) 在相同硬盘数量下, 在 8 块盘以内, RAID6 的可靠性最高, 其次为 RAID10, 超过 8~10 盘, RAID10 的可靠性超过 RAID 6, 其余依次为 RAID01、

## RAID2/3/4/5 和 RAID0.

3) RAID+HotSpare 模式从概率角度来看,并不能增加系统的可靠性,只是减少了 RAID 系统平均故障维修时间,但是从运维角度来看,热备盘的增加,非常大的提高了系统的可用性.

## 4 结束语

本文研究的目的在于通过研究 RAID 系统工作原理,结合可靠性理论,比较系统的给出常见的各种技术方案的可靠性数学模型,并在相同条件下,进行了分析与对比,从可靠性角度出发,对系统的配置给出一定的建议,对 RAID 设备的采购、日常运维具有一定的意义;但是,系统的可靠性是一个非常复杂的问题,不仅仅涉及到硬盘及其组成方式,同时牵涉到 RAID 系统的控制器、操作系统等诸多方面,不同的系统级冗余容错结构具有不同的工作原理、工作特点和故障容忍特性.由于这些冗余结构通常都是与重组、恢复技术结合在一起构成动态冗余结构使用,因此针对具体的 RAID 系统的可靠性分析还应在通用性理论的基础上结合具体情况进行分析,使得其可靠性模型更符合实际情况.

(上接第 92 页)

## 7 总结

光伏阵列单元的设计性能直接关系到整个光伏电站的整体性能.光伏电站阵列单元优化设计软件,为工程设计人员提供了一个交互式的设计平台.设计人员通过该软件完成阵列模型的建立及性能分析,进而实现优化设计,从而保证了整个电站的设计合理性.

## 参考文献

- 1 李安定,吕全亚.太阳能光伏发电系统工程.北京:化学工业出版社,2012.
- 2 太阳光发电协会.太阳能光伏发电系统的设计与施工.北京:科学出版社,2011.
- 3 Mayfield R. Photovoltaic design & Installation. WILEY Press.

## 参考文献

- 1 刘军平,周可,雷栋梁,庞丽萍.基于随机理论的 RAID 可靠性仿真方法.计算机研究与发展,2011,48(z1):12-16.
- 2 王新春,洪明.基于单物理硬盘的磁盘阵列研究.楚雄师范学院学报,2010,25(12):48-51.
- 3 葛杰.RAID 应用的性能及可靠性分析.南通职业大学学报,2007,21(4):87-91.
- 4 陈华英.磁盘阵列 RAID 可靠性分析.电子科技大学学报,2006,35(3):403-405.
- 5 章宏灿,薛巍.集群 RAID5 存储系统可靠性分析.计算机研究与发展,2010,47(4):727-735.
- 6 Chen PM, Lee EK, Gibson GA, et al. RAID: high-performance, reliable secondary storage. ACM Computing Surveys, 1994,26(2):145-185.
- 7 Gibson G, Patterson D. Designing disk arrays for high data reliability. Journal of Parallel and Distributed Computing, 1993, 17: 4-27.
- 8 李琼,汪审权,庞征斌,刘光明.高可靠磁盘阵列的设计.计算机研究与发展,2003,20(7):49-51.
- 9 赵亮.高性能磁盘阵列(RAID)关键技术的研究[硕士学位论文].长沙:国防科学技术大学,2002.

2011.

- 4 周治,吕康,范小苗,等.光伏系统设计软件简介.西北水电,2009,(6):76-79.
- 5 Photovoltaic Power Systems Program-Me. Worldwide overview of design and simulation tools for hybrid PV systems, Report IEA-PVPS T11-01.2011.
- 6 Kerekes T, Koutroulis E. An Optimization Method for Designing Large PV Plants. IEEE Journal of Photovoltaics, 2013, (3).
- 7 Messenger RA, Ventre J. Photovoltaic Systems Engineering. 3rd Ed. CRC Press, 2010.
- 8 Habberlin H. Photovoltaics System Design and Practice. WILEY Press, 2012.