MODELING THE RELIABILITY OF RAID SETS

By Abraham Long Jr.

By evaluating the expected reliability of different RAID levels alongside that of other system components and key factors such as performance, enterprises can choose a data protection configuration appropriate to their environment and overall IT strategy.

Related Categories:

Dell PowerEdge tower servers

RAID

Visit DELL.COM/PowerSolutions for the complete category index.

Creating a robust infrastructure that can handle hardware failures without data loss is critical to the success of many enterprises, and implementing RAID configurations can be a key part of that effort. However, choosing an appropriate configuration can be complex—different RAID levels require trade-offs in performance, level of data protection, and overall reliability, requiring enterprises to accurately gauge the level of risk they can tolerate and evaluate the trade-offs for each RAID level.

This article details the mathematical models of the relationships between hard disk drives (HDDs) in RAID sets using RAID levels 6, 0+1, 0+3, 30, 50, 60, and 100, of which RAID levels 6, 50, and 60 are supported on Dell™ PowerEdge™ servers. These mathematical models can help enterprises evaluate the overall reliability of different standard and nested RAID configurations. Although a single disk array can contain multiple RAID sets of different RAID levels and disk capacities, the examples in this article are based on an array of eight HDDs dedicated to one RAID level at a time. All examples use a hypothetical Serial Attached SCSI (SAS)/Serial ATA (SATA) HDD with a reliability figure of merit of 0.90 over five years in a 100 percent duty cycle (indicating a 90 percent chance that the HDD will not fail over that period).

EVALUATING STANDARD RAID CONFIGURATIONS: RAID LEVEL 6

RAID-6 augments the RAID-5 distributed parity scheme by implementing two independent parity computations and storing the results on different HDDs; it requires a minimum of four HDDs to implement, and can tolerate two HDD failures without data loss. Consequently, RAID-6 is designed to provide the highest reliability of the standard (non-nested) RAID levels—if two HDDs fail simultaneously, resulting in the loss of data and one set of parity information, the lost data is still recoverable with the remaining parity information.

The mathematical relationship that evaluates the reliability of *n* HDDs in a RAID-6 configuration is identical to that of RAID-5 (see Figure 1). In the figure, A1, B1, and so on each represent one data block, while A_p, B_p, and so on represent parity information and A_q, B_q, and so on represent redundant parity information.

For a Dell PowerEdge server with eight HDDs, a typical configuration would be to dedicate all eight HDDs to RAID-6. The reliability calculation for this

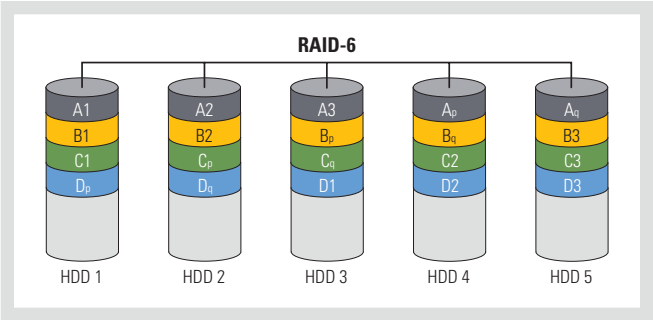

Figure 1. *Example RAID-6 configuration*

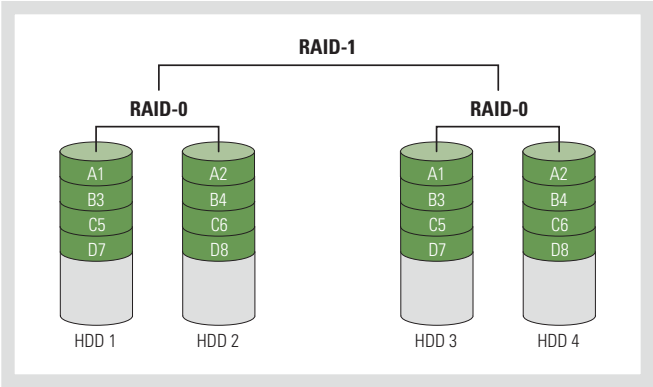

Figure 2. *Example RAID-0+1 configuration*

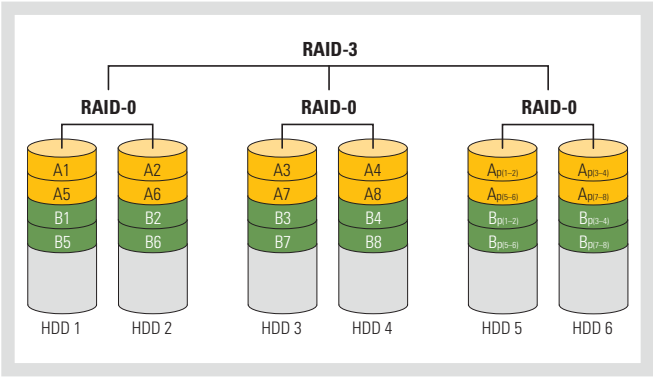

Figure 3. *Example RAID-0+3 configuration*

configuration is as follows (where *k* is the number of HDDs that must operate out of *n* HDDs—in this case, six out of eight):

$$R_{array} = \sum_{j=6}^{8} \binom{n}{j} R_{HDD}^{j} \left(1 - R_{HDD}\right)^{(n-j)} = \frac{8!}{6!(8-6)!} (0.9)^6 (1-0.9)^{(8-6)}$$

$$+ \frac{8!}{7!(8-7)!} (0.9)^7 (1-0.9)^{(8-7)} + \frac{8!}{8!(8-8)!} (0.9)^8 (1-0.9)^{(8-8)}$$

$$= 0.15 + 0.38 + 0.43 = 0.96$$

This result indicates that the probability of no data loss during a five-year period is approximately 96 percent, or, conversely, that the probability of losing data during a five-year period is approximately 4 percent.

EVALUATING NESTED RAID CONFIGURATIONS: RAID LEVELS 0+1, 0+3, 30, 50, 60, AND 100

Nested RAID levels are combinations of standard RAID levels, and include RAID-0+1 (a mirror of stripes), RAID-0+3 (a dedicated parity array across striped HDDs), RAID-30 (a RAID-0 array striped across RAID-3 elements), RAID-50 (a RAID-0 array striped across RAID-5 elements), RAID-60 (a RAID-0 array striped across RAID-6 elements), and RAID-100 (a RAID-0 array striped across RAID-10 elements).

RAID-0+1: Mirror of stripes

In RAID-0+1, data is striped to one disk set array and then mirrored to another disk set array, which helps provide good I/O performance and reliability. It requires a minimum of four HDDs to implement (see Figure 2). RAID-0+1 is not as fault tolerant as RAID-10, and cannot tolerate two simultaneous disk failures unless the second failed disk is in the same striped disk set array. If one HDD in one disk set array fails, data in that disk set array is lost, but all data remains available from the mirrored disk set array. However, an HDD failure in the remaining disk set array (the mirror) before the failed HDD is replaced results in data loss.

RAID-0+1 is not supported by Dell PowerEdge Expandable RAID Controllers (PERCs) and SAS RAID controllers, but a typical configuration for a server with eight HDDs would be to dedicate all eight HDDs to RAID-0+1. In this case, the eight HDDs would be divided into two disk arrays of four HDDs each. The reliability calculation for this configuration is as follows:

$$R_{RAID-0+1} = \prod_{i=1}^{\#\ of\ disk\ sets} \left[1 - \left(1 - R_{HDDi}^2\right)\left(1 - R_{HDDi}^2\right)\right]$$

$$= \left[1 - \left(1 - 0.90^2\right)\left(1 - 0.90^2\right)\right] \times \left[1 - \left(1 - 0.90^2\right)\left(1 - 0.90^2\right)\right] = 0.93$$

RAID-0+3: Dedicated parity array across striped HDDs

In RAID-0+3, blocks of data are striped across RAID-0 sub-arrays configured in a RAID-3 array with dedicated parity in one of the RAID-0 sub-arrays. RAID-0+3 requires a minimum of six HDDs to configure an array in sub-arrays of two or more. Figure 3 shows 8-byte blocks of data striped across two RAID-0 sub-arrays with the dedicated 4 parity bytes in one of the RAID-0 sub-arrays; in this figure, $A_{p(1-2)}$ applies to the parity created from the two RAID-0 HDDs that contain the blocks of data A1 and A2.

If one HDD fails in one of the RAID-0 sub-arrays, the corrupted data can be replaced by using exclusive OR (XOR) operations with the remaining HDD and the parity HDD. However, if two HDDs fail in any given RAID-0 sub-array, the corrupted data

is typically not recoverable. Furthermore, losing one of the RAID-0 sub-assemblies would also result in the loss of a block of data—thus losing all data in the entire RAID-0+3 array.

RAID-0+3 is not supported by Dell PERCs or SAS RAID controllers, but a typical configuration for a server with eight HDDs would be to dedicate six HDDs to RAID-0+3 and designate the remaining two HDDs as spares. Evaluating RAID-0+3 reliability entails two calculations. First, the reliability calculation for each RAID-0 sub-array in this configuration would be as follows:

$$R_{RAID-0} = \sum_{j=2}^{3} \binom{n}{j} R_{HDD}^{j} \left(1 - R_{HDD}\right)^{(n-j)}$$

$$= \frac{3!}{2!(3-2)!} 0.9^2 \left(1-0.9\right)^{(3-2)} + \frac{3!}{3!(3-3)!} 0.9^3 \left(1-0.9\right)^{(3-3)}$$

$$= 0.24 + 0.73 = 0.97$$

The overall reliability of the entire RAID-3 array, then, is the product of the reliabilities of the two RAID-0 sub-arrays:

$$R_{RAID-0+3} = \prod_{i=1}^{n} R_{RAID-0_i} = 0.97 \times 0.97 = 0.94$$

RAID-30: RAID-0 array striped across RAID-3 elements

RAID-30 is a combination of a RAID-3 array and a RAID-0 array, and requires a minimum of six HDDs to implement. RAID-30 arrays typically consist of two RAID-3 disk sub-arrays with data striped across both disk sub-arrays. The parity bit or byte is calculated by performing an XOR operation on the blocks of data for each RAID-3 sub-array, with the parity bit or byte created with the XOR operation then written to the last HDD in each RAID-3 sub-array. One HDD in each RAID-3 sub-array can fail without loss of data; however, if two HDDs fail in a given RAID-3 sub-array, all data in the entire RAID-30 array is lost.

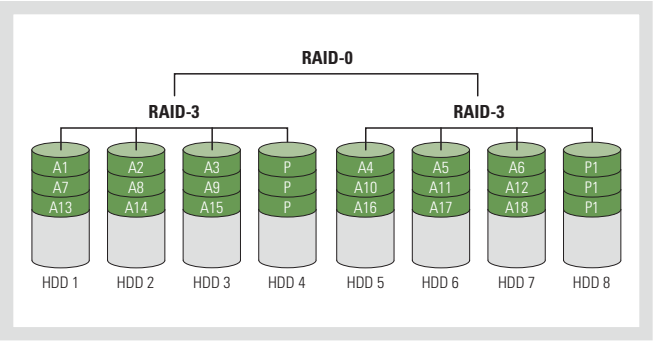

Figure 4. *Example RAID-30 configuration*

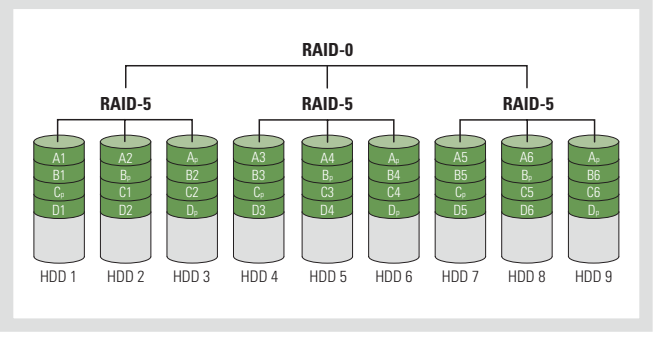

Figure 5. *Example RAID-50 configuration*

RAID-30 is not supported by Dell PERCs or Dell SAS RAID controllers, but a typical configuration for a server with eight HDDs would be to dedicate all eight HDDs to RAID-30 (see Figure 4). The reliability calculation for each RAID-3 sub-array in this configuration would be as follows:

$$R_{RAID-3} = \sum_{j=3}^{4} \binom{n}{j} R_{HDD}^{j} \left(1 - R_{HDD}\right)^{(n-j)} = \frac{4!}{3!(4-3)!} 0.9^3 \left(1-0.9\right)^{(4-3)}$$

$$+ \frac{4!}{4!(4-4)!} 0.9^4 \left(1-0.9\right)^{(4-4)} = 0.29 + 0.66 = 0.95$$

The overall reliability of the entire RAID-30 is the product of the reliabilities of the two RAID-3 sub-arrays:

$$R_{RAID-30} = \prod_{i=1}^{2} R_{RAID-3_i} = 0.95 \times 0.95 = 0.90$$

RAID-50: RAID-0 array striped across RAID-5 elements

RAID-50 combines the block-level striping of RAID-0 with the distributed parity of RAID-5, and requires a minimum of six HDDs to implement. One HDD in each RAID-5 sub-array can fail without loss of data; however, if two HDDs fail in a given RAID-5 sub-array, all data in the entire RAID-50 array is lost. Figure 5 shows a RAID-50 array comprising nine HDDs, in which a total of three HDDs (one in each RAID-5 sub-array) can fail without data loss.

For a Dell PowerEdge server with eight HDDs, a typical configuration would be to dedicate six HDDs to RAID-50: each of the two RAID-5 sub-arrays would contain three HDDs, with two hot spares that could immediately start rebuilding the sub-array following a failure. The reliability calculation for each RAID-5 sub-array in this configuration would be as follows:

$$R_{RAID-5} = \sum_{j=2}^{3} \binom{n}{j} R_{HDD}^{j} \left(1 - R_{HDD}\right)^{(n-j)} = \frac{3!}{2!(3-2)!} 0.9^2 \left(1-0.9\right)^{(3-2)}$$

$$+ \frac{3!}{3!(3-3)!} 0.9^3 \left(1-0.9\right)^{(3-3)} = 0.24 + 0.73 = 0.97$$

The overall reliability of the entire RAID-50 array is the product of the reliabilities of the two RAID-5 arrays:

$$R_{RAID-50} = \prod_{i=1}^{2} R_{RAID-5_i} = 0.97 \times 0.97 = 0.94$$

RAID-60: RAID-0 array striped across RAID-6 elements

RAID-60 combines the block-level striping of RAID-0 with the distributed parity of RAID-6, and requires a minimum of eight HDDs to configure (see Figure 6). Two HDDs from each RAID-6 sub-array can fail without data loss; however, if more than two HDDs fail in a given RAID-6 sub-array, all data in the entire RAID-60 array is lost.

For a Dell PowerEdge server with eight HDDs, a typical configuration would be to dedicate all eight HDDs to RAID-60. The reliability calculation for each RAID-6 sub-array in this configuration would be as follows:

$$R_{RAID-6} = \sum_{j=2}^{4} \binom{n}{j} R_{HDD}^{j} \left(1 - R_{HDD}\right)^{(n-j)}$$

$$= \frac{4!}{2!(4-2)!} 0.9^2 (1-0.9)^{(4-2)} + \frac{4!}{3!(4-3)!} 0.9^3 (1-0.9)^{(4-3)}$$

$$+ \frac{4!}{4!(4-4)!} 0.9^4 (1-0.9)^{(4-4)} = 0.049 + 0.291 + 0.656 = 0.996$$

The overall reliability of the entire RAID-50 array is the product of the reliabilities of the two RAID-6 arrays:

$$R_{RAID-60} = \prod_{i=1}^{2} R_{RAID-6_i} = 0.996 \times 0.996 = 0.99$$

RAID-100: RAID-0 array striped across RAID-10 elements

RAID-100 is a stripe of RAID-10 arrays, and has failure characteristics identical to RAID-10—that is, all but one HDD from each RAID-1 sub-array can fail without loss of data. If two HDDs fail in any given RAID-1 sub-array, however, all data stored in the entire RAID-100 array is lost.

RAID-100 is not supported by Dell PERCs and Dell SAS RAID controllers, but a typical configuration for a server with eight HDDs would be to dedicate all eight HDDs to RAID-100 (see Figure 7). The RAID-100 reliability solution is as follows:

$$R_{RAID-100} = P(1234 \cup 5678 \cup 1678 \cup 5234 \cup 1674 \cup 1634 \cup 1638 \cup 1278 \cup 1274 \cup 1238 \cup 5674 \cup 5634 \cup 5638 \cup 5274 \cup 5238 \cup 5278)$$

This equation can be solved through cut-sets, tie-sets, or a Boolean truth table approach evaluating 256 combinations. The resulting reliability for RAID-100 with eight HDDs is 0.96.

IDENTIFYING APPROPRIATE RAID CONFIGURATIONS

Figure 8 summarizes the reliability expectations for the RAID configurations discussed in this article—that is, the probability of not losing data over a five-year period. In terms of reliability engineering, enterprises should combine these probabilities with the expected reliability of other system elements to help gauge overall system availability. Evaluating these expectations alongside other factors such as system performance and hardware resources can enable enterprises to identify appropriate RAID configurations as part of their overall IT strategy.

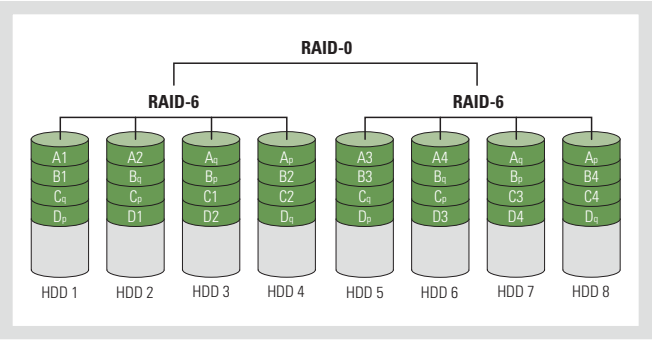

Figure 6. *Example RAID-60 configuration*

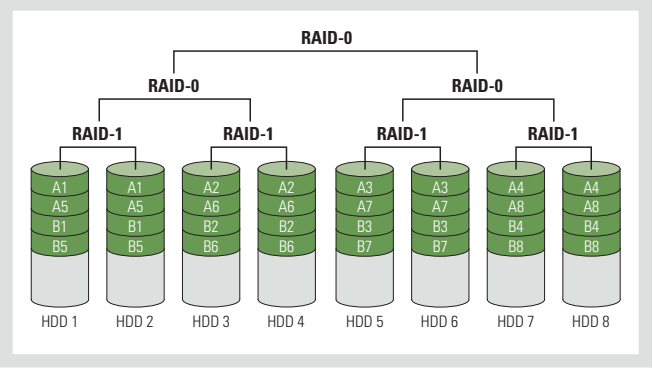

Figure 7. *Example RAID-100 configuration*

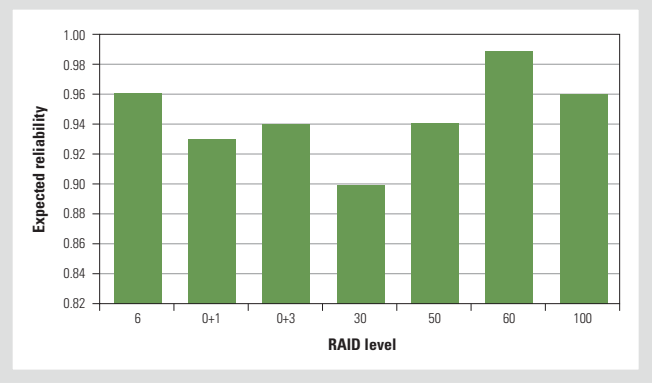

Figure 8. *Expected reliability for different RAID levels*

Abraham Long Jr. is a senior engineering consultant in the Dell Product Group Server Reliability Engineering team in the Dell Enterprise Systems Group. He has a B.S. in Industrial Engineering from New Mexico State University and an M.S. in Systems Engineering from California State University, Northridge. He is a senior member of the American Society for Quality (ASQ), has ASQ certifications in Reliability Engineering and Quality Engineering, and is a Six Sigma Black Belt.