

WHITEPAPER

BASIC INFORMATION

HDD & RAID

INFORMATION ON SERVICE LIFE AND FUNCTIONALITY
OF HARD DISKS AND REDUNDANT STORAGE SYSTEMS

Copyright © 2019 Dallmeier electronic GmbH & Co.KG

The reproduction, distribution and utilization of this document as well as the communication of its contents to others without express authorization is prohibited. Offenders will be held liable for the payment of damages.

All rights reserved in the event of the grant of a patent, utility model or design.

The manufacturer accepts no liability for damage to property or pecuniary damages arising due to minor defects of the product or documentation, e.g. print or spelling errors, and for those not caused by intention or gross negligence of the manufacturer.

Figures (e.g. screenshots) in this document may differ from the actual product. Specifications subject to change without notice. Errors and misprints excepted.

All trademarks identified by ® are registered trademarks of Dallmeier.

Third-party trademarks are named for information purposes only. Dallmeier respects the intellectual property of third parties and always attempts to ensure the complete identification of third-party trademarks and indication of the respective holder of rights. In case that protected rights are not indicated separately, this circumstance is no reason to assume that the respective trademark is unprotected.

TABLE OF CONTENTS

- CHAPTER 1: ABSTRACT 4**

- CHAPTER 2: HDD DURABILITY 5**
 - 2.1 AFR, MTBF and MTTF 5
 - 2.1.1 Definition 5
 - 2.1.1.1 AFR 5
 - 2.1.1.2 MTBF 5
 - 2.1.1.3 MTTF 6
 - 2.1.2 Relation 6
 - 2.1.3 Interpretation 6
 - 2.1.4 Practical Relevance 8
 - 2.1.4.1 Failure Rate - A Google Inc. Study 8
 - 2.1.4.2 Failure Time - A Carnegie Mellon University Study 9
 - 2.1.4.3 Failure Rate - A Dallmeier electronic Analysis 11
 - 2.1.5 Conclusion 12
 - 2.2 SMART 12
 - 2.2.1 Definition 12
 - 2.2.2 Interpretation 12
 - 2.2.3 Practical Relevance 13
 - 2.2.4 Conclusion 14

- CHAPTER 3: STORAGE TECHNOLOGIES 15**
- 3.1 JBOD 15
 - 3.1.1 Capacity and Costs 15
 - 3.1.2 Safety and Rebuild 15
 - 3.1.3 Conclusion 16
- 3.2 3.2 RAID 1 16
 - 3.2.1 Capacity and Costs 16
 - 3.2.2 Safety and Rebuild 17
 - 3.2.3 Conclusion 17
- 3.3 RAID 5 17
 - 3.3.1 Capacity and Costs 18
 - 3.3.2 Safety and Rebuild 18
 - 3.3.3 Conclusion 18
- 3.4 RAID 6 19
 - 3.4.1 Capacity and Costs 19
 - 3.4.2 Safety and Rebuild 20
 - 3.4.3 Conclusion 20

- CHAPTER 4: RECOMMENDATIONS 21**

- CHAPTER 5: REFERENCES 22**

ABSTRACT

Hard disk drives (**HDDs**) are storage media that write and read binary data on rotating magnetic disks using a sophisticated mechanism [1]. They represent a central component of all digital recording systems and are used for the continuous storage of audio and video data. Despite the highest quality standards a natural wear of HDDs must be considered, which is even more increased by usually uninterrupted operation (24/7).

A **RAID** system is a storage technology that combines multiple HDDs into a single logical unit, in which all data are stored redundantly [2]. This redundancy allows for the failure and replacement of individual HDDs without data loss. In digital recording systems, RAID systems are used to absorb the natural wear and failure of hard disk drives.

Even advanced RAID systems in conjunction with highest quality HDDs can not guarantee complete data security. This conclusion is based on the limitations of the RAID technology [3] and the lifetime of hard disk drives observed in practice [4]. The common view, that a RAID system provides the same security as a backup is not applicable [5]. Important recordings should always be secured with a backup.

This document contains various explanations of terms that are relevant in assessing the reliability of hard disk drives. Then their actual relevance to practice is derived based on two highly regarded studies of Google Inc. [4] and Carnegie Mellon University [6]. Based on latter, the functioning of various RAID systems and their advantages and limitations are considered.

HDD DURABILITY

Usually various information (AFR, MTBF, MTTF) from the respective manufacturer's data sheets is used for assessing the reliability of hard disk drives. Moreover, common methods (SMART) for early failure detection are considered.

As shown in the following, these theoretical considerations are limited to the assessment and planning of storage systems. They must be expanded to include observations and experiences from practice.

2.1 AFR, MTBF AND MTTF

AFR, MTBF and MTTF are the most common manufacturer information on the reliability of hard disk drives. They are based on experience, estimates and projections. Therefore, they can not be interpreted as absolute figures, but only as expectation or probability values.

2.1.1 Definition

2.1.1.1 AFR

The **AFR** (Annualized Failure Rate) is the percentage failure share of a certain amount of hard disks, which is extrapolated to one year based on expectation values [6].

2.1.1.2 MTBF

The **MTBF** (Mean Time Between Failures) specifies the expected operating time **between two consecutive failures** of a device type in hours (definition according to IEC 60050 (191)) [7].



Fig. 2-1 MTBF - Time between failures

The MTBF considers the life cycle of a device that fails repeatedly, is repaired and returned to service again.

This consideration can also be related to a failure without repair, as it is typically assumed for hard disks. In this case, the average operating time until the failure of the device is considered.

2.1.1.3 MTTF

The **MTTF** (Mean Time To Failure) specifies the expected operating time **until the failure** of a device type in hours (definition according to IEC 60050 (191)) [8].

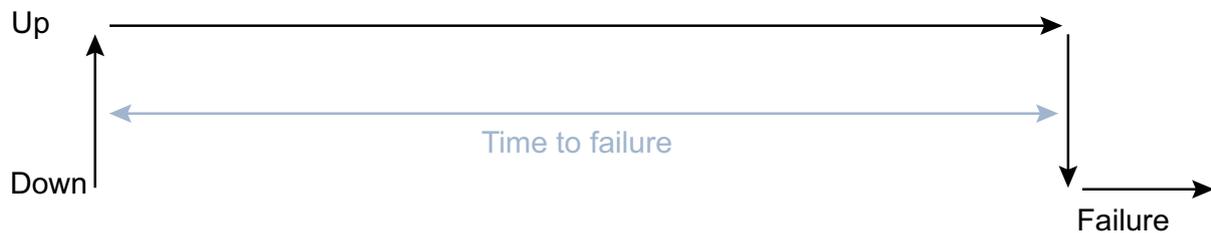


Fig. 2-2 MTTF - Time to failure

The expressions MTBF and MTTF are often used synonymously in terms of drives. Even the backronym mean time before failure has become customary for MTBF [7].

Seagate Technology LLC, for example, estimates the **MTBF** of its hard drives as the number of operating hours per year divided by the projected failure rate **AFR** [9]. This view is based on a failure without repair. **MTTF** would, thus, be the correct term.

2.1.2 Relation

AFR, MTBF and MTTF are values based on experience, estimates and projections. They can be set based on a singular view or calculated in relation and using various estimation methods (e.g. Weibull, WeiBayes) [9] page 1.

Simplified, and sufficient for the purposes of this document, the **MTTF** is the number of operating hours per year divided by the projected **AFR** [6].

2.1.3 Interpretation

Most manufacturers specify the reliability of their hard drives like Seagate by providing **AFR** and **MTTF**. As shown in the following, this value is always much better than the observations in practice.

The differences result from different **definitions of a failure**. A hard drive that is replaced by a customer due to peculiar behaviour, can be considered to be fully functional by the manufacturer. Seagate, for example, notes that 43% of the returned hard drives are functional [6]. Values between 15% and even 60% can be proven for other manufacturers [4].

In addition, the **environment conditions** in practice and in the manufacturer specific tests usually differ. In particular, higher temperature and humidity as well as increased utilization due to continuous read-write operations result in much higher failure rates in practice [6].

A useful interpretation of **AFR** and **MTTF** therefore only succeeds by considering the manufacturers procedures. As Adrian Kingsley-Huges [10] notes in his remark, the difference between observed and specified MTTFs can be found in their determination.

Simplified, the MTTF can therefore be calculated as follows:

$$\text{MTTF} = ([\text{test period}] \times [\text{number of HDDs}]) / [\text{number of failed HDDs}]$$

With a test period of 1,000 hours (approx. 41 days) and 1,000 HDDs and one failed HDD this results in:

$$\text{MTTF} = (1,000 \text{ hours} \times 1,000 \text{ HDD}) / 1 \text{ HDD} = 1,000,000 \text{ hours}$$

The projected annual failure rate AFR results from the reciprocal value:

$$\text{AFR} = ([\text{number of failures}] / ([\text{MTTF}] / [8,760 \text{ hours per year}]]) \times 100\%$$

$$\text{AFR} = (1 \text{ failure} / ([1,000,000 \text{ hours}] / [8,760 \text{ hours per year}])) \times 100\%$$

$$\text{AFR} = (1 \text{ failure} / 114.16 \text{ years}) \times 100\%$$

$$\text{AFR} = 0,86 \%$$

Conversely, one could calculate the MTTF based on an estimated AFR of 0.86%:

$$\text{MTTF} = (1,000 \text{ HDD} \times 8,760 \text{ hours}) / (1,000 \text{ HDD} \times 0.86\%)$$

$$\text{MTTF} = 8,760,000 \text{ HDD hours} / 8.6 \text{ HDD}$$

$$\text{MTTF} = 1,018,604 \sim 1,000,000 \text{ hours (114 years)}$$

What does a **MTTF of 1,000,000 hours** (114 years) express?

The assumption that an HDD can be operated 114 years is absolutely wrong! This value indicates **that the first failure would have to be expected after 1,000 hours (42 days) of operation if 1,000 hard drives were launched simultaneously.**

Looking at it another way you could say that **114 hard drives can be operated for a year and only one failure would have to be expected.**

2.1.4 Practical Relevance

MTTF and AFR are projected expectation values and, thus, represent a consistent failure behaviour. Observations in practice, however, show, that a relatively large proportion of the HDDs does not fail as expected, but much sooner or later.

2.1.4.1 Failure Rate - A Google Inc. Study

In their detailed study on HDD failure trend Pinheiro, Weber und Barroso (Google Inc.) evaluated more than 100,000 HDDs in a period of 9 month. The test group comprises HDDs with 80 to 400 GB, which were used in various systems of Google Inc. [4].

The authors note a trend towards uniformity with respect to the hard drive models in particular age groups. This might influence the absolute AFR slightly, but does not change the observable trend.

The hard disks were divided into groups according to their age. The hard drives that failed with the appropriate age were then put into relation with the corresponding group. [4].

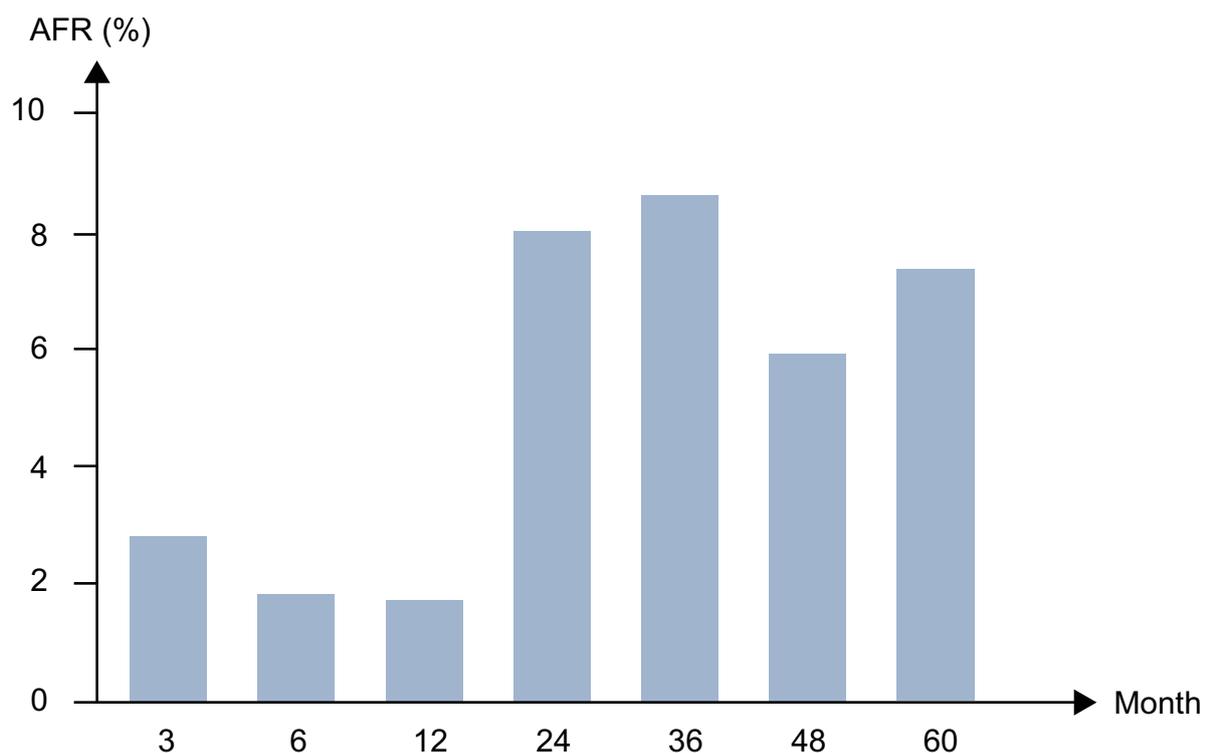


Fig. 2-3 AFR and age groups (Google Inc.)

Fig. 2-4The **failure rate** was between 1.7% for HDDs that failed with an age of 1 year and 8.6% for HDDs that failed with an age of 3 years. The observed **AFRs are consistently much higher than the values provided by the manufacturer.**

Another interesting aspect is, that **a relatively high proportion of the HDDs failed very early**: with 3 months (2.3%) or 6 months (approx. 1.9%). This result already shows the phenomenon of hard drives' "infant mortality", which is discussed in the following.

2.1.4.2 Failure Time - A Carnegie Mellon University Study

Schroeder and Gibson of Carnegie Mellon University found similar results in their study “Disk failures in the real world”. They evaluated the data of about 100,000 hard drives that were used in several large-scale systems [6].

They also found a large deviation of the manufacturer’s information (0,58% to 0,88%) from the observed failure rates (approx. 3% in average, up to 13% for single systems).

The **average failure rate** of all hard drives was **3.4 times higher than the highest specified AFR** of 0.88% [6]. The failure rate was 6 times higher for systems with an operating time of less than 3 years and even 30 times higher for systems with an operating time of 5-8 years.

The authors assessed that one-dimensional values such as MTTF and AFR do not reproduce the observations. Therefore, they focused on a more detailed analysis of the temporal distribution of failures.

First, they pointed to the generally accepted theory of the **bathtub curve**. This curve shows the theoretical failure rates of hardware products throughout the product life cycle [6] and could permit the prediction of the failure trend of hard drives in large-scale systems.

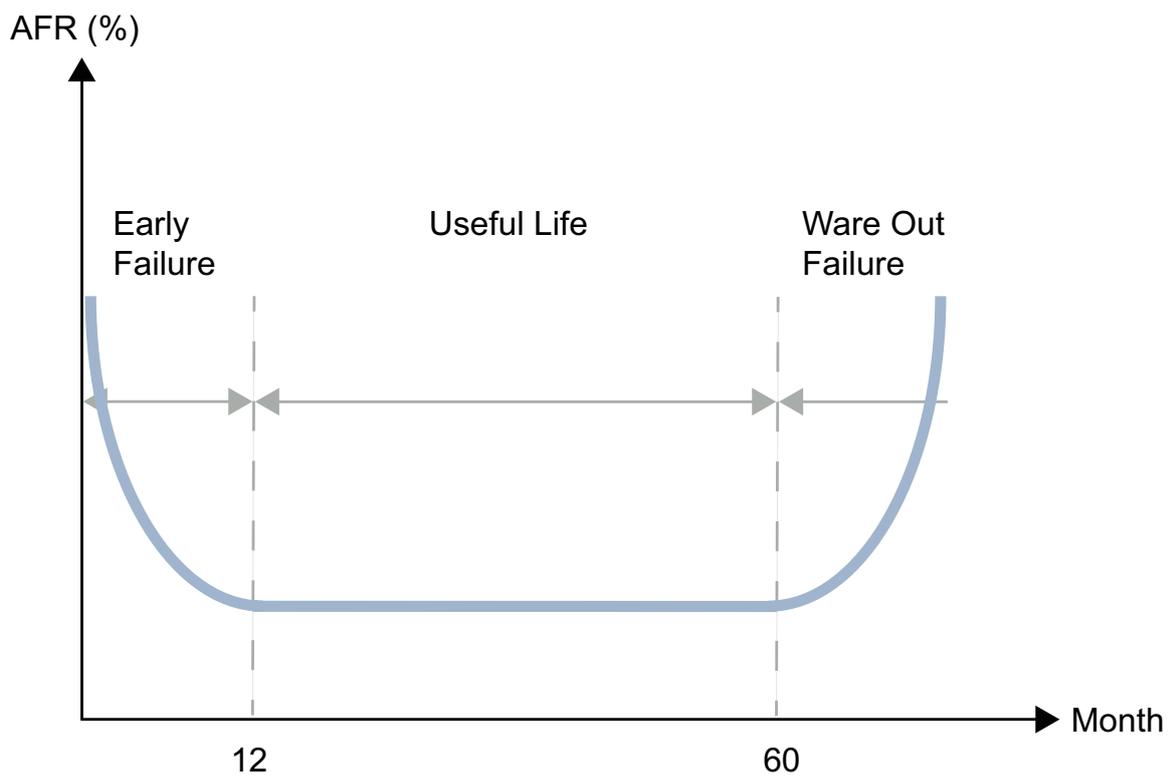


Fig. 2-5 Bathtub curve (theoretical model)

According to this graph, an increased failure rate in the first year would be observable for hard disk drives, followed by a period with a failure rate at a constant and low level. Towards the end of product life cycle the wear out has a strong effect, which would again lead to a rapidly increasing failure rate.

This **theoretical consideration** was **only partly confirmed by the trend in practice**. The following graph of the monthly distribution of failures in one of the evaluated systems shows a **relatively sharp delimitation of early failures** (infant mortality).

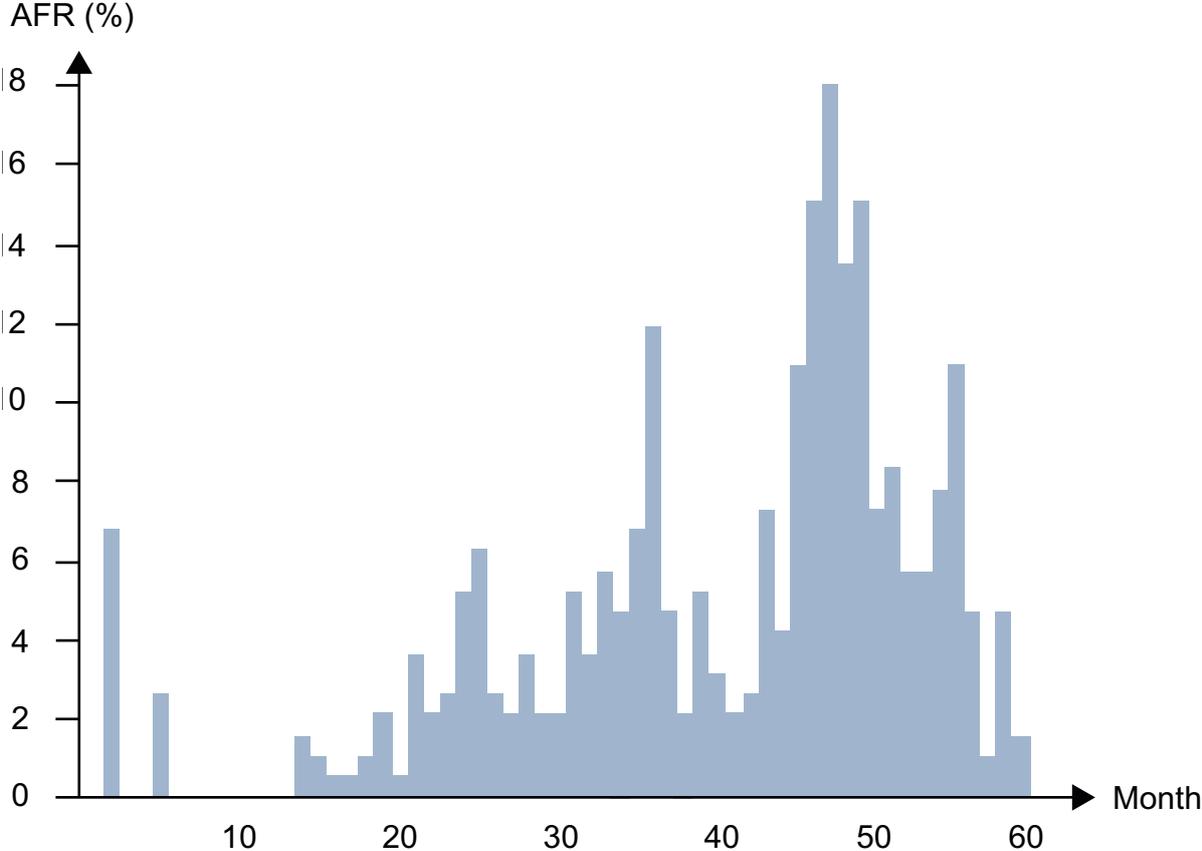


Fig. 2-6 Distribution of failure

It is, however, striking that the failure rate in the middle years does not settle to a relatively low value. The AFR begins to rise early and relatively constant. This observation suggests that **wear out has an early impact** and the failure rate linearly increases up to the end of the product life cycle [6].

2.1.4.3 Failure Rate - A Dallmeier electronic Analysis

The findings of the Google Inc. study were also confirmed by an internal evaluation of Dallmeier electronic. Here, the failures of hard disks that had been put on the market between January 2000 and December 2005 (74,000 units) were evaluated. First, the monthly failures for Wavelet and MPEG systems were identified.

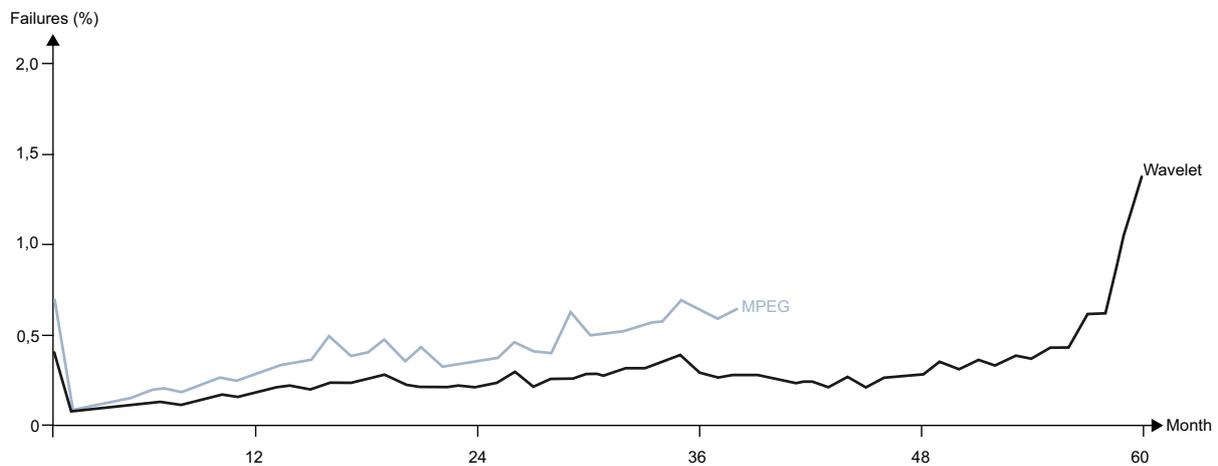


Fig. 2-7

In contrast to the average MTTF of 550,000 hours specified by the manufacturer, an actually observed mean value (87% Wavelet with 247 000, 13% MPEG with 111,000 hours) of 220,00 hours was calculated. This corresponds to an actual observed AFR of about 3.9%.

Based on this analysis, the availability of a system with 1,250 MPEG channels (on DIS-2 modules with a total of 2,500 hard disks) was considered. In this respect 200 disk failures per year would mathematically be expected. For the replacement of a DIS-2 module with a faulty HDD max. 2 minutes were supposed. The availability results from:

$$\begin{aligned} & ([\text{total operating time} - \text{repair time}] / \text{total operating time}) \times 100\% \\ & ([8.760 \text{ hours} \times 60 \text{ minutes} \times 1.250 \text{ channels} - 200 \times 2 \text{ minutes}] / \text{total operating time}) \times 100\% \\ & ([657.000.000 - 400] / 657.000.000) \times 100\% \\ & 99,99994\% \text{ availability} \end{aligned}$$

Referred to one channel of this system, an availability of 99.99994% results in an uptime of an annually 525,599.68 minutes and **a downtime of only 19.2 seconds** (0.32 minutes).

How favourable this value is, can be determined by using an example with an availability of only 99.5%. In this respect, one would have to expect an uptime of only 522,972 minutes and a downtime of already 2628 minutes (43.8 hours).

As this example illustrates as well, the MTTF specified by the manufacturers is substantially higher than the observed MTTF. However, a **system with mature technology and intelligent design** can be implemented with **high availability**.

2.1.5 Conclusion

The failure rate specified by the manufacturers with **MTTF/MTBF or AFR** is **usually too low**. A conservative planning should account for an **in average at least 3 times higher AFR**.

MTTF/MTBF provide no information on the distribution of failures. A conservative planning should always bear a **child mortality in the first months of operation** and increased **wear out failures towards the end of the product's life cycle** in mind.

The low and constant failure rate in the middle years of operation implied by the theoretical bathtub curve can not be confirmed in practice. A conservative planning should consider a **linearly increasing failure rate** due to wear out, which already **starts in the middle of the product's life cycle**.

2.2 SMART

2.2.1 Definition

SMART (Self-Monitoring, Analysis and Reporting Technology) is an industry standard that is built into almost all hard disks. SMART enables the permanent monitoring of important parameters, and, thus, early detection of an impending hard disk failure [11].

As a function SMART must be enabled for each hard disk drive in BIOS. The provided SMART values are evaluated by a software that is installed in addition to the operating system. The software can display warnings when manufacturer specific thresholds of individual parameters are exceeded. After prolonged usage expected failures can be predicted as well [11].

2.2.2 Interpretation

SMART provides values for many parameters, but Pinheiro, Weber and Barroso considered only four as significant for a failure prediction in their study for Google Inc. [4].

Scan Errors

Hard disks constantly test the surface of the magnetic discs and count the detected errors with a background function. A high number of errors is an indicator for a defective surface and, thus, for less reliability.

Reallocation Counts

When the hard disk finds a bad sector on the magnetic disk (during a read / write process or with a background function), the corresponding sector number is assigned to a new sector of the sector reserve. A high number of new assignments is an indicator for the wear of the magnetic disks.

Offline Reallocations

Offline Reallocations are a subset of the above described Reallocation. Only new assignments of bad sectors that are found by a background function are counted. Bad sectors and assignments that are discovered during a read / write process are not considered.

Probational Counts

Hard disks can put suspicious sectors “on probation”, until they fail permanently and are reassigned or until they continue to work without a problem. A high number of probations can be regarded as a weak error indicator.

When and if the values of a parameter trigger an alert of the monitoring software depends on the software and on the specifications of the manufacturer. To illustrate this complex analysis, the greatly simplified and reduced example of a 250 GB SATA is used in the following [11].

Parameter	Value (normalizes current measured value)	Worst (currently worst value)	Threshold (value should be greater)	RAW Value (actual measured value)	Remark
Relocation Counts	100	100	005	55	55 sectors have been replaced with reserve sectors due to failure. The drive estimates that there is no problem (the value is still 100) 0).
Seek Errors	100	100	067	0	Currently no read / write errors occurred.

Tabelle 2-1

The normalized measurement **Value** is counted down and triggers a warning upon reaching the **Threshold** limit. Although this example shows that 55 sectors have already been reallocated, the hard disk drive is still considered to be absolutely fine.

The failure detection by the SMART function of the hard disk drive is independent of the evaluation of the values by the SMART software, but decisive for a reliable failure forecast. If the detection does not work reliably, SMART may not be used as the sole tool for failure prediction of hard disk drives.

2.2.3 Practical Relevance

In their study for Google Inc. the authors evaluated SMART log files of more than 100,000 HDDs [4]. Nevertheless, they were **not able to develop a meaningful statistical model** for failure prediction [4].

Hereinafter, the possibility to create a **more simple prediction model** solely on the basis of SMART parameters was considered. But the analysis of the corresponding SMART values revealed, that they were not able to achieve a sufficient accuracy.

Of all failed hard disk drives, **56% showed no detected error** regarding all four strong SMART parameters. A forecast on this basis could therefore never predict more than half of the failures. Taking all other SMART parameters into consideration, 36% of the failed hard disk drives showed absolutely no errors (with no parameters!).

2.2.4 Conclusion

The conclusion of the authors is clear: „We conclude that it is unlikely that SMART data alone can be effectively used to build models that predict failures of individual drives.“ [4]. **SMART data alone can not be used to predict the failure of single hard disk drives.**

High values of a single parameter may cause an unnecessary exchange and therefore costs. Sudden failures without prior notification could lead to data loss due to lack of backups. This can lead to doubts about the reliability of the overall system, though actually only the SMART function failed.

As an alternative , the conservative maintenance based on the findings in point 1 remains. For systems with multiple hard disk drives, some protection may be achieved by using RAID systems, as described in the following.

STORAGE TECHNOLOGIES

Standard recording systems usually have one or several hard disk drives (JBOD) and are not able to compensate for the failure of a HDD.

High-quality recording systems redundantly store the audio and video data on several HDDs using a special storage technology (RAID) and are usually able to compensate for the failure of a HDD without data loss.

No matter what storage technology is used, it must always be taken into account, that a system can never be a substitute for a backup [5].

3.1 JBOD

JBOD (Just a Bunch Of Disks) refers to an undetermined number of hard disk drives (array) that are connected to a system (computer, recording system). They can be used by the operating system as individual drives or combined into one single logical drive [12].

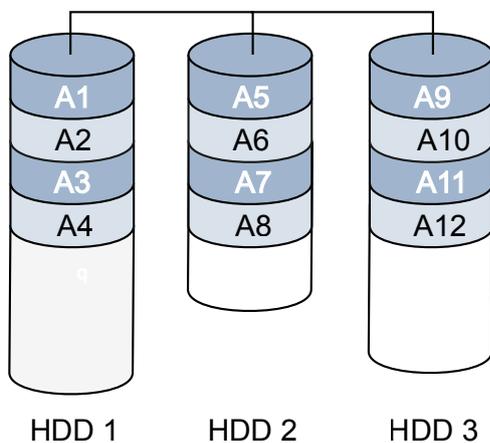


Fig. 3-1 JBOD with data (A1 to Ax)

Because a JBOD lacks all redundancy the expression is often used to distinguish a regular system from a RAID system.

3.1.1 Capacity and Costs

The net capacity of a JBOD array is as large as the sum of the capacities of the single hard disk drives. The net capacity, thus, corresponds to the total capacity of a system. A system of 8 hard disk drives with 2TB has a net capacity of 16TB. A JBOD system is the most economical system.

3.1.2 Safety and Rebuild

The behaviour when a hard disk fails varies with the JBOD systems from different manufacturers. Dallmeier JBOD recording systems have the advantage of continuing the recording when a HDD fails. The recordings on the remaining disks can be evaluated and secured as before.

3.1.3 Conclusion

JBOD is a simple and very cost-efficient storage system. But when a single hard disk drive fails, its recordings are lost.

3.2 3.2 RAID 1

A RAID 1 system consists of a combination of at least two hard disk drives (RAID array). The same data is simultaneously stored on all disks (mirroring). A RAID 1 system offers full redundancy [13].

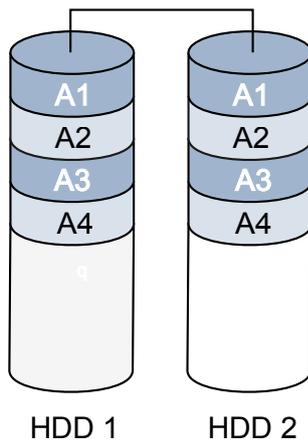


Fig. 3-2

3.2.1 Capacity and Costs

The net capacity of a RAID1 array is as large as the smallest hard disk drives. The total capacity of a system ideally is halved by mirroring the data, the storage costs are doubled. A system of 8 hard disk drives with 2TB has a net capacity of 8 TB.

3.2.2 Safety and Rebuild

If one of the mirrored HDDs fails, the recording is continued on the remaining HDD. After replacing the failed hard disk drive a rebuild process is started and the data is mirrored to the new HDD.

The failure of the intact HDD during the replacement or rebuild of the defective HDD inevitably leads to the loss of the data (if not mirrored on more than 2 HDDs). Since only a few HDDs are involved in a RAID 1, the probability of a simultaneous failure is relatively low, but can not be excluded.

3.2.3 Conclusion

RAID1 is a simple and relatively robust storage subsystem. The storage costs are relatively high, since the total capacity is always halved.

3.3 RAID 5

A RAID5 system consists of an array of at least three hard disk drives. The data is distributed and stored across all hard disk drives. In addition, parity data is generated and also distributed and stored.

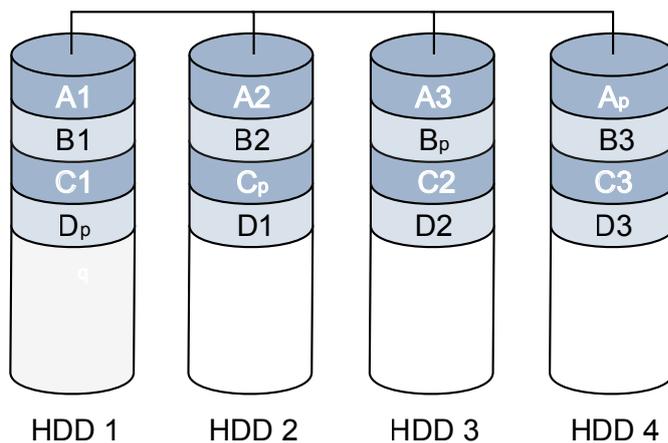


Fig. 3-3 RAID5 with distributed data (e.g. A1 to A3) and parity data (e.g. A_p)

If a hard disk drive fails, the parity data allows for the reconstruction of the lost data in combination with the remaining data [14].

3.3.1 Capacity and Costs

The capacity of a RAID5 array can be calculated as follows:

$$(\text{number of hard drives} - 1) \times (\text{capacity of the smallest hard drive})$$

A RAID5 array of 8 hard disk drives with 2TB has a net capacity of:

$$(8-1) \times 2\text{TB} = 14\text{TB}$$

If a spare HDD is used (see below), the formula must be adjusted:

$$(\text{number of hard drives} - 2) \times (\text{capacity of the smallest hard drive})$$

Unlike RAID1 a RAID5 system offers a better utilization of the total capacity of a system. Thus, a redundant data storage can be realized at relatively low cost.

3.3.2 Safety and Rebuild

If a HDD fails, the parity data allows for the reconstruction of the lost data on the replacement HDD in combination with the remaining data. The rebuild process starts automatically if a replacement HDD (spare HDD) is already integrated in the system. If this is not the case, it is started after the replacement of the defective HDD.

If a further hard disk drive fails during the replacement or rebuild of the defective HDD, the rebuild process can not be completed. This will lead to the loss of all data.

A RAID 5 usually consists of several HDDs. **The failure probability of a further HDD increases proportionally with their number.** It must also be noted that a rebuild may take several hours to days when using high-capacity HDDs. The critical period is relatively long.

In addition to the failure of another hard disk, an **unrecoverable read error** (URE) can also cause the failure of a rebuilding process. If a single fraction of the remaining or parity data can not be read, it can not be rebuild and the process usually stops.

The URE rate is the mean value for a hard disk drive model (not for a single drive) stated by the manufacturers. A typical **value of 10^{-14} Bit** means, that one unrecoverable read error occurs during the processing of 100.000.000.000.000 Bits (12TB).

Even with smaller RAID5 systems (e.g. RAID5 with 3 × 500 GB hard drives) the consideration of a URE of 10^{-14} Bit alone leads to a statistical **failure of the rebuild process in 8% of the cases** [17]. If larger HDDs are used, the occurrence of a URE is much more likely.

During the rebuild of a RAID array with 7 × 1TB hard disk drives the content of 6 HDDs (6TB) has to be read. With a URE of 10^{-14} Bit, the failure of the rebuild process would have to be expected in 50% of the cases [15].

3.3.3 Conclusion

RAID5 is a storage technology that allows for redundant data management at relatively low cost. But the risk of data loss is relatively high.

3.4 RAID 6

A RAID 6 system consists of an array of at least four hard disk drives. The data is distributed and stored on all hard drives. In the same way as with RAID 5, parity data is generated, distributed and stored, but in this case twice.

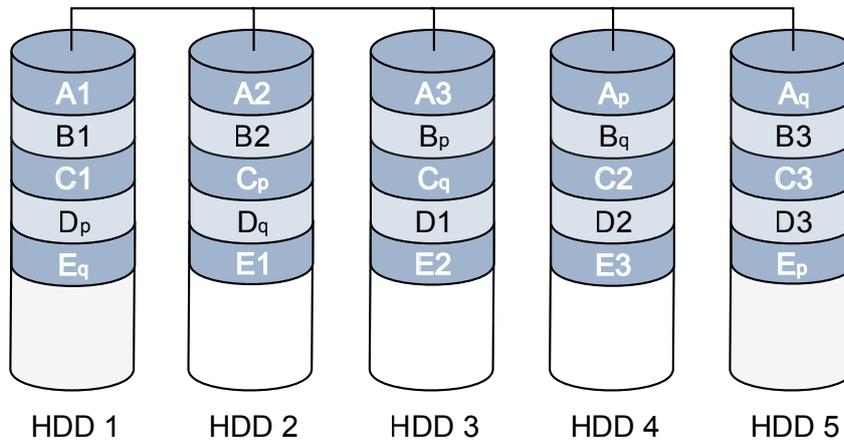


Fig. 3-4 RAID 6 with distributed data (e.g. A1 to A3) and double parity data (e.g. 2 x A_p)

The double parity data allow for a RAID 6 to compensate for the failure of up to two hard disk drives [16].

3.4.1 Capacity and Costs

The capacity of a RAID 6 array can be calculated as follows:

$$(\text{number of hard drives} - 2) \times (\text{capacity of the smallest hard drive})$$

A RAID 6 array of 8 hard disk drives with 2TB has a net capacity of :

$$(8-2) \times 2\text{TB} = 12\text{TB}$$

Unlike RAID 5 without a spare HDD, a RAID 6 system utilizes the total capacity of the system not quite as good. Nevertheless, a redundant data storage can be realized at relatively low cost.

However, if comparing a **RAID 6** system with a **RAID 5 system with spare HDD** (in practice most frequently used), the capacity consideration to be modified. In this case, both systems have an identical net capacity (in the chosen example 12TB) and **can be realized with the same storage costs**.

3.4.2 Safety and Rebuild

Generally, the problem of disk failure or unrecoverable read errors during the rebuild must be taken into account, as well, when regarding a **RAID 6** system. But the big advantage of a RAID 6 is its **tolerance of two failures**.

If a HDD fails, the parity data allows for the reconstruction of the lost data on the replacement HDD in combination with the remaining data. If a further hard disk drive fails during the replacement or rebuild of the defective HDD, this will not lead to the loss of the data. Simply put: the second set of parity data now allows for the reconstruction of the lost data to a second replacement HDD.

Like with RAID 5 systems, the probability of another HDD failure increases with the HDD number and the duration of the rebuild, which can take longer due to the RAID 6 dual parity calculation.

For all RAID systems, **the duration of the rebuild** depends on a variety of factors. Crucial aspects are the number and capacity of the hard disk, of course. Considering recording systems with comparable equipment, it also depends on the type of recording (SD or HD cameras, permanent or event-driven) and whether the recording is continued or discontinued.

A test series of Dallmeier electronic with comparable IPS systems at full capacity showed a **rebuild duration of approx. 2 hours per TByte for RAID 5 as well as for RAID 6 systems**. In practice hardly relevant, one can nevertheless find a slightly longer rebuild duration for RAID 6 systems. As a rule of thumb 25% to 35% can be expected.

Despite the slightly longer rebuild duration, RAID 6 has the distinct advantage of tolerating two hard disk failures. The risk of the losing of all data during a longer rebuild is so much lower than with RAID 5.

3.4.3 Conclusion

RAID 6 is a safer storage technology that allows for redundant data management at still relatively low cost. But the risk of data loss is relatively high. Compared to a RAID 5 system, however, the risk of data loss is relatively low. In general, RAID 6 can be considered as the superior storage system.

RECOMMENDATIONS

Planing

1. Note that JBOD systems are inexpensive, but offer no failure protection for individual hard disk drives.
2. RAID 1 systems are simple and relatively robust, but cause high storage costs.
3. RAID 5 systems cause lower storage costs than RAID 1 systems and tolerate the failure of one hard disk drive.
4. Note that a RAID 6 system causes the same storage costs as a RAID 5 system with spare HDD, but tolerates the failure of two hard disk drive.
5. RAID 6 currently is the superior storage system and offers maximum safety at a reasonable cost.
6. Note that no RAID system provides the same security as a backup of crucial recordings.

Maintenance

1. Include an at least three times higher failure rate than specified by the manufacturers in your planning.
2. Take the life time of the hard disk drives specified by the manufacturers into account, and replace still functioning hard disk drives early enough.
3. Consider a linearly increasing failure rate which already starts in the middle of the product life cycle of the hard drives.
4. Keep in mind, that failures increase in the first operating months and towards the end of the product life cycle of the hard drives.
5. Note that SMART data are not suitable for the failure prediction of individual hard disk drives.

Backup

1. No RAID system provides the same security as a backup. Important recordings should always be secured with a backup.
2. Backups can be conveniently carried out with the software SMAVIA Viewing Client. Some recording systems offer the possibility to remove all hard disks. They can be stored in a safe place and reconnected later.
3. Only backups provide an effective securing of recordings in cases such as:
 - defective files due to memory errors
 - accidental deletion of recordings
 - theft of the recording system
 - disasters such as fire, water damage, etc.
 - system malfunctions due to defective components or failure of the RAID controller

REFERENCES

- [1] Various authors, Hard disk drive, in http://en.wikipedia.org/wiki/Hard_disk_drive (2012.09.03)
- [2] Various authors, RAID, in <http://en.wikipedia.org/wiki/RAID> (2012.09.03)
- [3] Various authors, RAID, Data backup in <http://en.wikipedia.org/wiki/RAID> (2012.09.03)
- [4] Eduardo Pinheiro, Wolf-Dietrich Weber, Luiz André Barroso (Google Inc.), in Failure Trends in a Large Disk Drive Population (Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST'07), February 2007)
- [5] Christopher Negus, Thomas Weeks. The Mythos of RAID Backups, in Linux Troubleshooting Bible, Seite 100, Wiley Publishing Inc., 2004
- [6] Bianca Schroeder, Garth A. Gibson (Computer Science Department, Carnegie Mellon University), Age-dependent replacement rates, in Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you? (5th USENIX Conference on File and Storage Technologies, San Jose, CA), February 2007
- [7] Various authors, Mean Time Between Failures, in http://de.wikipedia.org/wiki/Mean_Time_Between_Failures (2012.08.16)
- [8] Various authors, Mean Time To Failures, in <http://de.wikipedia.org/wiki/MTTF> (2012.08.16)
- [9] Gerry Cole. (Seagate Personal Storage Group), Estimating Drive Reliability in Desktop Computers, Longmont Colorado, November 2000
- [10] Adrian Kingsley-Hughes, Making sense of „mean time to failure“ (MTTF), in <http://www.zdnet.com/blog/hardware/making-sense-of-mean-time-to-failure-mttf/310> (2012.08.20)
- [11] Various authors, Self-Monitoring, Analysis and Reporting Technology, in http://de.wikipedia.org/wiki/Self-Monitoring_Analysis_and_Reporting_Technology (2012.08.22)
- [12] Various authors, RAID / JBOD, in <http://de.wikipedia.org/wiki/RAID> (2012.08.28)
- [13] Various authors, RAID / RAID 1: Mirroring – Spiegelung, in <http://de.wikipedia.org/wiki/RAID> (2012.08.16)
- [14] Various authors, RAID / RAID 5: Leistung + Parität, Block-Level Striping mit verteilter Paritätsinformation, in <http://de.wikipedia.org/wiki/RAID> (2012.08.16)
- [15] Robin Harris, Why RAID 5 stops working in 2009, in <http://www.zdnet.com/blog/storage/why-raid-5-stops-working-in-2009/162> (2012.08.31)
- [16] Various authors, RAID / RAID 6: Block-Level Striping mit doppelter verteilter Paritätsinformation, in <http://de.wikipedia.org/wiki/RAID> (2012.08.16)
- [17] Various authors, RAID / Statistische Fehlerrate bei großen Festplatten, in <http://de.wikipedia.org/wiki/RAID> (2012.08.31)



HEAD & ACCOUNTS OFFICE

Dallmeier electronic GmbH & Co.KG
Bahnhofstr. 16
93047 Regensburg
Germany

tel +49 941 8700 0
fax +49 941 8700 180
mail info@dallmeier.com

www.dallmeier.com