Research Paper

2002-10-12

RAID Systems

Jimmy Persson (pt98jpr)
Gustav Evertsson (pt99gev)

Blekinge Institute of Technology, Sweden

Index

Introduction

Overall description

The purpose of this document is to give an over view of what RAID is, why we need it and how it works.

The research was performed during the course “Systems Engineering” at Blekinge Institute of Technology in Sweden.

History

[1]Back in the middle eighties SLED:s (Single Large Expensive Disk) were the most popular media for storing data. At that time, the disk drives did not by far have the storing capacity or the performance that disks have today. To be able to provide a large amount of data one had to have lots of disk drives, which all had to be mounted in a single file tree[2]. This was an extremely messy and inconvenient way of handling data. Disks these days were also very expensive, hence the meaning of SLED. Another big problem was, and still is, loss of data because of disk failure. A solution for this was much needed.

[3]To take care of these problems IBM co-sponsored Berkeley University of California to build a disk array subsystem to which IBM had received a patent in 1978[4]. In 1987 Randy Katz and Dave Patterson, both working at Berkeley University of California, had succeeded. They called the solution RAID[5]. RAID stand for “Redundant Array of Inexpensive Disks”, although some people chose to change “… Inexpensive …”, which is the original word, to “… Independent …”. Randy and Dave clustered multiple smaller and less expensive disks into an array. By doing this, all disks appear to the rest of the world as if there was just one single large disk. The result was compared to SLED:s according to cost versus performance. It turned out to that RAID had the same or superior performance as SLED, but with a theoretical Meantime Before Data Loss (MTBDL) that was reduced to an acceptable level.

[6]The search for a way of decreasing the MTBDL now started. There was need for a way to prevent single disk drive failures from causing data loss within the array of disks. The result was the six RAID levels 0 through 5. The different RAID levels determine how data is distributed across the drives in the array and the level of redundancy. A RAID system does not prevent hard drive failure. Hard drives always fail. RAID level 1 through 5 offer protection against the data loss caused by drive failure. RAID level 0 is a non-redundant method which only increases transfer speed.

[7]IBM shipped RAID 1 in 1990, and it was not until then RAID had its breakthrough. The year after (1991) IBM launched RAID 3 and 1992 they released RAID 5.

---

[1] De Re IT Corp, White Paper: RAID Systems.
[2] Rockwood, RAID Theory: An Overview, 2002.
[3] De Re IT Corp, White Paper: RAID Systems.
[4] Storage search, RAID manufacturers on STORAGE search .com.
IBM, IBM Hard disk drives | A history of IBM "firsts" in storage technology | 1973.
[5] Berkeley, RAID,
De Re IT Corp, White Paper: RAID Systems.
[6] De Re IT Corp, White Paper: RAID Systems.
[7] Storage search, RAID manufacturers on STORAGE search .com.
IBM, IBM Hard disk drives | A history of IBM "firsts" in storage technology | Timeline.

Issues

[8]So, basically, what RAID is all about, is to connect multiple disk drives and make them appear as if there was only one disk. Then better performance, storage capability and reliability are achieved, than can be done by using just a single disk drive.

[9]Now days there are three ways of accomplish a RAID system. It can be either hardware driven, software driven or a combination of both. The performance of hard ware driven RAID systems are generally higher than software driven ones.

[10]One issue with RAID systems is that all disks must have the same storing capacity to work properly. If they have not, the storage capacity of all disks will be restricted to have the same capacity as the smallest disk drive. If for we for instance have 3 disks with a capacity of 20 GB respectively, and one disk with a storing capability of 10 GB, the total storing capacity will be 40 GB (4x10 GB), since all disks will be restricted to a maximum storage of 10 GB.

Problem description

Problem

- Why RAID?
- How does RAID work?
- Are there any alternatives to RAID?

There are probably more questions to consider, but in this research these are the ones we have chosen.

[8] Solinap, RAID: An In-Depth Guide To RAID Technology, 2001.
[9] De Re IT Corp, White Paper: RAID Systems.
[10] De Re IT Corp, White Paper: RAID Systems.

Research

Why RAID?

RAID has for a long time been something that you only find in large server systems, but lately cheaper RAID controller card have made it possible to get a RAID system even for small servers and home computers. These will of course not have all the features, which the more expensive ones have.

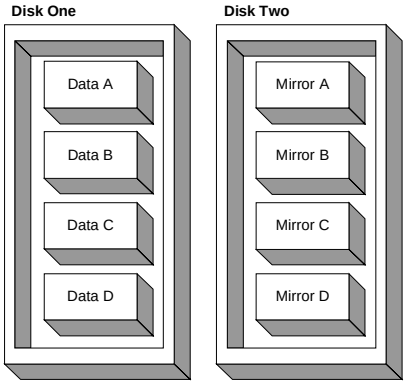
Different levels of RAID have different advantages and disadvantages. Therefore one must make an analysis of the workload before deciding what to buy. The choice also much depend on the quality attributes needed. Some examples of quality attributes one can get by using a RAID system is data redundancy, fault tolerance, increased capacity and increased performance[11].

How does RAID work?

The main idea behind RAID is, as mentioned in the introduction, to take some inexpensive disks and group them together, which will make the system see them as one single disk. This is done by using a RAID controller card that handle all I/O to the disks, and which knows where the stored data can be found.

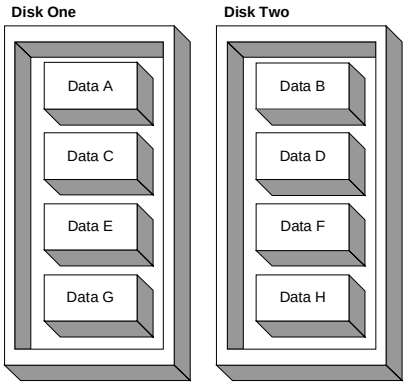
RAID works in three different ways to provide the quality attributes mentioned above. These ways are mirroring, striping and parity, of which each can be used either separately or mixed with one or more of the others. This is why RAID is divided into different levels.

Mirroring



The easiest way to get both availability and fault tolerance is to make a copy of all data on a second disk. This is called mirroring and you normally get one MB for every two MB of physical disk space. You will always have the second disk to read from if the other disk fails. The disadvantages of this method are waste of disk space and that you will not get higher write performance. You can however get higher read performance because reads can occur simultaneously on every drive[12].

Striping

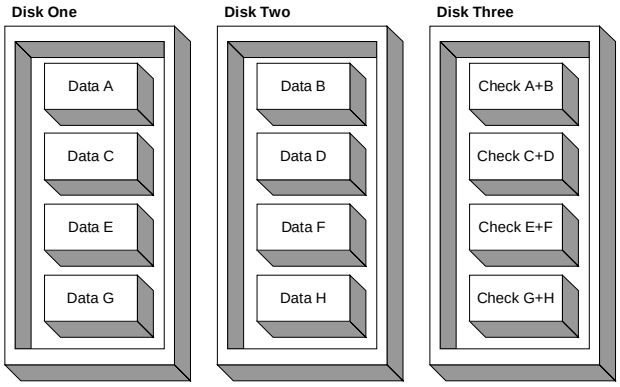


Whereas mirroring and parity deal with improvement of reliability, striping is used to get higher performance. The idea is to split data into small pieces, which then are distributed across the disks. This way the disks can work in parallel with different pieces of data. You will not lose any disk space as with mirroring and striping. One big disadvantage with this method is that if one disk breaks, all data will be lost. Therefore it is most often not used alone but in combination with mirroring or parity. Only such data that can be recreated by the application, such as

[11] RAID: An In-Depth Guide To RAID Technology, p 8.
[12] Adaptec, RAID White Papers and Benchmark Papers, What Is RAID?

cache or other temporary stored data is recommended to store using only striping.

Parity



Mirroring and striping are fairly easy to understand. Parity however is a bit more complicated. In the same way as with mirroring, it is used to improve the availability but without the waste of space. If you have X number of data elements, they can be used to create a parity. Then you end up with X+1 data elements. It is always possible to recover a lost element by using the others. The advantage with parity is of course that you have no single point of failure. However, to achieve this, it will cost a lot of computing power.

Different levels of RAID

Now when we know what RAID can do, we will look at how it is implemented in the different levels of RAID. The ones described below are the standard RAID levels. Some companies though, have developed their own levels of RAID.

RAID 0

Raid 0 is the simplest RAID level. It uses striping only to get higher performance. Therefore it is not really a RAID system, because it has no data redundancy. Because of that, it is not recommended to store important data with RAID 0. If one disk is lost, so will all data stored on that disk be as well.

RAID 1

RAID 1 is the easiest way to get data redundancy. It uses mirroring to store two copies of all data on two separate disks. Many computer systems need high availability without the requirement of more performance. In that case RAID 1 can be a good solution. Neither RAID 0 nor RAID 1 uses much computing power. The RAID controllers that handle these two RAID levels are therefore cheaper than controllers for higher RAID levels.

RAID 2

RAID 2 use striping and a special kind of redundancy technique, which is not described above. The technique used is bit level striping with Hamming code ECC. Separate disks are used for data storage and ECC. The Hamming codes are calculated and written to the ECC disks at the same time as data is written to its specific disk. The code is calculated again when data is read from the disks. This is done to check that it has not been changed since the time it was written. The complicated and expensive RAID controller hardware needed for this level of RAID, and the minimum number of hard drives required, is the reason this level is not used today[13].

RAID 3

This level uses data striping with parity. The parity is stored on dedicated disks. Parity makes it possible to, for example, store data on two disks with striping and have a third disk to store

[13] 4RAID, What is a RAID.

the parity. The performance is good but when data is written, the parity data will slow down the writing process.

RAID 4

RAID 4 is very similar to RAID 3. The difference is that it use larger blocks of data than RAID 3 does. This makes it possible to change the data stripe size to suit the applications needs. In the same way as with RAID 3, however, the extra parity disk will have a negative impact on performance.

RAID 5

This level is similar to RAID 4 in such a way that it uses data striping with larger blocks of data. The difference is that it tries to remove the bottleneck that RAID 4 has. That is done by skipping the dedicated parity disks. The parity is instead distributed amongst all disks. When the bottleneck is removed still one problem remains. It still needs to calculate the parity, and that is still more costly than for example mirroring. For fault tolerance and performance reasons, the data and parity are never stored on the same disk. This means that if one disk goes down, that data can always be recovered by using data on other disks to calculate what data disappeared. This level is one of the most used levels today, because many think it is the best combination of the quality attributes such as performance and fault tolerance.

Combinations of different levels

As explained above the different levels have different advantages. Then why not combine two levels and get the advantages from both? Well, that was the idea behind combinations such as 0+1, 1+0, 0+3, 3+0, 0+5, 5+0, 1+5, and 5+1. The most common one is 1+0 because it gives you both performance and good data redundancy without the need of complicated and expensive hardware that other combinations needs. The extra waste of space may be cheaper than the more expensive RAID controller.

Are there any alternatives to RAID?

Now when we know what RAID is and how it works, one may wonder if there are any other techniques that do the same or similar thing. As mentioned earlier in this document there are obvious advantages to RAID compared to other techniques, for instance SLED (Single Large Expensive Disk). However, RAID is not all good though. Since a RAID system is placed at one single geographical location, it is very vulnerable to disasters. If the room is flooded, all data on the disks may be lost. Moreover, if there is a temporary failure, because of a power outage, a hardware or software failure, etc., then the data on a RAID will be unavailable for the duration of the outage[14].

[15]At the Berkeley University of California they perform researches about alternative solutions. Such a solution is RADD, or Redundant Array of Distributed Disks. RADD: s can support redundant copies of data across a computer network at the same space cost as RAID: s do for local data. Such copies increase availability in the presence of both temporary and permanent failures (*disasters*) of single site computer systems as well as disk failures. As such, RADD: s should be considered as a possible alternative to traditional multiple copy techniques.

[14] Stonebraker, Schloss, DISTRIBUTED RAID -- A NEW MULTIPLE COPY ALGORITHM, http://s2k-ftp.cs.berkeley.edu:8000/postgres/papers/ERL-M89-56.pdf.
[15] Stonebraker, DATA BASE RESEARCH AT BERKELEY, http://s2k-ftp.cs.berkeley.edu:8000/postgres/papers/sigmodr90-ucb.pdf.

Furthermore, RADD: s are also candidate alternatives to high availability schemes such as *hot standbys*.

[16]This is a new technology, and it is still under construction. There are some issues that must be handled to make RADD useful. A good thing is that all RAID algorithms work directly in a distributed environment; however distribution generates its own significant problems, which must be handled. Other things to consider are the fact that algorithms must be able to handle unequal amount of storage at each site, it must be possible to mix different group sizes in the same network and final, it must be possible to subtract disk storage from RADD without the need of a global reconstruction.

[17]There are other technologies that may be alternatives to RAID. These are for instance: 2D-RADD, C-RAID, RO WB-optimistic and RO WB-pessimistic.

[16] Stonebraker, DATA BASE RESEARCH AT BERKELEY, http://s2k-ftp.cs.berkeley.edu:8000/postgres/papers/sigmodr90-ucb.pdf.
[17] Stonebraker, DATA BASE RESEARCH AT BERKELEY, http://s2k-ftp.cs.berkeley.edu:8000/postgres/papers/sigmodr90-ucb.pdf.

Conclusion

There are some disadvantages about RAID. They are however few compared to the advantages, like the high availability and performance. These are things that have become more important, especially since companies to day use Internet for making business on a global market. One must not forget that technology and techniques are developed and fine tuned every day. This means that alternative technologies, like RADD, may be a complement to RAID or even replace it.

The idea behind RAID is rather simple. It is important to understand it, at least a high level. This is important knowledge have to be able to make a good decision about what RAID level to use. The choice of level may have great impact of how well the system will work.

Nowadays RAID controllers have become cheaper. Therefore RAID has become available to a bigger market. It is even possible to find it in ordinary workstations and small company servers.

Source

Internet

4RAID, What is a RAID, http://www.4raid.com/raidlevels.htm.

Adaptec, RAID White Papers and Benchmark Papers, What Is RAID?, http://www.adaptec.com/worldwide/product/markeditorial.html?sess=no&cat=%2fTechnology%2fRAID&prodkey=raid_wp_benchmark&type=Technology.

Berkeley, RAID, http://www.sims.berkeley.edu/courses/is257/f99/Lecture10_257/sld025.htm.

IBM, IBM Hard disk drives | A history of IBM "firsts" in storage technology | 1973, http://www.storage.ibm.com/hdd/firsts/n1973.htm.

IBM, IBM Hard disk drives | A history of IBM "firsts" in storage technology | Timeline, http://www.storage.ibm.com/hdd/firsts/timeline.htm.

De Re IT Corp, White Paper: RAID Systems, http://www.dere.com/en/wp/raid.html.

Rockwood, Ben, RAID Theory: An Overview, The Cuddletech Veritas Volume Manager Series CDLTK Cuddletech, http://www.cuddletech.com/veritas/raidtheory/x26.html, April 08th 2002.

Solinap, Tom, RAID: An In-Depth Guide To RAID Technology, http://www.slcentral.com/articles/01/1/raid/, January 24th 2001.

Storagesearch.com, RAID manufacturers on STORAGE search.com, http://www.storagesearch.com/raid.html.

Written

Stonebraker, Michael, DATA BASE RESEARCH AT BERKELEY, Department of Electrical Engineering and Computer Science University of California Berkeley, CA 94720, http://s2k-ftp.cs.berkeley.edu:8000/postgres/ papers/sigmodr90-ucb.pdf.

Stonebraker, Michael † - Schloss, Gerhard A. ‡, DISTRIBUTED RAID -- A NEW MULTIPLE COPY ALGORITHM, † EECS Department, CS Division, ‡ Walter A. Haas School of Business, University of California, Berkeley Berkeley, CA 94720, http://s2k-ftp.cs.berkeley.edu:8000/postgres/papers/ERL-M89-56.pdf.