

Internal distribution code:

- (A) Publication in OJ
(B) To Chairmen and Members
(C) To Chairmen
(D) No distribution

**Datasheet for the decision
of 12 April 2013**

Case Number: T 0743/11 - 3.5.01

Application Number: 05738175.8

Publication Number: 1745395

IPC: G06F 17/30

Language of the proceedings: EN

Title of invention:

Extension of a system and method for write allocation within a write anywhere file layout file system

Applicant:

Network Appliance, Inc.

Headword:

Write allocation/NETWORK APPLIANCE

Relevant legal provisions:

EPC Art. 123(2)

Relevant legal provisions (EPC 1973):

EPC Art. 84

EPC Art. 56

Keyword:

"Inventive step - (yes)"



Case Number: T 0743/11 - 3.5.01

D E C I S I O N
of the Technical Board of Appeal 3.5.01
of 12 April 2013

Appellant: Network Appliance, Inc.
(Applicant) 495 East Java Drive
Sunnyvale, CA 94089 (US)

Representative: Rupprecht, Kay
Meissner, Bolte & Partner GbR
Widenmayerstraße 47-50
D-80538 München (DE)

Decision under appeal: Decision of the Examining Division of the
European Patent Office posted 11 October 2010
refusing European patent application
No. 05738175.8 pursuant to Article 97(2) EPC.

Composition of the Board:

Chairman: S. Wibergh
Members: P. Scriven
P. Schmitz

Summary of Facts and Submissions

- I. The appeal is against the Examining Division's decision to refuse European patent application 05738175.8. The Examining Division decided not to admit the appellant's main, and first and second auxiliary requests, and refused the application because there were no admissible requests on file. The main request was not admitted because it failed to overcome previous objections as to lack of clarity (Article 84 EPC), added subject-matter (Article 123(2) EPC), and lack of inventive step (Article 56 EPC). The two auxiliary requests, which were filed during oral proceedings before the division, were not admitted because they were filed late, and, prima-facie, seemed not to overcome the objections of lack of clarity, added subject-matter, and lack of inventive step.
- II. With the statement setting out its grounds of appeal, the appellant filed a new main and two new auxiliary requests. The appellant also requested reimbursement of the appeal fee on the basis of a number of alleged procedural errors, and, if the Board were unable to order the grant of a patent on the basis of its written submissions, that oral proceedings be held.
- III. The Board sent a communication under the terms of Rule 100(2) EPC, in which it presented its provisional analysis of the alleged procedural violations. A number of issues regarding clarity and added subject-matter were also identified. In particular, the Board suggested that the invention might have to be limited to the context of write anywhere file layout (WAFL)

- systems, and that, if that were done, the prior art set out on page 5 of the description might be the most appropriate starting point for the assessment of inventive step.
- IV. The appellant responded by filing a new request and arguments regarding clarity, support for amendments, novelty, and inventive step. In particular, the appellant discussed the disclosures of documents D1 - D4, mentioned in the decision under appeal.
- V. The Board arranged for oral proceedings to be held, and sent a summons accordingly. In the accompanying communication, the Board raised a number of issues concerning compliance with Articles 84 and 123(2) EPC, and made some observations on the disclosures of documents D1 - D4.
- VI. With the letter dated 12 March 2013, the appellant filed a new main and a new auxiliary request. The second auxiliary request, filed with the statement of grounds, was maintained.
- VII. Oral proceedings were held as scheduled on 12 April 2013. After discussion of various issues with the Board, the appellant filed an amended set of claims 1 - 16 and a complete set of amended description pages 1-7, 7a, 8 - 31, as a new main request.
- VIII. The appellant's final requests were that the decision under appeal be set aside, and that a patent be granted on the basis of claims 1 to 16 and the description

filed during oral proceedings before the Board (main request), or of the first auxiliary request filed with the letter dated 12 March 2013. The second auxiliary request and the request for reimbursement of the appeal fee were withdrawn.

IX. Claim 1 according to the main request reads as follows:

A computer-implemented method for performing write allocation in a storage system (120), the storage system storing files using buffer trees as an internal representation of blocks for a file, the buffer trees comprising parent blocks which reference data blocks used to store data of the file, the method comprising the steps of:

layering one aggregate on top of a RAID plex (550) comprising at least one RAID group (560), each RAID group comprising a plurality of storage disks (530), the aggregate having its own physical volume block number (pvbn) space and configured to maintain metadata for the aggregate in the form of block allocation structures within the pvbn space;

organizing a plurality of virtual volumes (vvol) on the aggregate (500) in container files, each vvol having its own virtual volume block number (vvbn) space and configured to maintain metadata for the vvol in the form of block allocation structures within the vvbn space of the vvol, wherein each container file in the aggregate contains all blocks owned by one vvol, wherein for each vvol a vvbn identifies a file block number (fbn) location within the container file of the vvol, wherein when operating in a vvol the storage system uses indirect blocks of the aggregate for the respective container file to translate the fbn into a pvbn; and

allocating a first block to a vvol (510) of the plurality of vvols and to said aggregate, wherein allocating a first block to a vvol (510) further comprises selecting a physical volume block number (pvbn) for the first block from the pvbn space (720) of the aggregate (500) and a virtual volume block number (vvbn) for the first block from the vvbn space (730) of the vvol; adjusting the block allocation structures of the aggregate (500) to record the selected pvbn and the block allocation structures of the vvol to record the selected vvbn; and updating a parent block of the first block with the selected pvbn as a block pointer, the parent block being an indirect block of a file of the vvol or an inode of the file of the vvol.

X. The appellant's arguments can be summarised as follows.

The invention built on the prior art "write anywhere file layout" (WAFL) system, which was set out in the description. In that system, each logical disk was assigned a fixed area of "physical" storage, and that lacked flexibility. It was, for example, difficult to change the size of a logical disk.

The invention added an "aggregate", which was the totality of the "physical" storage and which had its own address space. The addresses of the aggregate had fixed allocation to the "physical" storage, but the virtual volumes could be assigned flexibly on the aggregate.

The aggregate used a buffer tree and maintained metadata for keeping track of the allocations of its blocks, just as the logical disks did in the prior art.

Two allocation structures were therefore maintained, together with a mapping between them. However, in order to avoid conversions when reading data, pointers to the aggregate addresses were maintained in the inodes and indirect blocks of the virtual volumes.

The prior art disclosed neither the aggregate, nor the storing of pointers to its address space in the inodes or indirect blocks of a virtual volume.

The word "physical" in the term "physical volume block number" was not to be taken literally. The history of data storage had repeatedly seen physical storage replaced by something that looks the same to the system that used it, but to which the term "physical" no longer applied in any strict sense. That was something the skilled person understood. All that mattered for the present invention was that the virtual volumes each have an address space, and that the aggregate has its own distinct address space.

Reasons for the Decision

1. *Introduction*

- 1.1. The invention concerns an extension to the write anywhere file layout (WAFL) system. In that system, a user's data are stored in logical volumes. A logical volume looks, to the user, like a storage disk, but it is not. Underlying the logical volume is a further storage device.

- 1.2. When data are written in a WAFL system, blocks of the data are assigned block numbers in the logical volume. These logical block numbers are associated with block numbers in the underlying storage. Similarly, when data are read, the user looks for the contents of some particular logical blocks, but actually gets the contents of the underlying blocks.
- 1.3. The underlying storage may, for example, be a physical disk, in the sense of a rotating device with magnetic domains that can be put in one or another state so as to represent stored data. It may, however, be anything that presents the logical volume with the same behaviour. So long as the logical volume sees the behaviour it expects, it will happily "store" data for the user. It does not matter whether the underlying storage really is physical in the sense just outlined, although there must ultimately be some physical storage.
- 1.4. One possibility for the underlying storage is a RAID system. In such a system, data are distributed redundantly over several disks. A user's data block would then be "written" to a logical block, which corresponds to a RAID block, which corresponds to a number of copies in possibly different forms at a number of different locations in different disks.
- 1.5. As explained by the appellant, in prior art WAFL implementations, there was a fixed relationship between logical blocks numbers and RAID addresses. Each logical volume would use a fixed part of the RAID system.
- 1.6. The WAFL system uses a buffer tree to keep track of how data are stored. Each data file is associated with an

inode, which comprises pointers to the logical blocks at which the data blocks are stored. If the file is big enough, some of the pointers will not point directly to data, but to "indirect blocks", containing further pointers. By following the pointers from an inode, if necessary passing via indirect blocks, the logical blocks in which the data are stored can be found. Via the fixed relationship between logical blocks and RAID addresses, the logical blocks so identified correspond to specific locations in the RAID system.

2. *Main request, Article 123(2) EPC.*

2.1. In the communication accompanying the summons to oral proceedings, the Board objected that the application as filed did not, apparently, support parent blocks that were neither inodes nor indirect blocks. The parent blocks, according to the claims of the present main request, are restricted to inodes or indirect blocks, and the Board considers that this objection has been overcome.

2.2. The Board also objected that the application as filed supported the layering of an aggregate on a RAID system, but not on a single RAID plex. The amended independent claims still refer to one plex, but the Board is satisfied that there is no added subject matter. That is because each RAID plex is a complete copy of every other, so that if the aggregate is layered on one, it is, by implication, layered on all.

2.3. In its first communication, the Board also noted that the appellant had cited bases in different parts of the description for individual features of the independent

claims. The Board is now satisfied that this does not create a problem under Article 123(2) EPC. The description sets out the basic idea of the invention in terms of the aggregate and its buffer tree structure (see the paragraph bridging pages 18 and 19 of the published application). It further sets out two embodiments of the buffer tree structure, the "hybrid vvol embodiment" and the "dual vbn hybrid vvol embodiment" (see page 19 of the published application, first and second complete paragraphs). Apart from some hints at a third embodiment (page 30, line 10 - page 31, line 2), which have now been removed, the rest of the description, from the middle of page 18 on, deals with those two embodiments. The present independent claims cover both embodiments, and do not combine features that belong only to the one with features that belong only to the other.

2.4. The Board sees no other reason to doubt that the amendments comply with Article 123(2) EPC.

3. *Main request, Article 84 EPC, clarity*

3.1. The Board is satisfied that the terminology of WAFL systems was known to, and would have been understood by, the skilled person. Thus, terms such as "inode" do not require definition in the claims.

3.2. The meanings of "physical" and of "aggregate" do require some extra comment. The former can be confusing, because it is sometimes used with its everyday meaning, so that "physical disk" would be something one could pick up, and sometimes with a more abstract meaning (see point 1.3, above), as something that presents

appropriate behaviour to some higher layer. The Board is satisfied that the skilled person, who is familiar with WAFL storage, would understand the distinction, and that the more abstract meaning is appropriate for something layered on a RAID plex. Once that has been grasped, the skilled person would correspondingly interpret the term "physical volume block number space," and understand that the aggregate is nothing other than an additional virtualization layer, which covers the whole of the underlying RAID plex, which has its own space of block numbers, and which maintains its own block allocation structures, just as the virtual volumes do.

3.3. The Board, therefore, considers that the claims are clear.

4. *Main request, Article 56 EPC 1973*

4.1. Claim 1 defines a method of write allocation in which two virtualization layers (the aggregate and the virtual volumes) sit on a RAID plex. Each of them has its own space of block addresses and each maintains its own block allocation structure. When a block is allocated in a virtual volume, a parent block (which will be an inode or an indirect block) is updated so as to include the corresponding block number in the aggregate.

4.2. The Board and the appellant agree that the prior art WAFL system, as set out in the application, is the most appropriate starting point for the assessment of inventive step. That prior art system had a single virtualization layer, and, therefore, a single block

allocation structure. The pointers in its inodes and indirect blocks were to blocks within the same virtualization layer, and it was the job of the underlying RAID system to translate those blocks numbers to its own representation.

4.3. In contrast to that prior art, the invention defined in claims 1 and 10 has the aggregate layered on top of the RAID system. The virtual volumes now sit inside container files on the aggregate. In effect, there are, according to the invention, two layers of virtualization. That allows a more flexible use of the underlying RAID system by the virtual volumes.

4.4. Such a two-layer structure is not part of the prior art WAFL system. Nor do any of the documents cited in the decision under appeal disclose one. In this respect, it is significant to note that the underlying RAID system cannot itself be regarded as a virtualization layer in the way the aggregate is. That is because the RAID system does not concern itself with block allocation structures. That is the job of the file system that makes use of the RAID system.

4.5. In the Board's view, the simple addition of a new layer of virtualization would not involve an inventive step, because successive virtualization has been an important factor in the development of storage devices. Nor, once the decision to add a new virtualization layer has been taken, would the use of the same sort of block allocation structures as those already used be anything more than an obvious choice. However, the invention does more than add a virtualization layer using the same sort of allocation structures as are already in

use. During write allocation, it puts (cross-layer) pointers to aggregate allocation structures in the inodes and indirect blocks of the virtual volumes. As a result, when the data are later read, there is no need to convert from the block structures of the virtual volume to the underlying block structures of the aggregate. The saving during reading is at the cost of more processing during writing, and of some storage capacity. That is part of the sort of trade-off that an engineer routinely makes, but the Board does not consider that every manifestation of a trade-off is obvious simply because the factors traded off are known. In the present case, while it would have been obvious to the skilled person to trade off ease of reading against ease of writing and cost in storage, there is nothing to suggest that she would have considered using cross-layer pointers in doing so. None of the prior art suggests cross-layer pointers at all.

4.6. The Board concludes that the subject matters defined in claims 1 and 10 according to the main request do involve an inventive step (Article 56 EPC 1793).

5. *Main request, further requirements*

The Board is satisfied that the application and the invention to which it relates also meet the other requirements of the EPC.

Order

For these reasons it is decided that:

1. The decision under appeal is set aside.

2. The case is remitted to the department of first instance with the order to grant a patent on the basis of the following documents:
 - claims 1 to 16 as filed during oral proceedings before the board,
 - description as filed during oral proceedings before the board, and
 - drawings as originally filed.

The Registrar:

The Chairman:

T. Buschek

S. Wibergh