

**Internal distribution code:**

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

**Datasheet for the decision  
of 11 September 2023**

**Case Number:** T 0790/20 - 3.5.07

**Application Number:** 15911754.8

**Publication Number:** 3264285

**IPC:** G06F17/30, G06F3/06, H03M7/30

**Language of the proceedings:** EN

**Title of invention:**  
Data deduplication method and storage device

**Applicant:**  
Huawei Technologies Co., Ltd.

**Headword:**  
Data de-duplication/HUAWEI TECHNOLOGIES

**Relevant legal provisions:**  
EPC Art. 84  
RPBA 2020 Art. 12(4), 13(2)

**Keyword:**

Claims - clarity - main request, first, second and fourth auxiliary requests (no)

Amendment to case - third auxiliary request - amendment admitted (no)

Amendment after summons - fourth auxiliary request - exceptional circumstances (yes)



**Beschwerdekammern**  
**Boards of Appeal**  
**Chambres de recours**

Boards of Appeal of the  
European Patent Office  
Richard-Reitzner-Allee 8  
85540 Haar  
GERMANY  
Tel. +49 (0)89 2399-0  
Fax +49 (0)89 2399-4465

Case Number: T 0790/20 - 3.5.07

**D E C I S I O N**  
**of Technical Board of Appeal 3.5.07**  
**of 11 September 2023**

**Appellant:** Huawei Technologies Co., Ltd.  
(Applicant) Huawei Administration Building  
Bantian  
Longgang District  
Shenzhen, Guangdong 518129 (CN)

**Representative:** Körber, Martin Hans  
Mitscherlich PartmbB  
Patent- und Rechtsanwälte  
Karlstraße 7  
80333 München (DE)

**Decision under appeal:** **Decision of the Examining Division of the European Patent Office posted on 11 December 2019 refusing European patent application No. 15911754.8 pursuant to Article 97(2) EPC**

**Composition of the Board:**

**Chair** J. Geschwind  
**Members:** P. San-Bento Furtado  
R. de Man

## **Summary of Facts and Submissions**

- I. The appeal lies from the decision of the examining division to refuse European patent application No. 15911754.8.
- II. The following document was cited in the appealed decision:  
D1: WO 2012/067805 A2, 24 May 2012.
- III. The examining division decided that the subject-matter of the independent claims of the main request was not novel over document D1, that the subject-matter of the dependent claims of the main request lacked novelty or inventive step, and that the subject-matter of the claims of the first and second auxiliary requests was not inventive.
- IV. With its statement of grounds of appeal, the appellant maintained the refused requests and submitted a set of claims of a third auxiliary request.
- V. In a communication accompanying the summons to oral proceedings, the board expressed its preliminary opinion that the subject-matter of claim 1 of each of the main request and first and second auxiliary requests was not clear and not inventive over document D1 and that the third auxiliary request should not be admitted into the proceedings.
- VI. With a letter dated 5 May 2023, the appellant filed a fourth auxiliary request and arguments in favour of patentability of the fourth auxiliary request.
- VII. In a further letter, the appellant withdrew its request for oral proceedings, which were then cancelled.

VIII. The appellant's final requests are that the contested decision be set aside and that a patent be granted on the basis of the main request or one of the first to fourth auxiliary requests.

IX. Claim 1 of the main request reads as follows (itemisation added by the board):

"A deduplication method, comprising:

- (a) receiving, by a storage device, a first data stream;
- (b) dividing, by the storage device, the first data stream to obtain  $n$  data blocks, wherein logical addresses of the  $n$  data blocks are contiguous, the  $n$  data blocks comprise a first data block, a logical address of the first data block is a head address in the logical addresses corresponding to the  $n$  data blocks, and  $n$  is an integer not less than 2;
- (c) calculating, by the storage device, the  $n$  data blocks to obtain fingerprints of the  $n$  data blocks;
- (d) contiguously storing, by the storage device, the  $n$  data blocks in a first storage area in a sequence of the logical addresses of the  $n$  data blocks when the fingerprints of the  $n$  data blocks are not found in fingerprints in the storage device, wherein
  - (d1) a physical address of the first data block stored in the first storage area is a first physical address;
- (e) contiguously storing, by the storage device, metadata of each of the  $n$  data blocks in a second storage area in the sequence of the logical addresses of the  $n$  data blocks, wherein

- (e1) metadata of each of the n data blocks comprises a respective fingerprint in the fingerprints of the n data blocks and a physical address of the respective data block, corresponding to the fingerprint, that is stored in the first storage area;
- (f) establishing, by the storage device, a mapping between an address identifier of the metadata of each of the n data blocks and the metadata of the respective data block; and
- (g) establishing, by the storage device, a mapping between the logical address of the first data block and an aggregation address, wherein
  - (g1) the aggregation address comprises a physical address of an aggregation data block and an address identifier of metadata of an aggregation fingerprint,
  - (g2) the aggregation data block comprises the n data blocks,
  - (g3) the physical address of the aggregation data block comprises the first physical address and physical address lengths of the n data blocks stored in the first storage area, and
  - (g4) the address identifier of the metadata of the aggregation fingerprint comprises an address identifier of metadata of the first data block and a quantity of address identifiers of metadata of the n data blocks."

X. Claim 1 of the first auxiliary request differs from claim 1 of the main request in that it adds the following text at the end of the claim:

"wherein the establishing, by the storage device, a mapping between the logical address of the first data block and an aggregation address specifically comprises:

establishing, by the storage device, a mapping between the logical address of the first data block and both the physical address of the aggregation data block and the address identifier of the metadata of the aggregation fingerprint."

XI. Claim 1 of the second auxiliary request differs from claim 1 of the main request in that it adds the following text at the end of the claim:

"wherein the method further comprises:

establishing, by the storage device, an index of a first fingerprint in the fingerprints of the n data blocks in the first data stream, wherein the index of the first fingerprint comprises a mapping between the first fingerprint and an address identifier of metadata of the first fingerprint."

XII. Claim 1 of the third auxiliary request differs from claim 1 of the main request in that it adds the following text at the end of the claim:

"wherein the method further comprises:

before the establishing, by the storage device, a mapping between the logical address of the first data block and an aggregation address, determining, by the storage device, that the physical address lengths of the n data blocks stored in the first storage area do not exceed a compression window of the storage device, wherein the compression window refers to a length of data blocks that can be compressed at a time."

XIII. Claim 1 of the fourth auxiliary request differs from claim 1 of the main request in that the text "A deduplication method" has been replaced with "A method for storing non-duplicate data" and in that the text items (d) and (g3) have been replaced with the following text items (d') and (g3'):

(d') "contiguously storing, by the storage device, the  $n$  data blocks in a first storage area in a sequence of the logical addresses of the  $n$  data blocks when no identical fingerprint is found in fingerprints in the storage device after querying whether a fingerprint of the fingerprints of the  $n$  data blocks is the same as any fingerprint that is stored in the storage device, wherein"

(g3') "the physical address of the aggregation data block comprises the first physical address and an address length of physical blocks in which the  $n$  data blocks stored in the first storage area, wherein the address length of the physical blocks is a range from the first physical address to an end physical address wherein a  $n$ th data block of the  $n$  data blocks [sic] stored, and"

## Reasons for the Decision

### *Application*

1. The application concerns data de-duplication for reducing the amount of storage capacity needed to store data (see translation of original description, page 1, lines 3 to 8).
2. *Clarity - claim 1*
  - 2.1 Claim 1 specifies a method which divides a first data stream into  $n$  blocks and calculates their fingerprints, the  $n$  blocks having contiguous logical addresses (features (a) to (c)). According to claim 1, the method then stores data blocks in a first storage area "in a sequence of the logical addresses of the  $n$  data blocks when the fingerprints of the  $n$  data blocks are not



found in fingerprints in the storage device" (feature (d)) and contiguously stores metadata of each of the  $n$  data blocks in a second storage area. The metadata of a block consists of the fingerprint and physical address in the first storage area of the data block corresponding to the fingerprint (features (e) and (e1)).

- 2.2 According to feature (g2), the  $n$  data blocks are included in an "aggregation data block".
- 2.3 Mappings are established between different addresses and metadata (features (f) to (g4)).
- 2.4 Taking the description into account, the board understands that, in view of the purpose of these features, i.e. the claimed "deduplication" purpose, the method is to store the data of a block each time a block of the stream is received which is not a duplicate of a previously received block and to store the metadata of each block. In this manner, the data of repeating blocks is stored only once (deduplication). Similarly, according to the description, the purpose of the aggregation data blocks is to avoid storing duplicate sequences of data chunks (see e.g. page 10, line 27, to page 11, line 15; page 14, line 20 to 34).
- 2.5 In the board's understanding of the invention, in order to ensure that a data block is not stored twice, for each incoming block it should be tested whether its fingerprint corresponds to the fingerprint of any of the blocks already stored in the first storage area (see also the description, page 10, line 31, to page 11 line 5).

Similarly, in order to deduplicate sequences of data blocks, it would have to be tested whether an incoming sequence of blocks of the data stream corresponds to a stored sequence of blocks.

- 2.6 However, these two tests are performed neither by feature (d), which specifies the only test mentioned in the claim, nor by any of the other features of the claim.

According to feature (d), the  $n$  data blocks are stored in the first storage area "when the fingerprints of the  $n$  data blocks are not found in fingerprints in the storage device". It is unclear from this definition whether it is tested that none of the  $n$  fingerprints of the data blocks of the stream is in the storage device or whether at least one of the fingerprints is missing. Furthermore, the test does not take into account the sequence in which data blocks occur in the stream and the sequence in which corresponding data blocks are stored in the first storage area.

The mappings defined in features (f) to (g4) connect the corresponding data in the two storage areas together and establish the relationship between logical and physical addresses of the data blocks. Features (g) to (g4) further describe an "aggregation data block" and its associated metadata. However, these features do not refer to the detection of duplicate data blocks or duplicate sequences of data blocks.

- 2.7 The board concludes from the above that claim 1 does not clearly specify what happens when individual data blocks or sequences of data blocks are received in the data stream which have already been stored (Article 84 EPC).

It is thus not clear from the claim either how the claimed "data de-duplication" is achieved and why the two levels of data blocks and aggregation of data blocks are needed (Article 84 EPC).

- 2.8 In its reply to the board's preliminary opinion, the appellant submitted amended claims but did not present counter-arguments with regard to these objections against the main request.
- 2.9 In view of the above, claim 1 of the main request is unclear and does not fulfil the requirements of Article 84 EPC.

*First and second auxiliary requests*

3. Claim 1 of the first auxiliary request further specifies that "a mapping between the logical address of the first data block and an aggregation address" comprises establishing "a mapping between the logical address of the first data block and both the physical address of the aggregation data block and the address identifier of the metadata of the aggregation fingerprint".
4. The features added in claim 1 of the second auxiliary request concern establishing "an index of a first fingerprint" comprising "a mapping between the first fingerprint and an address identifier of metadata of the first fingerprint".
5. However, none of these features clarifies how the method uses the aggregate data blocks to avoid storing duplicate sequences of data blocks or overcomes the deficiencies mentioned above with regard to claim 1 of the main request.

5.1 In its reply to the board's preliminary opinion, the appellant did not deal with these objections to the first and second auxiliary requests.

5.2 Therefore, the first and second auxiliary requests do not fulfil the requirements of Article 84 EPC, either.

*Third auxiliary request*

6. The appellant submitted that claim 1 of the third auxiliary request was based on originally filed claims 1 and 5 and overcame the grounds for refusal. The appellant has, however, not provided, as required by Article 12(4) RPBA 2020, reasons for submitting the third auxiliary request in the appeal proceedings, i.e. at this late stage of the proceedings and not before.

6.1 The features added in claim 1 of the third auxiliary request specify a "mapping between the logical address of the first data block and an aggregation address" and the use of a compression window. In its preliminary opinion, the board stated that these features did not seem to render the subject-matter of claim 1 inventive, since the use of compression windows was well known. Besides, the features did not solve the clarity objections raised in the board's communication with regard to claim 1 of the main request. Claim 1 of the third auxiliary request was thus *prima facie* not allowable. For these reasons, the board was inclined not to admit the third auxiliary request into the proceedings.

6.2 In reply to the board's communication, the appellant did not contest the board's reasoning on the admissibility of the third auxiliary request.

6.3 The board further notes that the appellant could have filed the third auxiliary request together with the main request and first and second auxiliary requests filed in reply to the examining division's novelty objection raised in the annex to the summons to oral proceedings. The novelty objection was based on document D1, i.e. the same starting point as used in the inventive step reasoning of the decision under appeal.

6.4 In view of this, using its discretion under Article 12(4) RPBA 2020, the board does not admit the third auxiliary request into the proceedings.

*Fourth auxiliary request*

7. *Admissibility*

7.1 The amendments introduced by the fourth auxiliary request (see section XIII. above) attempt to overcome the clarity objections raised by the board for the first time in its communication without significantly changing the subject-matter claimed. In view of that, the board recognises that exceptional circumstances under Article 13(2) RPBA 2020 are present which justify taking the request into account. The fourth auxiliary request is thus admitted into the proceedings.

8. *Clarity - claim 1*

8.1 The appellant argued that the amendment of the designation of the invention to "A method for storing non-duplicate data" clarified that the method of claim 1 was for storing non-duplicate data and thus overcame the objection raised in point 7.6 of the board's preliminary opinion.

From the text of feature (d') it was clear that only non-duplicate data blocks were contiguously stored in a first storage area in a sequence of the logical addresses of the  $n$  data blocks.

The appellant further argued that feature (g3') made it clear that "the physical address of an aggregation data block" was presented using "the first physical address" (i.e. the physical address of the first data block of the  $n$  data blocks stored in the first storage area) and the number of physical blocks that the  $n$  data blocks occupied when the  $n$  data blocks were stored.

8.2 The board notes that the test in feature (d) of the main request was amended to the test in feature (d') of whether "no identical fingerprint is found in fingerprints in the storage device after querying whether a fingerprint of the fingerprints of the  $n$  data blocks is the same as any fingerprint that is stored in the storage device" (see point XIII. above).

8.3 In the board's opinion, feature (d') clarifies that only non-duplicate data blocks are stored in the first storage area and thus overcomes part of the objections raised in point 7.6 of the board's communication and maintained for the main request in this decision (see point 2.7 above). However, for the reasons given in the following, neither the amendment to the designation of the invention nor any of the other amendments introduced with claim 1 of the fourth auxiliary request overcome the objections raised in point 7.6 of the board's communication (and maintained in point 2.7 above) with regard to the occurrence of duplicate sequences of data blocks and the unclear role of the aggregation data blocks.

8.4 Claim 1 specifies, in features (d') and (d1), that each non-duplicate data block is stored in the first storage area in sequence and, in features (e) and (e1), that the metadata of each data block is contiguously stored in a second storage area. However, in these features the method does not check the occurrence of duplicate sequences of data blocks, nor does it store metadata for sequences of data blocks.

The mappings defined in features (f) to (g4) connect the corresponding data in the two storage areas together and establish the relationship between logical and physical addresses of the data blocks. Features (g) to (g4), including amended feature (g3'), further describe an "aggregation data block" and its associated metadata.

However, these features do not refer to the detection of duplicate sequences of data blocks and do not clearly specify how aggregation data blocks are used to avoid the duplicate storage of sequences of data blocks. Feature (g) specifies that a mapping is established between the logical address of the first data block (of a sequence of data blocks) and an aggregation address. But none of the claim features specifies that this mapping, the aggregation address or other metadata of the aggregation data block are used when a duplicate sequence is detected. Since in features (e), (e1) and (f) metadata of each of the  $n$  data blocks is stored, it is not clear why the metadata for an aggregation data block is needed.

8.5 Therefore, claim 1 of the fourth auxiliary request does not satisfy the requirements of Article 84 EPC.

*Concluding remark*

9. Since none of the admitted requests is allowable, the appeal is to be dismissed.

**Order**

**For these reasons it is decided that:**

The appeal is dismissed.

The Registrar:

The Chair:



S. Lichtenvort

J. Geschwind

Decision electronically authenticated