Internal distribution code:

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

Datasheet for the decision of 22 April 2025

Case Number: T 0181/24 - 3.5.07

Application Number: 18892475.7

Publication Number: 3717997

IPC: G06F16/2453

Language of the proceedings: EN

Title of invention:
CARDINALITY ESTIMATION IN DATABASES

Applicant:
HUAWEI TECHNOLOGIES CO., LTD.

Headword:
Cardinality estimation in databases/HUAWEI

Relevant legal provisions:
RPBA 2020 Art. 13(1), 13(2)

Keyword:
Amendment after summons - exceptional circumstances (no) - sole request

Europäisches Patentamt
European Patent Office
Office européen des brevets

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

Case Number: T 0181/24 - 3.5.07

D E C I S I O N
of Technical Board of Appeal 3.5.07
of 22 April 2025

Appellant: (Applicant)	HUAWEI TECHNOLOGIES CO., LTD. Huawei Administration Building, Bantian, Longgang District, Shenzhen, Guangdong 518129 (CN)
Representative:	Gill Jennings & Every LLP The Broadgate Tower 20 Primrose Street London EC2A 2ES (GB)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 9 October 2023 refusing European patent application No. 18892475.7 pursuant to Article 97(2) EPC**

Composition of the Board:

Chair	M. Jaedicke
Members:	C. Barel-Faucheux
	E. Mille

Summary of Facts and Submissions

I. The appellant (applicant) appealed against the examining division's decision refusing European patent application No. 18 892 475.7, published as international application number WO 2019/120093 A1.

II. The examining division decided that the subject-matter of claim 1 of the sole request was unclear (Article 84 EPC) and the subject-matter of the independent claims, as far as they could be understood, lacked an inventive step in view of the disclosure of the following document:

D1: Hien To et al.: "Entropy-based Histograms for Selectivity Estimation", Proceedings of the 22nd ACM international conference on Information & Knowledge Management, CIKM '13, San Francisco, CA, USA, 27 October 2013, pages 1939 to 1948

and the common general knowledge of the skilled person illustrated by the following document:

D4: Paul White: "Cardinality Estimation: Combining Density Statistics", SQLPerformance.com, 19 September 2017, https://web.archive.org/web/20170919134545/https://sqlperformance.com/2017/08/sql-optimizer/combining-density

III. In its statement of grounds of appeal, the appellant requested that the decision under appeal be set aside and that a patent be granted on the basis of a new set of claims filed with the statement of grounds of appeal.

IV. In a communication pursuant to Article 15(1) RPBA accompanying the summons to oral proceedings, the board cited the following document:

D5: Josien P. W. Pluim: "Mutual-Information-Based Registration of Medical Images: A Survey", IEEE Transactions on Medical Imaging, Vol. 22, No. 8, August 2003, pages 986 to 1004

The board stated that it was inclined not to admit the sole request into the proceedings (Article 12(4) and (6) RPBA), that the subject-matter of claim 1 of the sole request appeared to lack clarity (Article 84 EPC) and that it was not apparent to the board what technical problem was solved by the method of claim 1 of the sole request. It appeared, in the light of documents D1, D4 and D5, that the method of claim 1 of the sole request resulted from known mathematical expressions concerning entropy and therefore did not appear to be inventive (Article 56 EPC).

V. By letter dated 27 February 2025, the appellant filed a new set of claims "to replace the claims currently on file" together with arguments with regard to admissibility, clarity and inventive step.

VI. By a further letter dated 24 March 2025, the representative informed the board that the appellant would not be represented at the oral proceedings. The oral proceedings were then cancelled.

VII. Claim 1 of the sole request reads as follows:
(itemisation of the features added by the board, the underlined part was added with respect to claim 1 of the sole request to which the appealed decision

relates, the part in italics was added in claim 1 of the previous sole request filed with the statement of grounds of appeal on which the current sole request is based):

"A method applied to a database management system, comprising:

A calculating, by a server in the database management system (102), a data skew for a first column based on an entropy for the first column in a database (101),

B wherein the first column includes a plurality of distinct values wherein the skew k is given by $k=2^{\log n-H}$, where H is entropy for the first column and n is the number of total distinct values in the first column;

C storing, by the server, the data skew to a catalogue in the database;

D estimating, by the server, a selectivity for the first column by multiplying a first selectivity with the data skew where the selectivity is given by $s = 2^{-H} = (1/n)*k$, where k is the data skew and (1/n) is the first selectivity; and

wherein the method further comprises:

E calculating, by the server, a joint entropy for the first column and a second column in the database, *wherein the first and second columns are columns from the same database table,*

F wherein joint entropy is defined as:

$$H(x,y) = -\sum_{x,y} p_{(x,y)} \times logp_{(x,y)},$$

where p(x,y) is the joint selectivity function of X and Y, wherein X and Y are first and second columns;

G calculating, by the server, mutual information I defined as:
I(x,y) = H(x) + H(y) - H(x,y), wherein H(x) is the entropy for the first column, H(y) is an entropy for the second column, and H(x,y) is the joint entropy;

H calculating, by the server, a degree of correlation c between the first and second columns based on the mutual information, wherein $c = 2^I$; and

I storing, by the server, the degree of correlation in the catalogue; and further

wherein [*sic*] the method further comprises:

J estimating, by the server, a joint selectivity between the first and second columns by multiplying the selectivity for the first column, a selectivity for the second column given by

J1 the selectivity is [*sic*] given by $s = 2^{-H(y)}$,

J2 and the degree of correlation; and

L generating, by the server, a query plan based on the estimated joint selectivity

M calculating, by the server, a cost of the query plan based on the degree of correlation and the data skew stored in the catalogue;

N wherein generating the query plan comprises minimizing the cost of the query plan using a cost model."

Reasons for the Decision

1. The application relates to a method for improving cardinality estimation (CE) in a database management system (DBMS) (paragraph [0006] of the published application). By calculating the data skew and correlation based on entropy and estimating the selectivity by considering the data skew and correlation, CE in the DBMS is improved. The improved cardinality estimates in turn improve the query plan leading to better query performance (paragraph [0015]).

Sole request

2. Admissibility

2.1 The sole request constitutes an amendment to the appellant's appeal case which may be admitted only at the discretion of the board (Article 12(4) RPBA).

2.2 The appellant argued, in the statement of grounds of appeal, that the amendments introduced by the previous sole request, on which the current sole request is

based, served to overcome the objections raised in point 15.1.1 of the decision under appeal.

2.3 The board notes that the amendments introduced by claim 1 of the sole request are not merely minor clarifications but instead change the claimed subject-matter. Moreover, they do not address the part of the objection raised in point 15.1.1 of the decision under appeal relating to the unclear relationship of the first and second columns X and Y (from the same database table) with the query for which the query plan is generated, since no query is stipulated in claim 1. In addition, they do not address the part of the objection raised in point 15.1.1 of the decision under appeal relating to the unclear relationship between the query plan (and a corresponding "query" which is however not specified in claim 1) referred to in features L to N of claim 1 and the first and second columns X and Y, since this relationship is not specified in claim 1.

2.4 In their letter of reply dated 27 February 2025, the appellant argued that the first and second columns were analysed to understand their individual and combined statistical properties (e.g. entropy, skew, correlation), and this information was then used to refine the query optimiser's understanding of the data distribution and relationships within the table, resulting in an efficient query execution plan. The relationship between the first and second columns and the query revolved around their contribution to query optimisation through the estimation of selectivity and correlation. In particular, the selectivity of the first column was adjusted based on its data skew, and a joint selectivity for the first and second columns was calculated by incorporating the degree of correlation

between the two columns, ensuring that the query optimiser had an accurate estimate of the combined filtering effect of the two columns.

2.4.1 The board first notes that there is neither an explicit "filtering" step or effect (it might be that the estimation of the joint selectivity of feature J is considered by the appellant as corresponding to a "filtering step" but this is not clear from the wording of claim 1 alone) nor an explicit query optimiser (it might be that the calculation and minimisation of the cost of the query plan of features M and N is considered by the appellant to be "a query optimisation" but this is not clear from the wording of the claim alone) present in claim 1 of the sole request and that claim 1 does not in any way specify the query for which a query plan is generated. It remains unclear how an arbitrary query would relate to the first and second columns X and Y of claim 1. For example, claim 1 does not specify any query conditions regarding these columns.

2.5 These clarity objections raised in the decision under appeal were similar to those raised by the examining division in its preliminary opinion of 21 September 2023 (see point 4.1.1) in advance of the date scheduled for oral proceedings and to those raised in the annex to the summons to attend oral proceedings (see point 2.1.1). In reply to the examining division's preliminary opinion, the appellant chose not to file any amendments and announced, for the second time, that it would not attend the oral proceedings scheduled on 5 October 2023. The amendments introduced by the previous sole request and present in the current sole request could thus have been submitted during the examination proceedings.

2.6 Moreover, the appellant did not provide reasons for submitting amendments M and N at this stage of the appeal proceedings. In the board's view, there are no exceptional circumstances which have been justified with cogent reasons by the appellant, as required by Article 13(2) RPBA.

2.6.1 In particular, the objection raised in point 16 of the board's communication stating that it was not clear in feature L how the query plan was generated based on the estimated joint selectivity was already present in point 15.1.3 of the examining division's decision.

2.6.2 In its reply to the board's communication, the appellant maintained that the claim was clear and that the skilled person would know how to implement this feature. For example, the joint selectivity between the first and second columns, calculated in step J, was "used as a key input" for query optimisation. This value reflected how the combined filtering effect of the two columns impacted the query result size. The joint selectivity helped refine the cardinality estimate, and more accurate cardinality estimates led to better cost estimates for different execution strategies. The plans could be evaluated based on their estimated cost (see feature N), which depended on the joint selectivity, and the final query plan could be selected, stored or directly executed by the database engine.

2.6.3 However, since no query is specified in claim 1, the skilled person reading claim 1 cannot understand how an unspecified query would relate to columns X and Y and why the estimated joint selectivity would be useful for generating a plan for this query. In addition, the following aspects are not reflected in the claim: there

are no explicit filtering, no cardinality estimates and no selection, storage or execution of the (final) query plan.

2.7 In the board's communication, point 17 (see also decision, points 15.1.3 and 17.2, last paragraph), the board stated that it was unclear how the query plan could be generated when the relationship between two columns and the query being evaluated was undefined.

In their letter of reply, the appellant argued that, in this case, a query plan could still be selected or generated based on joint selectivity by using joint selectivity "as a statistical tool to improve the optimiser’s estimation of intermediate results". All potential execution plans for the query were evaluated and joint selectivity was incorporated as part of cost estimation. For example, if joint selectivity was low, this implied that combined filtering or joining would produce fewer rows, potentially favouring more sequential filtering approaches. If the joint selectivity was high, this indicated less selective operations, which could lead to favouring broader scans or parallel processing. The final query plan was still generated taking into account both the explicit query conditions and the underlying statistics (including joint selectivity) for all relevant columns. The plan with the lowest estimated cost was still selected, even if joint selectivity was used in a secondary or indirect manner. Thus, the joint selectivity of two columns provided useful information that could be used to generate the query plan, even in the absence of a direct or explicit relationship to the query. It helped ensure a more accurate model of data distribution, which was applied broadly to all potential query paths, improving overall plan selection.

2.7.1 The board, however, is of the opinion that the claim does not explicitly specify that the query is for searching in the database having the first and second columns. In any event, it remains unclear why joint selectivity would be useful for any arbitrary and unspecified query (for example, for a query relating to a single column of the database) and how this selectivity is actually used.

2.8 Feature N now defines "generating the query plan" as comprising "minimizing the cost of the query plan using a cost model". Feature M stipulates a cost of the query plan based on the degree of correlation and the data skew stored in the catalogue, while feature L stipulates "generating, by the server, a query plan based on the estimated joint selectivity".

2.9 Therefore, the clarity objection under Article 84 EPC in point 16 of the board's communication (i.e. how the query plan is generated based on the estimated joint selectivity, see point 2.6.1 above and what are the relationships between the first and second columns and the query for which the query plan is generated, see point 15.1.1 of the decision) still *prima facie* applies to claim 1 of the sole request.

2.10 **In view of the above, the board does not admit the sole request into the proceedings (Article 13(1) and (2) RPBA). Since the appellant's sole request is not admitted, the appeal is to be dismissed.**

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chair:

S. Lichtenvort M. Jaedicke

Decision electronically authenticated