

PATENTS ACT 1977

APPLICANT Hitachi Vantara LLC

ISSUE Whether patent application GB2003326.2 is excluded under section 1(2)(c)

HEARING OFFICER H Jones

DECISION

Background

- 1 Patent application GB2003326.2, now in the name of Hitachi Vantara LLC, was filed on 24 July 2018 via the international route and has a priority date of 10 August 2017. The international application was published as WO 2020/074932 A2 on 16 April 2020 and subsequently as GB 2580559 A.
- 2 The original 'compliance period', that is the period by the end of which the application needs to comply with all the requirements of the Patents Act 1977 ("the Act") and the rules, ended on 10 December 2022. This period was extended by two months to end on 10 February 2023.
- 3 Throughout the examination process, the examiner objected that the application relates to a computer program and is therefore excluded from patentability. Objections were also raised relating to clarity. Following several rounds of correspondence and amendments to the claims, the clarity issues have been resolved, but the applicant's attorney, Dr John Addiss of Mewburn Ellis, has not been able to persuade the examiner that the application is not excluded. The examiner consequently offered a hearing, which took place on 12 January 2023 via video conference. I am grateful to Dr Addiss for his clear and helpful presentation of the arguments.

The issue to be decided

- 4 The only outstanding matter is that of excluded matter. Thus, the issue to be decided is whether the invention consists solely of a program for a computer which the Act excludes from patentability under section 1(2)(c).

The invention

- 5 A well-known concept employed in relational database models is the primary key-foreign key relationship between datasets. When performing data integration, the structure of a database and the relationships between the elements of the database is required, but this data is sometimes lost or corrupted. In such circumstances the

relationships must be established, but this is time consuming and computationally intensive due to the vast amount of data involved.

- 6 The application relates to establishing these primary key-foreign key relationships in a database. Rather than comparing all the data in every column of a database to determine these relationships, which is computationally intensive and therefore impractical, the application uses an inclusion dependency system.
- 7 The invention aims to reduce computations and minimise disk operations when determining inclusion dependency and establishing primary-foreign key relationships. To do this the database is pruned by selecting a portion of the columns of the database (to discard any columns which won't have the primary key-foreign key relationships) and these columns are then sorted/ordered. Candidate primary key and candidate foreign key pairs are determined based on dependency characteristics (e.g. average step size in each column). The inclusion dependency is determined by looking at a possible primary key-foreign key relationship using search techniques that are dynamically chosen depending on the dependency characteristics. The number of fetches of candidate keys are computed and this value is then used to determine whether to retain or relinquish the relevant candidate primary keys or foreign keys in order to minimise the input and output operations that are performed on the file system.
- 8 The application includes two independent claims (1 and 14):
 1. A method of establishing primary key-foreign key relationships in a database stored in a file system or one or more solid state hard drives by determining inclusion dependency between a plurality of columns, said method employing an inclusion dependency determination system comprising at least one processor configured to execute computer program instructions for performing the method, the method comprising:
 - generating a pruned database by retaining only a portion of the plurality of columns having a predetermined data type;
 - sorting data in the pruned database;
 - determining a dependency characteristic that comprises at least a minimum value, a maximum value and an average step size of each column in the sorted pruned database;
 - arranging the columns in the sorted pruned database by applying at least one predefined rule based on a minimum value of the data in each of the columns;
 - identifying at least one candidate primary key and at least one candidate foreign key by determining which pairs of columns in the sorted pruned database demonstrate possible inclusion dependency based on the dependency characteristic; and
 - determining inclusion dependency by comparing data in the candidate primary key with data in the candidate foreign key by a search technique, wherein the search technique is dynamically selected, from multiple search techniques, based on the average step size;

wherein the step of arranging the columns comprises the step of creating a relationship mapping comprising a unidirectional chain, wherein at least one column occurring earlier in the chain fully contains at least one column occurring later in the chain;

wherein the method further comprises:

computing a number of fetches required to retrieve the candidate primary key and the candidate foreign key from the file system or one or more solid state hard drives based on the relationship mapping of the columns; and based on the computed number of fetches, performing one of retention and relinquishment of one or more of the candidate primary key and the candidate foreign key in a non-transitory computer readable storage medium, thereby minimizing input and output operations of said file system or one or more solid state hard drives.

14. A computer implemented method for determining inclusion dependency between a plurality of columns of a plurality of tables in a target database stored in a file system or one or more solid state hard drives, said method comprising the steps of:

obtaining sorted columns of at least a portion of said plurality of columns;

determining dependency characteristic data for each of said columns, the dependency characteristic data comprising at least an average step size, a minimum value and a maximum value of the data in each of said sorted columns;

arranging the sorted columns by applying at least one predefined rule based on a minimum value of the data in each of the sorted columns, wherein the step of arranging the sorted columns comprises the step of creating a relationship mapping comprising a unidirectional chain,

wherein at least one column occurring earlier in the chain fully contains at least one column occurring later in the chain;

determining possible inclusion dependency pairs in the sorted columns, based at least in part on said determined dependency characteristic data, and wherein a first column of said pair of columns is identified as a candidate primary key and a second column is identified as a candidate foreign key;

determining the presence of each of the candidate foreign key values in the candidate primary key values by a search technique, the search technique selected from one of multiple possible search techniques, wherein the selection is based at least in part on said dependency characteristic data;

and determining an inclusion dependency based on the presence of the candidate foreign key values in the candidate primary key values,

wherein the method further comprises the steps of:

computing a number of future fetches required of each candidate foreign key and each candidate primary key of each of the pairs of candidate foreign key

and candidate primary key from the file system or one or more solid state hard drives, based on the relationship mapping of the columns;

and based on the computed number of fetches, performing one of retention and relinquishment of one or more of the candidate foreign keys and candidate primary keys; wherein candidate foreign keys and candidate primary keys are retained or relinquished to minimize input and output operations of said file system or one or more solid state hard drives.

The Law

- 9 The examiner has raised an objection that the invention is not patentable because it relates to one or more of the categories of subject-matter which are not considered to be inventions under the Act. This 'excluded matter' is set out in section 1(2) of the Act:

1(2). It is hereby declared that the following (among other things) are not inventions for the purposes of this Act, that is to say, anything which consists of –

- (a) ...*
- (b) ...*
- (c) ... a program for a computer;*
- (d) ...*

but the foregoing provision shall prevent anything from being treated as an invention for the purposes of this Act only to the extent that a patent or application for a patent relates to that thing as such.

- 10 The Court of Appeal's judgment in *Symbian*¹ tells us that in order to determine whether an invention falls solely within the any of the exclusions listed in section 1(2), the four-step test set out in its earlier judgment in *Aerotel*² must be used. The four steps are:

- (i) properly construe the claim;
- (ii) identify the actual contribution;
- (iii) ask whether it falls solely within the excluded subject-matter;
- (iv) check whether the actual or alleged contribution is actually technical in nature.

- 11 To assist in identifying whether there is a technical contribution in computer related inventions, the signposts set out in *AT&T/CVON*³ and by the Court of Appeal in *HTC/Apple*⁴ act as guidelines. They provide a list of some of the factors that can indicate whether a contribution may be technical. They are:

- i) whether the claimed technical effect has a technical effect on a process which is carried on outside the computer.
- ii) whether the claimed technical effect operates at the level of the architecture of the computer; that is to say whether the effect is produced irrespective of the data being processed or the applications being run.

¹ *Symbian Ltd v Comptroller-General of Patents* [2008] EWCA Civ 1066

² *Aerotel Ltd v Telco Holdings Ltd and Macrossan's Application* [2006] EWCA Civ 1371

³ *AT&T Knowledge Venture/CVON Innovations v Comptroller General of Patents* [2009] EWHC 343 (Pat)

⁴ *HTC Europe Co Ltd v Apple Inc* [2013] EWCA Civ 451

iii) whether the claimed technical effect results in the computer being made to operate in a new way.

iv) whether the program makes the computer a better computer in the sense of running more efficiently and effectively as a computer.

v) whether the perceived problem is overcome by the claimed invention as opposed to merely being circumvented.

Argument and analysis

Step 1 - Properly construe the claim

12 Dr Addiss accepted the examiner's construction of the claims. I see no reason to consider this further, noting only that while there are some minor differences between claims 1 and 14, claim 14 has been construed to be essentially performing the same steps as those outlined in claim 1.

13 I therefore construe claim 1 as a method of establishing primary key-foreign key relationships in a database by determining inclusion dependency between a plurality of columns of the database, comprising:

generating a pruned database from a portion of a plurality of columns;

sorting the data in the pruned database;

determining a dependency characteristic that comprises at least a minimum value, a maximum value and an average step size of each column in the sorted pruned database;

arranging the columns in the sorted pruned database based on applying at least one predefined rule based on a minimum value of the data in each columns, the arranging further comprising creating a relationship mapping between the columns wherein the creating of the relationship mapping comprises a unidirectional chain, wherein at least one column occurring earlier in the chain fully contains at least one column occurring later in the chain;

identifying at least one candidate primary key and at least one candidate foreign key by determining which pairs of columns in the sorted pruned database demonstrate possible inclusion dependency based on the dependency characteristic;

determining inclusion dependency by comparing data in the candidate primary key with data in the candidate foreign key using a search technique which is dynamically selected, from multiple possible search techniques, based at least in part on said dependency characteristic;

computing a number of fetches required to retrieve the candidate primary key column(s) and the candidate foreign key column(s) from the file system or one or more solid state drives based on the relationship mapping of the columns;
and

based on the computed number of fetches, performing one of retention and relinquishment of one or more of the candidate primary key column(s) and candidate foreign key column(s).

Step 2 – Identify the actual (or alleged) contribution

- 14 Paragraph 43 of Aerotel suggests that the contribution is, in essence, that which has been added to the stock of human knowledge. Determining the contribution involves taking into account the problem to be solved, how the invention works and what the advantages are, and also looking at the substance and not the form of the claims.
- 15 Although Dr Addiss essentially accepted the examiner's detailed analysis of the contribution, he wished to emphasise the advantages of the invention (as set out his assessment of the contribution in his letter of 4 Jan 2023 and reiterated at the hearing). These advantages being the establishment of primary key-foreign key relationships in a fast and efficient manner while minimising disk input and output operations, for example leading to reduced wear of the memory unit and therefore longer time before servicing of the memory unit is required.
- 16 Fundamentally the invention attempts to reduce the amount of columns/data that are being compared in order to improve the efficiency of determining primary key-foreign key pairs. Reducing the amount of columns/data to compare is achieved by initially selecting only particular columns and sorting the data in these columns. The columns themselves are also arranged using the minimum values of the data in the columns and further arranged so that a relationship mapping between the columns can be determined using chaining. Arranging the columns and determining their relationship mapping should ensure that possible primary key-foreign key pair columns are determined efficiently. An appropriate search technique is also used based on a particular attribute of the data, e.g. average step size. A calculated number of fetches is based on the determined relationship mapping in order to determine whether to retain or relinquish one or more of the candidate primary and foreign keys. The hardware utilised is considered to be purely conventional.
- 17 I therefore assess the contribution to be as follows:
- 18 The determination of primary key-foreign key relationships in a database in a fast and efficient manner while minimising disk input and output operations by determining an inclusion dependency between a plurality of columns found in the database. This is done by determining inclusion dependency between a portion of columns selected from a plurality of columns of the database; using dependency characteristics to determine candidate primary key and foreign key pair columns from the portion of columns; arranging the columns by determining relationship mappings between the columns, the relationship mapping being determined using a unidirectional chain, wherein at least one column occurring earlier in the chain fully contains at least one column occurring later in the chain; using dynamically selected search techniques to compare data between the candidate primary and foreign key columns in order to determine the inclusion dependency, the dynamically selected search technique being selected based at least in part on the dependency characteristics; wherein the number of fetches of the candidate primary and foreign key columns from the file system or solid state drive(s) is calculated based on the relationship mapping between the candidate primary and foreign key columns, this calculated number of fetches being used to determine whether to retain or relinquish

the candidate primary and foreign key columns when determining the inclusion dependency of the candidate columns.

Steps 3 & 4 - Whether the actual or alleged contribution falls solely within the excluded matter and check whether it is actually technical

19 In considering whether the above contribution is technical, I will use the above-mentioned signposts which provide a list of some of the factors that can indicate whether a contribution may be technical in computer related inventions.

20 At the hearing, Dr Addiss asserted that the first, second and fifth signposts are relevant here. I agree, so will restrict my analysis to these.

Signpost i) *whether the claimed technical effect has a technical effect on a process which is carried on outside the computer;*

21 To meet the first signpost, the process carried out by the program must be, or must operate on, something external to the computer on which the program is being run.

22 The examiner was unable to identify a technical effect on a process which is carried on outside the computer. He noted that the invention relates to determining the primary key-foreign key relationships within a database of a computer, and also that any effect on the storage system is within the computer itself and not outside it.

23 Dr Addiss argued that the invention does provides a technical effect outside the computer because the reduced wear on the memory unit results in the need for less maintenance of the memory, maintenance being something which occurs outside the computer.

24 I am not persuaded by this argument. The process carried out by the program relates to determining the primary key-foreign key relationships within a database of a computer. The process does not itself operate on something external to the computer. Thus signpost 1) is not met.

Signpost ii) *whether the claimed technical effect operates at the level of the architecture of the computer;*

25 The examiner could not identify anything which would indicate that the contribution operates at an architectural level of the computer. He contended that because the contribution relates to the use of a database application and determining primary key-foreign key relationships in that application, the effect is produced only for the application in question and would not provide a benefit to any software program running on the system. He also argued that the calculation of fetches and the effect this has on read/write operations of the underlying storage system is not at the architectural level of the computer as it only has an effect for this particular database computer program.

26 Dr Addiss argued that a data repository is considered to be an infrastructure component and this works irrespective of the data being processed and the application being run. In his letter of 4 January 2023 he sets out that "...the data repository and the process it implements does not operate at the application level and are designed to operate independently of both the nature of the processed data and of the particulars of the applications accessing the data – instead they operate at

a lower level, providing an infrastructure component for efficient data retrieval upon which various software applications and services can be built”.

- 27 The contribution relates to the use of a database application and determining primary key-foreign key relationships in that application, and doing so in a more efficient/less computationally intensive way. It is less computationally intensive because of the way the memory is accessed (specifically fewer fetches), not because the memory itself works differently. I am therefore of the view that how the memory is used is completely tied to the application being run. Signpost ii) is not met.

Signpost v) whether the perceived problem is overcome by the claimed invention as opposed to merely being circumvented.

- 28 When looking to the fifth signpost, the examiner identified the existing problems with determining primary key-foreign key relationships in a database as computational overhead and the number of disk input and output operations required.
- 29 The invention attempts to overcome these problems by reducing the number of columns to be considered (pruning), determining the column's relationships based on their relation to each other and the attributes of the values they hold, and using different search techniques. The examiner asserted that none of these features overcome the problem that a computer system may not have sufficient processing or appropriate disk performance to determine the primary key-foreign key relationships. Instead, the contribution of the invention circumvents the issue of inadequate or inappropriately configured computer systems. He points out in his letter of 21 November 2022 that “...a situation could arise in that if a database is of a particularly large size with regards to columns and data in these columns and/or if the data types or values of the columns are of a particular type or structure then the invention may well not be able to prune sufficient columns, arrange the columns to provide a more efficient selection of columns or may not be able to utilise a more efficient search technique leading to no gain in the efficiency...”.
- 30 The invention also includes the calculation of the number of fetches based on relationship mapping between the candidate primary and foreign columns, and using this to determine whether to retain or relinquish columns. The examiner did not consider this to overcome the technical problems of storage system performance, but merely to improve how this particular program utilises the underlying storage system. He considered this to circumvent rather than solve the existing issues in storage system performance such as the number disk input and output operations it can support, or lifespan due to wear and tear from these operations.
- 31 Dr Addiss, takes a similar view to the examiner regarding the problems overcome by the invention. In his discussion of the invention at the hearing and in his letters of 7 February 2022, 7 July 2022 and 4 January 2023, he sets out how the invention reduces computations and minimises disk operations when determining inclusion dependency and establishing primary-foreign key relationships. This minimises unnecessary wear of the memory unit and reduces the servicing requirement of the memory unit. At the hearing he asserted that these problems are overcome and not circumvented because all the primary key-foreign key relationships are still identified, but in a more efficient way.
- 32 Dr Addiss sought support from the EPO Board of Appeal's decision [T1351/04](#) relating to the creation of an index file for searching a file which controls the retrieval

of data by directing the computer to a certain memory location. Dr Addiss pointed to paragraphs 7.2 and 9 where, and I summarise, the Board held that the index file was technical because it determined the way the computer searches (searching being controlling the computer along a path to retrieve to the desired data). Dr Addiss suggested that the primary key-foreign key relationships of the invention equate to indexing data and are therefore technical. In particular, he suggested in his letter on 4 January 2023 that the invention “provides an improved index method for data handling and specifically for searching a record in a database” and that “the primary key-foreign key relationship established...allows all information related to a searched entity to be simultaneously obtained”. He considered this to be technical because “...by using an index method in order to generate an improved search procedure, one is commonly able to control the way a...computer performs a search operation”.

- 33 The EPO has a different approach to patentability criteria, and while decisions on patentability given by the EPO Boards of Appeal can be of persuasive value, they are not binding. Furthermore, the details of the Board’s case can be distinguished from the application in suit, where the contribution isn’t directly concerned with how the computer searches the database but, instead, how it goes about finding relationships between data in a database. Thus, I do not consider the primary key-foreign key relationships of the invention to be equivalent to indexing, so cannot be considered to be technical based on this reasoning.
- 34 While the contribution does indeed provide all primary key-foreign key relationships and does this more efficiently as suggested by Dr Addiss, it does this by considering a subset of the overall data. It does not process all the data more efficiently, but avoids having to do this by selecting to analyse less data (by discarding irrelevant data). In effect the problem is avoided by looking at fewer columns, not by looking at all columns faster.
- 35 Furthermore, I do not consider the details regarding computation fetch numbers and retaining/relinquishing data within the storage system to overcome storage system performance issues. It might well improve how the program utilises the underlying storage system, but the capability of the memory in terms of read write operations and wear and tear is not changed. I cannot see how this is anything other than a circumvention of the problem. Signpost v) is not met.
- 36 I therefore conclude that the claimed invention is excluded as a program for a computer as such under section 1(2)(c).

Conclusion

- 37 Having carefully considered the arguments, I am of the view that the claimed invention is excluded by section 1(2)(c) as a program for a computer as such. I therefore refuse the application under section 18(3).

Appeal

38 Any appeal must be lodged within 28 days after the date of this decision.

Huw Jones

Deputy Director, acting for the Comptroller