PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 11/14 | A3

(11) International Publication Number: **WO 96/37837**

(43) International Publication Date: 28 November 1996 (28.11.96)

(21) International Application Number: PCT/NO96/00122

(22) International Filing Date: 21 May 1996 (21.05.96)

(30) Priority Data:
08/451,855 26 May 1995 (26.05.95) US

(71) Applicant: TELENOR AS [NO/NO]; N-7005 Trondheim (NO).

(72) Inventors: TORBJØRNSEN, Øystein; Gyldenløves gate 4, N-7014 Trondheim (NO). HVASSHOVD, Svein-Olaf; Klaebuveien 40B, N-7030 Trondheim (NO).

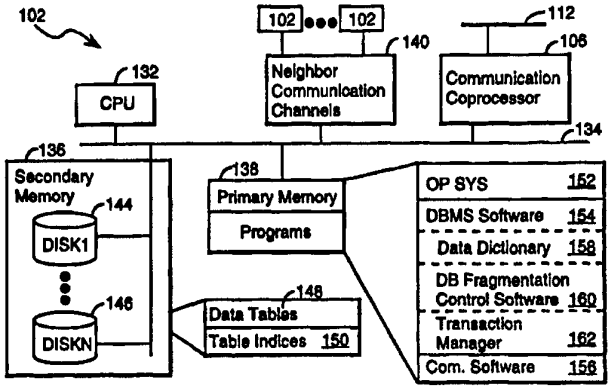
(74) Agent: OSLO PATENTKONTOR A/S; P.O. Box 7007 M, N-0306 Oslo (NO).

(81) Designated States: JP, NO, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
16 January 1997 (16.01.97)

(54) Title: CONTINUOUSLY AVAILABLE DATABASE SERVER HAVING MULTIPLE GROUPS OF NODES WITH MINIMUM INTERSECTING SETS OF DATABASE FRAGMENT REPLICAS



(57) Abstract

A database server with a "shared nothing" system architecture has multiple nodes, each having its own central processing unit, primary and secondary memory for storing database tables and other data structures, and communication channels for communication with other ones of the nodes. The nodes are divided into at least two groups that share no resources, including power supply and cooling system. Each database table in the system is divided into fragments distributed for storage purposes over all the nodes in the system. To ensure continued data availability after a node failure, a "primary replica" and a "standby replica" of each fragment are each stored on nodes in different ones of the groups. Database transactions are performed using the primary fragment replicas, and the standby replicas are updated using transaction log records. Every node of the system includes a data dictionary that stores information indicating where each primary and standby fragment replica is stored among the system's nodes. The records of each dtabase table are allocated as evenly as possible among the table fragments, for example, by hashing a primary key value for each record with a predefined hash function and using the resulting value to select one of the database table fragments. A transaction manager on each node responds to database queries by determining which fragment of a database is being accessed by the query and then forwarding the database query to the node processor on which the primary replica of that fragment is stored. Upon failure of any one of the data processors in the system, each node updates the information in its data dictionary accordingly. In addition, the fragment replicas made unavailable by the node failure are regenerated and stored on the remaining available nodes in the same node group as the failed node.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

INTERNATIONAL SEARCH REPORT

Inter..onal Application No
PCT/NO 96/00122

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F11/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO,A,94 14125 (TELEFONAKTIEBOLAGET L M ERICSSON) 23 June 1994 see claim 1	1-7
A	US,A,5 307 481 (SHIMAZAKI ET AL.) 26 April 1994 see abstract	1-7
A	US,A,5 379 418 (SHIMAZAKI ET AL .) 3 January 1995 see column 2, line 9 - line 23; figure 10 see column 3, line 14 - line 26	1-7

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
13 November 1996	26. 11. 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Corremans, G

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter. ...nal Application No
PCT/NO 96/00122

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9414125	23-06-94	SE-C-	500656	01-08-94
		AU-B-	670852	01-08-96
		AU-A-	5663294	04-07-94
		CA-A-	2151254	23-06-94
		CN-A-	1092886	28-09-94
		EP-A-	0673528	27-09-95
		FI-A-	952793	07-06-95
		JP-T-	8504529	14-05-96
		NO-A-	952248	02-08-95
		SE-A-	9203691	09-06-94
		US-A-	5548750	20-08-96
US-A-5307481	26-04-94	JP-A-	3256146	14-11-91
		JP-A-	3256143	14-11-91
		JP-A-	3256144	14-11-91
		US-A-	5379418	03-01-95
US-A-5379418	03-01-95	JP-A-	3256146	14-11-91
		JP-A-	3256143	14-11-91
		JP-A-	3256144	14-11-91
		US-A-	5307481	26-04-94