

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 February 2010 (04.02.2010)

PCT

(10) International Publication Number
WO 2010/014955 A3

(51) International Patent Classification:
G06F 17/30 (2006.01) G06F 17/00 (2006.01)

NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(21) International Application Number:
PCT/US2009/052490

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(22) International Filing Date:
31 July 2009 (31.07.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/085,022 31 July 2008 (31.07.2008) US
12/270,872 14 November 2008 (14.11.2008) US

Declarations under Rule 4.17:

(71) Applicant (for all designated States except US): MICROSOFT CORPORATION [US/US]; One Microsoft Way, Redmond, WA 98052-6399 (US).

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

(72) Inventors: NETZ, Amir; One Microsoft Way, Redmond, WA 98052-6399 (US). PETCULESCU, Cristian; One Microsoft Way, Redmond, WA 98052-6399 (US).

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(88) Date of publication of the international search report:
22 April 2010

(54) Title: EFFICIENT LARGE-SCALE PROCESSING OF COLUMN BASED DATA ENCODED STRUCTURES

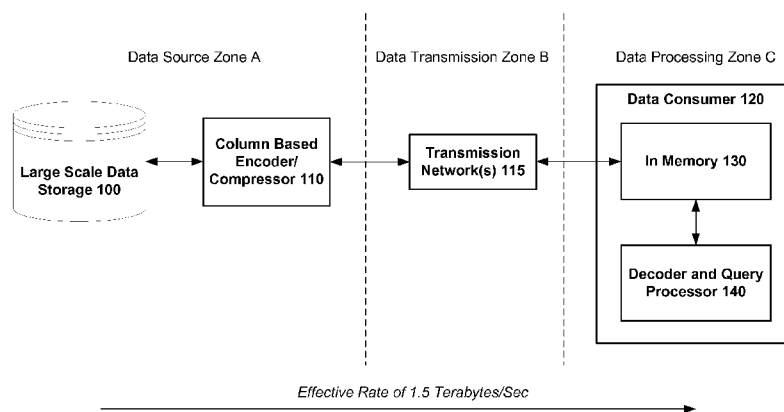


FIG. 1

(57) Abstract: The subject disclosure relates to efficient query processing over large scale data storage. An exemplary process includes retrieving a subset of columns implicated by a query as integer encoded and compressed sequences of values corresponding to different columns of data, defining query processing buckets that span over the subset of columns based on changes of compression type occurring in the integer encoded and compressed sequences of values of the subset of data and processing the query in memory on a bucket by bucket basis and processing the query based on type of current bucket when processing the integer encoded and compressed sequences of values. The column based organization of the data, and the application of a hybrid run length encoding and bit packing technique, enable a highly efficient and speedy query response in real-time.



WO 2010/014955 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2009/052490**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/30(2006.01)i, G06F 17/00(2006.01)i**

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)

G06F 17/30; G06F 7/00

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

Korean utility models and applications for utility models
Japanese utility models and applications for utility models
(Chinese Patents and application for patent)

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

eKOMPASS(KIPO internal) & Keywords: query, compress, large-scale

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2008-0071724 A1 (STEVEN W. PARKINSON et al.) 20 March 2008 See abstract; para.30-31, 34, 47-51; claims 1-3, 5-6 and figures 3-4.	1-21
A	US 2008-0133562 A1 (CHEONG CHI PING BESS et al.) 05 June 2008 See abstract; para.3-7, 20-25, 31-32, 72; claims 1-5 and figures 2-3, 8-9.	1-21
A	US 7024414 B2 (ADAM SAH et al.) 04 April 2006 See abstract; col.1 ln.49-col.2 ln.10, col.3 ln.22-col.4 ln.16, col.9 ln.31-67, col.14 ln.61-col.15 ln.41; claims 1-7 and figures 5-6.	1-21
A	US 2007-0143248 A1 (RADHA KRISHNA UPPALA) 21 June 2007 See abstract; para.19-22, 42-44, 47, 55-57, 75-78; claims 1-7 and figures 3-4, 6-8.	1-21

 Further documents are listed in the continuation of Box C. See patent family 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 application or patent 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 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

Date of the actual completion of the international search

25 FEBRUARY 2010 (25.02.2010)

Date of mailing of the international search report

26 FEBRUARY 2010 (26.02.2010)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Kim, Nam In

Telephone No. 82-42-481-8583



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2009/052490

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008-0071724 A1	20.03.2008	None	
US 2008-0133562 A1	05.06.2008	None	
US 7024414 B2	04.04.2006	EP 1283486 A2 EP 1283486 A3 US 2003-0028509 A1	12.02.2003 08.12.2004 06.02.2003
US 2007-0143248 A1	21.06.2007	None	