



(51) International Patent Classification:

G06F 12/00 (2006.01) G06F 11/14 (2006.01)
G06F 12/16 (2006.01) G06F 9/44 (2006.01)

(21) International Application Number:

PCT/US2011/065662

(22) International Filing Date:

16 December 2011 (16.12.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

12/971,694 17 December 2010 (17.12.2010) US

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

(72) Inventors: CHEUNG, Chun, Ho (Ian); c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). OLTEAN, Paul, Adrian; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). BENTON, James, Robert; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

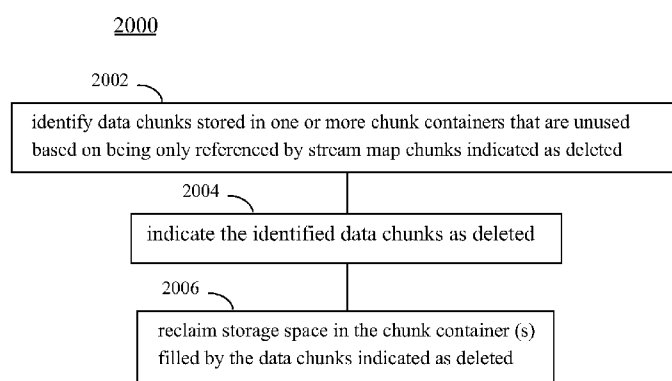
Published:

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

(88) Date of publication of the international search report:

27 December 2012

(54) Title: GARBAGE COLLECTION AND HOTSPOTS RELIEF FOR A DATA DEDUPLICATION CHUNK STORE

**FIG. 20**

(57) **Abstract:** Techniques for garbage collecting unused data chunks in storage are provided. According to one implementation, data chunks stored in a chunk container that are unused are identified based an analysis of one or more stream map chunks indicated as deleted. The identified data chunks are indicated as deleted. The storage space in the chunk container filled by the data chunks indicated as deleted may then be reclaimed. Techniques for selectively backing up data chunks are also provided. According to one implementation, a data chunk is received for storing in a chunk container. A backup copy of the received data chunk is stored in a backup container if the received data chunk is in a predetermined top percentage of most referenced data chunks in the chunk container and has a number of references greater than a predetermined reference threshold.

A. CLASSIFICATION OF SUBJECT MATTER**G06F 12/00(2006.01)i, G06F 12/16(2006.01)i, G06F 11/14(2006.01)i, G06F 9/44(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 12/00; G06F 12/02; H03M 7/00; H03M 7/46

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

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

eKOMPASS(KIPO internal) & Keywords: stream, chunk, deduplication, garbage collect, reference count

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	US 5990810 A (WILLIAMS, ROSS NEIL) 23 November 1999 See column 22, lines 22-41; figure 26.	1, 10, 11 2-9, 12-15
Y A	"Reclaiming space in a Backup Exec Deduplication Folder" , Article ID: TECH 130103, Created: 2010-01-17. Article URL http://www.symantec.com/docs/TECH130103	1, 10, 11 2-9, 12-15
Y A	US 7567188 B1 (ANGLIN, MATTHEW J. et al.) 28 July 2009 See column 7, lines 26-27; column 8, lines 18-61.	1, 10, 11 2-9, 12-15
A	Benjamin Zhu et al. "Avoiding the disk bottleneck in the data domain deduplication file system" Published in: Proceeding FAST'08 Proceedings of the 6th USENIX Conference on File and Storage Technologies. February 26-29 2008, SAN JOSE, CALIFORNIA. See pages 272 - 273, 3. Deduplication Storage System Architecture.	1-15
A	US 2010-0223441 A1 (LILLIBRIDGE, MARK DAVID et al.) 02 September 2010 See paragraphs [0009] - [0011], [0044] - [0048].	1-15



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

24 AUGUST 2012 (24.08.2012)

Date of mailing of the international search report

27 AUGUST 2012 (27.08.2012)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan
City, 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

JANG Ho Keun

Telephone No. 82-42-481-8187



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2011/065662

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5990810 A	23.11.1999	WO 96-25801 A1	22.08.1996
US 7567188 B1	28.07.2009	None	
US 2010-0223441 A1	02.09.2010	CN 101855619 A	06.10.2010
		US 2010-0198792 A1	05.08.2010
		WO 2009-054827 A1	30.04.2009