



- (51) **International Patent Classification:**
G06N 99/00 (2019.01) *G06N 3/04* (2006.01)
- (21) **International Application Number:**
PCT/CN2021/093933
- (22) **International Filing Date:**
14 May 2021 (14.05.2021)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
16/890,091 02 June 2020 (02.06.2020) US
- (71) **Applicant:** **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).
- (71) **Applicant (for MG only):** **IBM (CHINA) CO., LIMITED** [CN/CN]; 7F, Bldg 10, Zhangjiang Innovation Park, 399 Keyuan Road, Zhangjiang High-Tech Campus, Pudong New Area, Shanghai 201203 (CN).
- (72) **Inventors:** **ZHANG, Qi**; 1101 Kitchawan Rd., Room 33-153, Yorktown Heights, New York 10598 (US). **NOVOTNY, Petr**; T.J. Watson Research Center, 1101 Kitchawan Rd., Yorktown Heights, New York 10598 (US). **MIN, Hong**; P.O. Box 218, 1101 Kitchawan Rd., York-

town Heights, New York 10598 (US). **NAIR, Ravi**; P.O. Box 218, 1101 Kitchawan Rd., Yorktown Heights, New York 10598 (US). **RAMJI, Shyam**; 1101 Kitchawan Rd., Yorktown Heights, New York 10598 (US). **YU, Lei**; 1101 Kitchawan Rd., P.O. Box 218, Yorktown Heights, New York 10598 (US). **NAKAIKE, Takuya**; 19-21 Nihonbashi-Hakozakicho, Chuo-Ku, Tokyo 103-8510 (JP). **KAWAHITO, Motohiro**; 19-21 Nihonbashi-Hakozakicho, Chuo-Ku, Tokyo 103-8510 (JP).

- (74) **Agent:** **KING & WOOD MALLESONS**; 20th Floor, East Tower, World Financial Centre, No. 1 Dongsanhuan Zhonglu, Chaoyang District, Beijing 100020 (CN).

- (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (54) **Title:** STREAMLINING DATA PROCESSING OPTIMIZATIONS FOR MACHINE LEARNING WORKLOADS

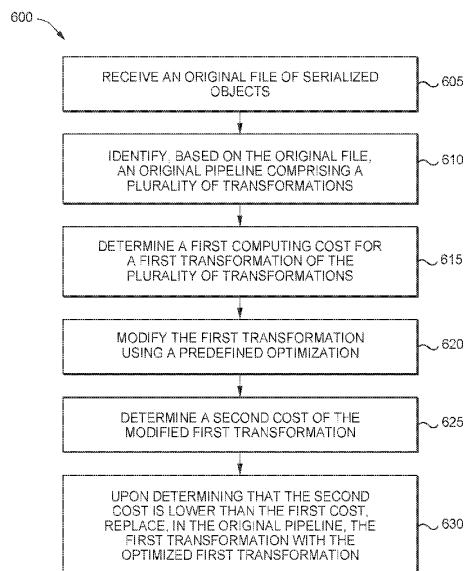


FIG. 6

- (57) **Abstract:** Techniques for refinement of data pipelines are provided. An original file of serialized objects is received, and an original pipeline comprising a plurality of transformations is identified based on the original file. A first computing cost is determined for a first transformation of the plurality of transformations. The first transformation is modified using a predefined optimization, and a second cost of the modified first transformation is determined. Upon determining that the second cost is lower than the first cost, the first transformation is replaced, in the original pipeline, with the optimized first transformation.

(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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Published:

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