

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号
WO 2022/252839 A1

(43) 国际公布日
2022 年 12 月 8 日 (08.12.2022)

(51) 国际专利分类号:
G06F 9/48 (2006.01)

(21) 国际申请号: PCT/CN2022/086761

(22) 国际申请日: 2022 年 4 月 14 日 (14.04.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110620358.4 2021 年 6 月 3 日 (03.06.2021) CN

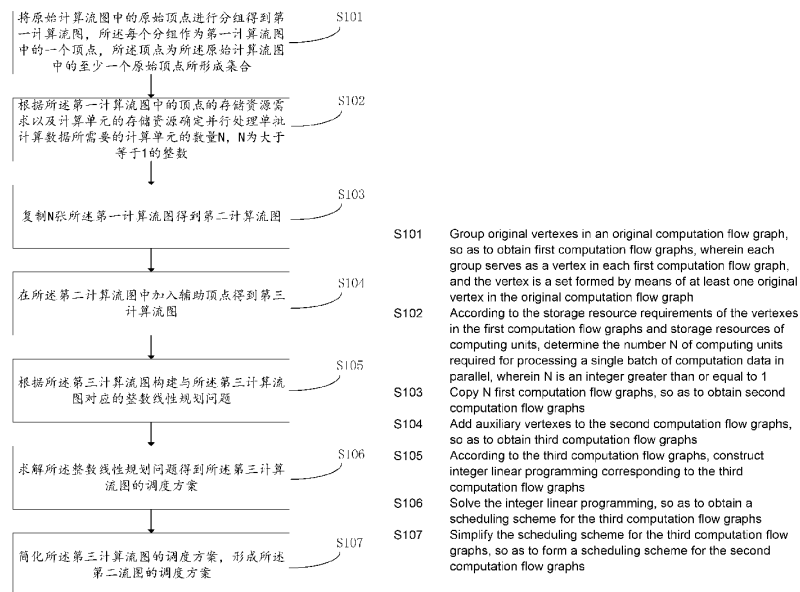
(71) 申请人: 北京希姆计算科技有限公司 (BEIJING STREAM COMPUTING INC.) [CN/CN]; 中国北京市朝阳区来广营西路 5 号路 5 号楼 8 层 801、802、803、808, Beijing 100012 (CN)。

(72) 发明人: 曹睿 (CAO, Rui); 中国北京市朝阳区来广营西路 5 号路 5 号楼 8 层 801、802、803、808, Beijing 100012 (CN)。 吕文媛 (LV, Wenyuan); 中国北京市朝阳区来广营西路 5 号路 5 号楼 8 层 801、802、803、808, Beijing 100012 (CN)。 淡孝强 (DAN, Xiaoqiang); 中国北京市朝阳区来广营西路 5 号路 5 号楼 8 层 801、802、803、808, Beijing 100012 (CN)。 刘雷 (LIU, Lei); 中国北京市朝阳区来广营西路 5 号路 5 号楼 8 层 801、802、803、808, Beijing 100012 (CN)。

(74) 代理人: 北京睿派知识产权代理事务所 (普通合伙) (WITPAT INTELLECTUAL PROPERTY LAW FIRM); 中国北京市朝阳区阜通东大街 6 号院 2 号楼 12 层 1508, Beijing 100102 (CN)。

(54) Title: METHOD AND APPARATUS FOR GENERATING COMPUTATION FLOW GRAPH SCHEDULING SCHEME, AND ELECTRONIC DEVICE AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 计算流图调度方案的生成方法、装置、电子设备及计算机可读存储介质



(57) Abstract: Disclosed in the embodiments of the present disclosure are a method and apparatus for generating a computation flow graph scheduling scheme, and an electronic device and a computer-readable storage medium. The method for generating a computation flow graph scheduling scheme comprises: grouping original vertexes in an original computation flow graph, so as to obtain first computation flow graphs; determining the number N of computing units required to process a single batch of computation data in parallel;

[见续页]

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): 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, JM, 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。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (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)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

copying N first computation flow graphs, so as to obtain second computation flow graphs; adding auxiliary vertexes to the second computation flow graphs, so as to obtain third computation flow graphs; constructing integer linear programming according to the third computation flow graphs; and solving the integer linear programming, so as to obtain a scheduling scheme for the third computation flow graphs. By means of the method for generating a computation flow graph scheduling scheme, an original computation flow graph is converted into third computation flow graphs, and integer linear programming is constructed, so as to obtain a scheduling scheme, thereby solving the technical problem in the prior art of a low data reuse ratio or a low degree of parallelism.

(57) 摘要: 本公开实施例公开了一种计算流图调度方案的生成方法、装置、电子设备和计算机可读存储介质。其中该计算流图调度方案的生成方法包括: 将原始计算流图中的原始顶点进行分组得到第一计算流图; 确定并行处理单批计算数据所需要的计算单元的数量N; 复制N张所述第一计算流图得到第二计算流图; 在所述第二计算流图中加入辅助顶点得到第三计算流图; 根据所述第三计算流图构建整数线性规划问题; 求解所述整数线性规划问题得到所述第三计算流图的调度方案。上述计算流图调度方案的生成方法通过将原始计算流图转换成第三计算流图并构建整数线性规划问题以求解调度方案, 解决了现有技术中数据重用率低或并行度低的技术问题。