論文 / 著書情報
Article / Book Information

題目(和文)	局所性を考慮した大規模グラフ解析向けグラフデータストア
Title(English)	Locality-aware Graph Data Store for Large-scale Graph Analytics
著者(和文)	岩渕圭太
Author(English)	Keita Iwabuchi
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第10435号, 授与年月日:2017年3月26日, 学位の種別:課程博士, 審査員:松岡 聡,南出 靖彦,渡辺 治,遠藤 敏夫,脇田 建
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第10435号, Conferred date:2017/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	数理・計算科学　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of　（　理学　）
学生氏名： Student's Name	岩渕　圭太	指導教員（主）： Academic Advisor(main)	松岡　聡　教授
		指導教員（副）： Academic Advisor(sub)	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Big data processing brings us many challenges attributed to not only its volume but also to the emergence of a new paradigm; that is, analyzing the data to discover knowledge, to understand behaviors, and to mine for patterns accompanied with complex memory access patterns on large volume of data. At the same time, demands for large-scale graph analytics has risen as an important kernel for high-performance computing (HPC) applications in various domains, such as WWW and social network analysis, network security, artificial intelligence and genomic analysis. Meanwhile, the interest in non-volatile random-access memory (NVRAM) such as NAND flash, phase change memory (PCM) and resistive RAM (ReRAM) has risen due to the cost and high power consumption of DRAM. However, large-scale graph analytics often presents challenging data-intensive workloads because of unstructured and random memory access patterns. Therefore, designing locality-aware data stores for graph analytics is an extremely important key factor to enable high-performance graph analytics.

To address the issues above, we explore techniques and designs of large-scale graph data stores from the perspective of static and dynamic graph analytics.

First, we explore important techniques for out-of-core static graph stores by developing a high performance out-of-core breadth-first search (BFS) implementation. Specifically, we propose NETALX, an extremely high performance BFS implementation using NVRAM for Hybrid BFS algorithm which devises the arrangement of graph data on DRAM and NVRAM to improve data locality (reduce the number of accesses to NVRAM) and sequential locality in NVRAM. Experimental results compliant to the Graph500 benchmark on a single compute node with arrays of NAND flash-based SSDs show that NETALX can achieve 4.14 Giga TEPS (Traversed Edges Per Second) for a graph with 2^{31} vertices and 2^{35} edges, whose size is 4 times larger than the size of graphs that the machine can accommodate only using DRAM, with only 14.99% performance degradation. We also demonstrate that NETALX can achieve a power efficiency of 11.8 Mega TEPS/W (Traversed Edges Per Second / Watt).

Second, for large-scale dynamic graph analytics, we propose DegAwareRHH, a high performance dynamic graph data store, which leverages a linear probing open addressing compact hash table that exhibits high spatial and sequential locality to increase graph update performance keeping graph analytics performance. To extend DegAwareRHH to distributed-memory platforms, we adopt an asynchronous communication framework aiming for localizing remote communication into the area where need to be updated. We demonstrate that DegAwareRHH is 212.2 times faster than a state-of-the-art shared-memory streaming graph processing framework on a single compute node to update a graph with 1 billion edge insertion requests and 54 million edge deletion requests. DegAwareRHH achieves a processing rate of over 1.8 billion edge insertion requests per second at 192 compute nodes on a massive-scale real graph that has 128 billion edges. We also show that DegAwareRHH can accelerate the performance of a large-scale dynamic graph colouring algorithm and achieve high performance on out-of-core graph update workloads including future NVRAM devices.

This thesis presents several contributions towards high performance data stores for large-scale graph analytics, in terms of locality awareness, on HPC platforms, including next generation supercomputers which will have locally-attached NVRAM.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).