Database Group

Architecture for Stream OLAP Exploiting SPE and OLAP Engine

Explosive increase of real-time data streams gives highly demands for real-time analysis over the streams. However, developing tailor-made systems for such applications is not always desirable due to high developing costs. To cope with this problem, we propose a novel architecture for online analytical processing (OLAP) over streams exploiting off-the-shelf stream processing engine (SPE) combined with OLAP engine. It allows users to perform OLAP analysis over streams for the latest time period, called Interval of Interest (IoI). The system in the meantime processes multiple continuous query language (CQL) queries corresponding to different aggregation levels in cube lattice. To cover arbitrary aggregation levels using limited system's memory, we propose to partially deploy CQL queries for those with higher reference frequencies, whereas the results are dynamically calculated using existing aggregation results with the help of OLAP engine. For optimal CQL query deployment, we propose a cost-based optimization method that maximizes the performance. The experimental results show that the proposed system significantly outperforms other comparative methods.

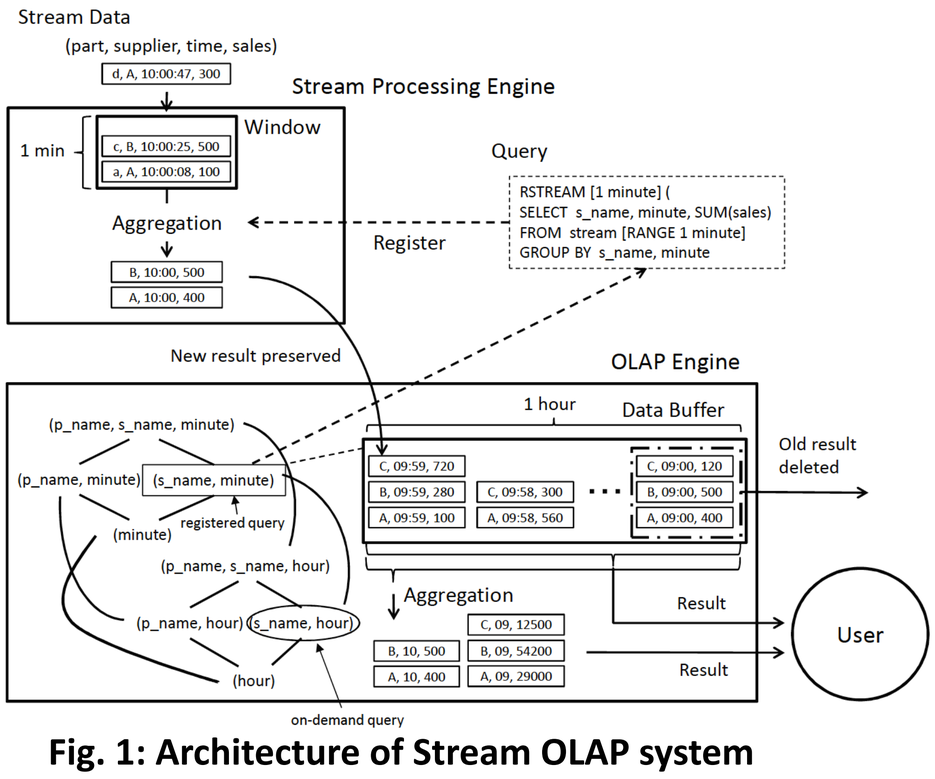

Fig. 1: Architecture of Stream OLAP system

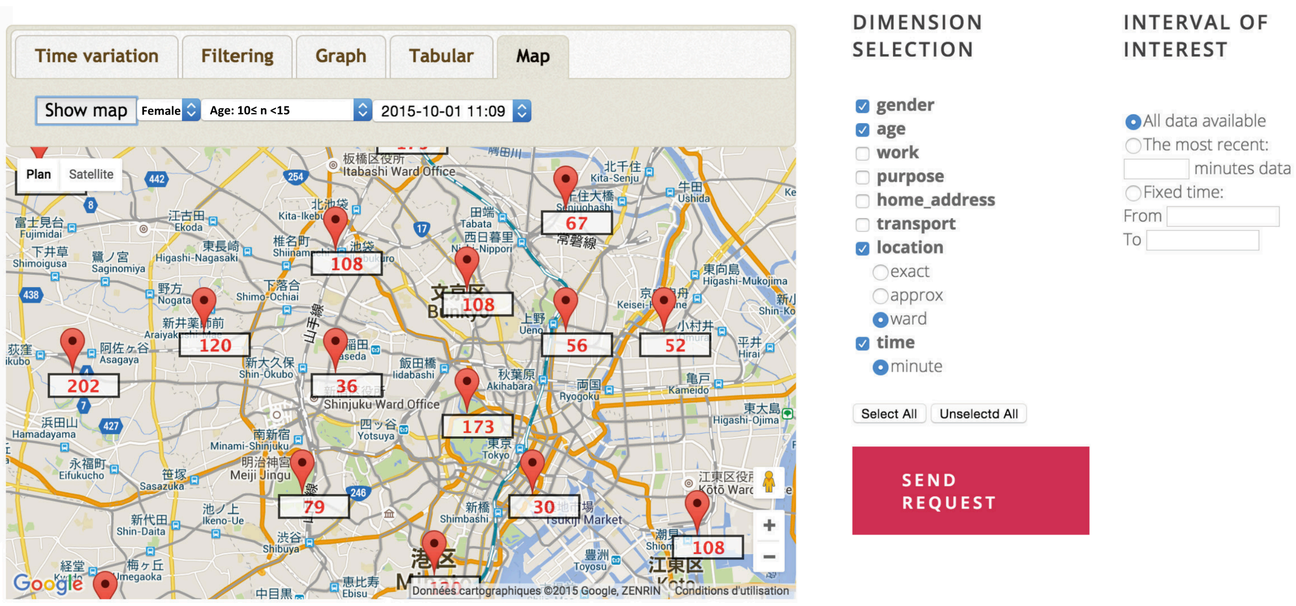

Fig. 2: Stream OLAP system analyzing "People Flow Data"

Scalable Graph Analysis over Intel Xeon Phi Coprocessors

The structural graph clustering method *SCAN* is successfully used in many applications since it detects not only densely connected nodes as clusters but also extracts sparsely connected nodes as hubs or outliers (Fig. 3) . However, it is difficult to apply SCAN to large-scale graphs since SCAN needs to evaluate the density for all adjacent nodes included in the graph. In this work, so as to address the above problem, we present a novel algorithm *SCAN-XP* that performs on Intel Xeon Phi coprocessors. We designed SCAN-XP to make best use of many cores in the Intel Xeon Phi by employing the following approaches: First, SCAN-XP avoids the bottlenecks that arise from parallel graph computations by providing good load balances among the cores. Second, SCAN-XP effectively exploits 512 bit SIMD instructions implemented in each core to speed up the density evaluations. As a result, SCAN-XP runs approximately 100 times faster than SCAN; for the graphs with 100 million edges, SCAN-XP is able to perform in a few seconds (Fig. 4).)

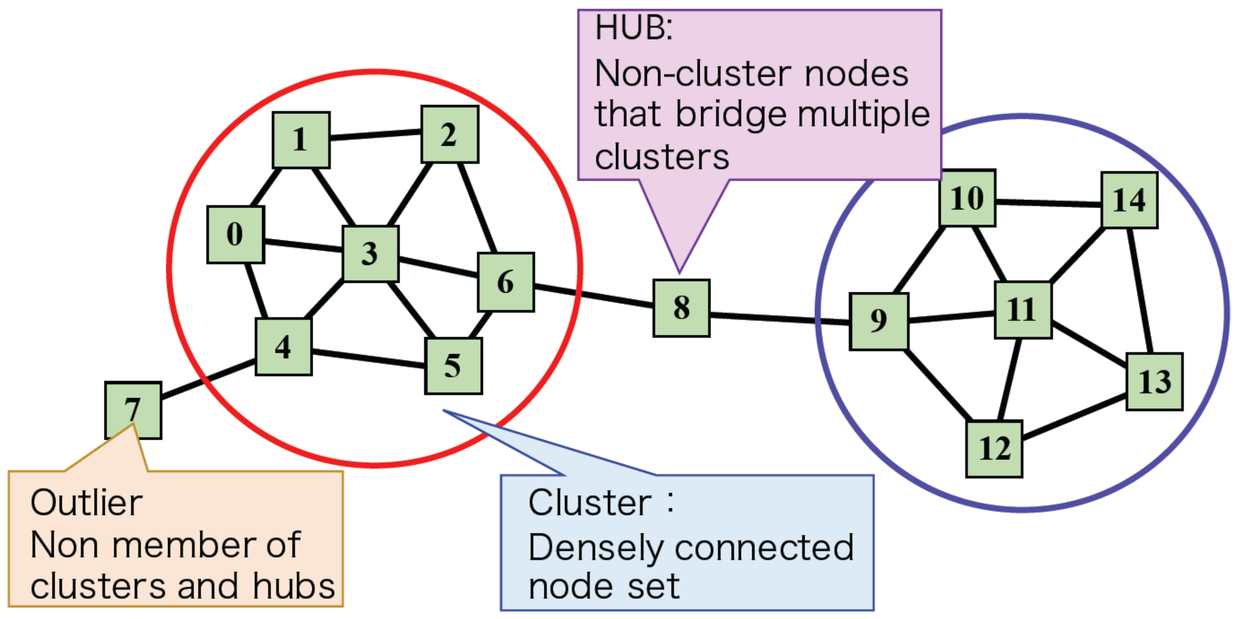

Fig. 3: Structural Graph Clustering SCAN

Table. 1: Real-world Dataset

Dataset	# of nodes	# of edges
com-youtube	1,134,890	2,987,624
web-BerkStan	685,230	6,649,470
soc-Pokec	1,632,803	22,301,964
com-LiveJournal	3,997,962	34,681,189
soc-LiveJournal1	4,846,609	42,851,237
com-Orkut	3,072,441	117,185,083
webbase2001	115,554,441	854,809,761

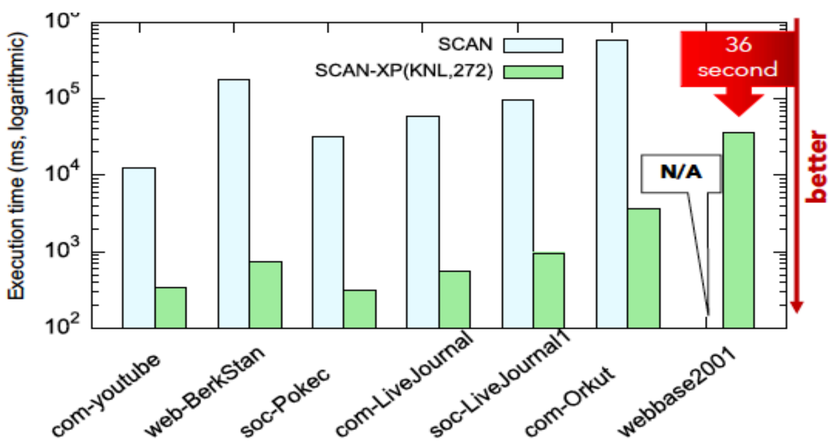

Fig. 4: Overall performances

contact address: {kitagawa, amagasa, shiokawa}@cs.tsukuba.ac.jp