



University of Electronic Science and Technology of China

Title: A Graph Database Storage Engine for Provenance Graphs



The First Author: Changhong Liu



The Second author: Hancong Duan



Presenter: Changhong Liu

Email: 314979677@qq.com





personal resume

Name: Changhong Liu

Organization: School of Computer Science and Engineering, University of Electronic Science and Technology, Chengdu, China.

I obtained my bachelor's degree from College of Computer & Information Science Southwest University in 2018. Now I'm studying for a master's degree in the school of computer science and Engineering of UESTC. My research field is the distributed database and storage. Recently, I have been working on graph computing and graph database.

Project experience:

- Developing a distributed memory column database
- Developing a distributed graph database



Challenges

Challenges

- 1. A great graph-partitioning algorithm can reduce the communication overhead in the distributed graph database. A great graph-repartitioning algorithm can keep the load balance of the distributed graph database.**
- 2. Ensuring that the proximity of graph-structured data is maintained on disk.**
- 3. It is a challenge to implement the distributed transaction because of the uniqueness of distributed graph database.**



Proposed Solution

About the first challenge:

- Large scale graph is partitioned in a distributed way.
- Streaming partition algorithm such as HDRF.

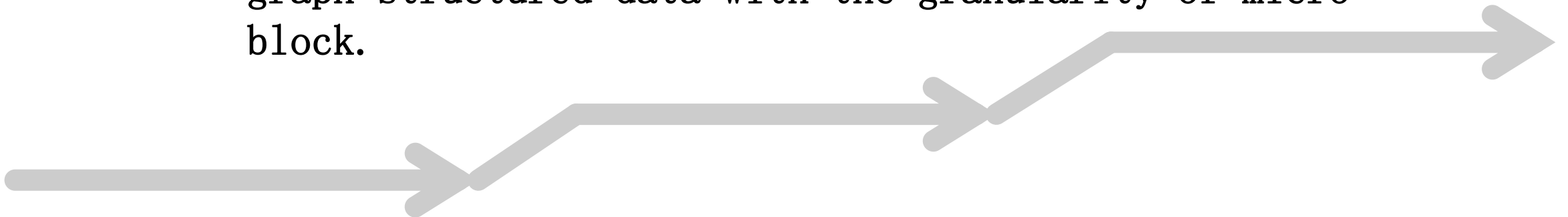




Proposed Solution

About the second challenge:

- Each node and edge is assigned an store ID. We have designed three algorithms to assign store ID: Hash, Layer Priority, deepwalk-partition.
- Each node and edge is sequentially stored in the file by sorting the store ID. Then the database reads the graph-structured data with the granularity of micro block.





Proposed Solution

About the third challenge:

We created several different types of locks to ensure that the transaction of our graph database can reach RR isolation level and try not to affect the parallelism of transactions.

	Node_Lock	Topo_Lock	Prop_Lock	Read
Node_Lock	x	x	x	x
Topo_Lock	x	✓	✓	✓
Prop_Lock	x	✓	✓	✓
Read	x	✓	✓	✓





Project

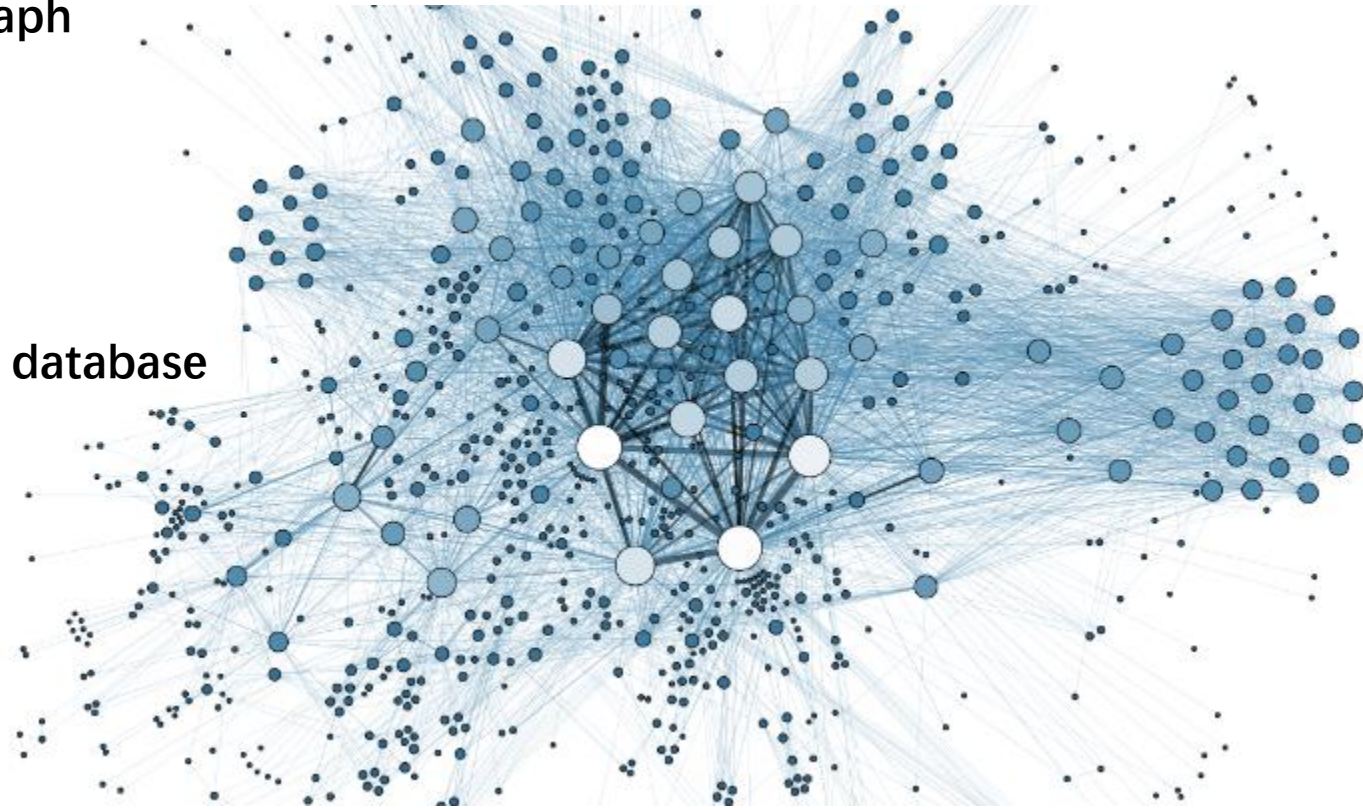
Distributed Graph Database

Background:

- Storing massive graph-structured data
- Meeting the requirements of knowledge graph
- Can query time-series data
- Distributed graph computing

Date: 2018.09 ~ up to now

Status: We have a complete distributed graph database system. We are further optimizing this system.





Future plans

- Researching efficient distributed computing scheduling algorithm.
- Researching a high performance distributed storage system based on new hardware and new technology.
- Researching the optimization of database query and storage through machine learning.





THANKS FOR YOUR WATCHING

