



UNIVERSITÄT
LEIPZIG

Big Data Visualisation

Viewport-Driven Graph Data Reduction

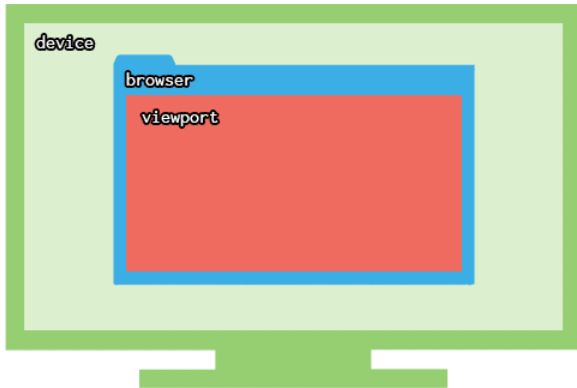
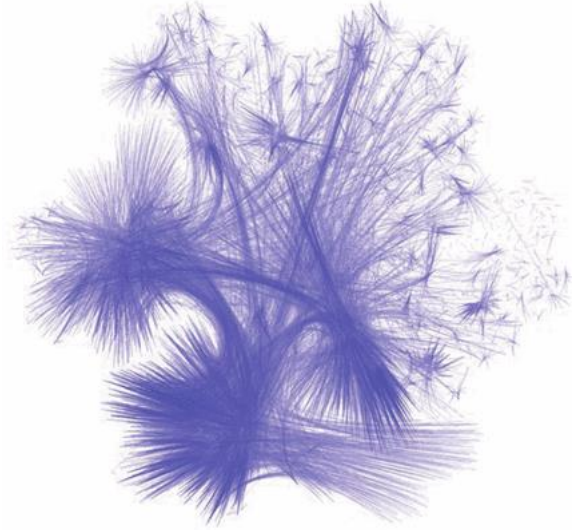
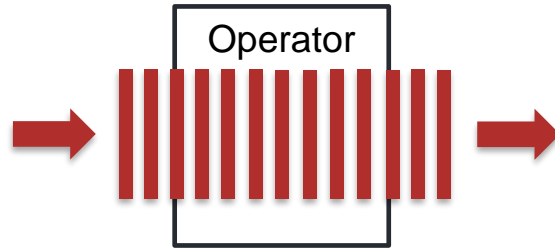
Leipzig, 11.12.2020

Aljoscha Rydzyk



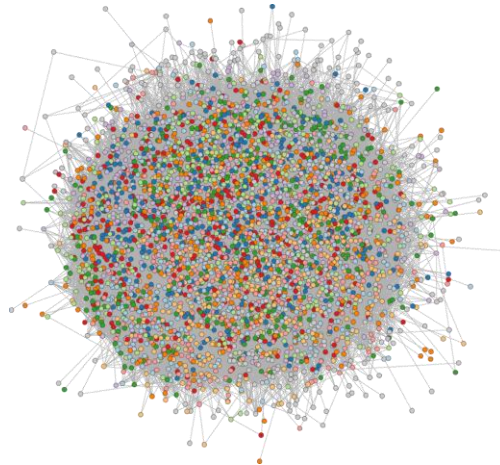
Overview

- Introduction and Related Work
 - Terms and Definitions
 - The Challenge of Big Data
 - Viewport-Driven Data Reduction (VDDR)
- VDDR on Graphs – a New Approach:
 - Application Setup
 - Back End Data Representation
- Evaluation
- To Do List
- Example Presentation



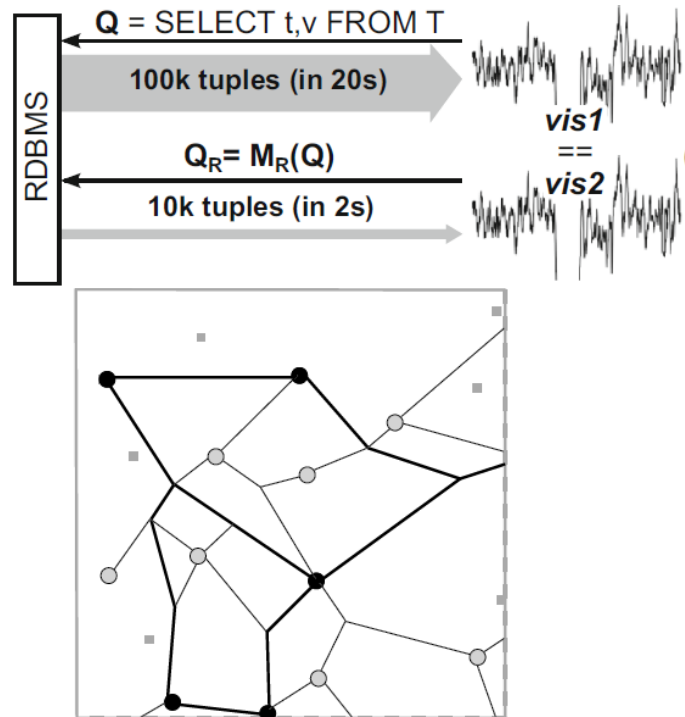
The Challenge of Big Data

- Increasing amount of network-type data
- Big Data Graph
 - Keep representation clear and focused
 - Data reduction

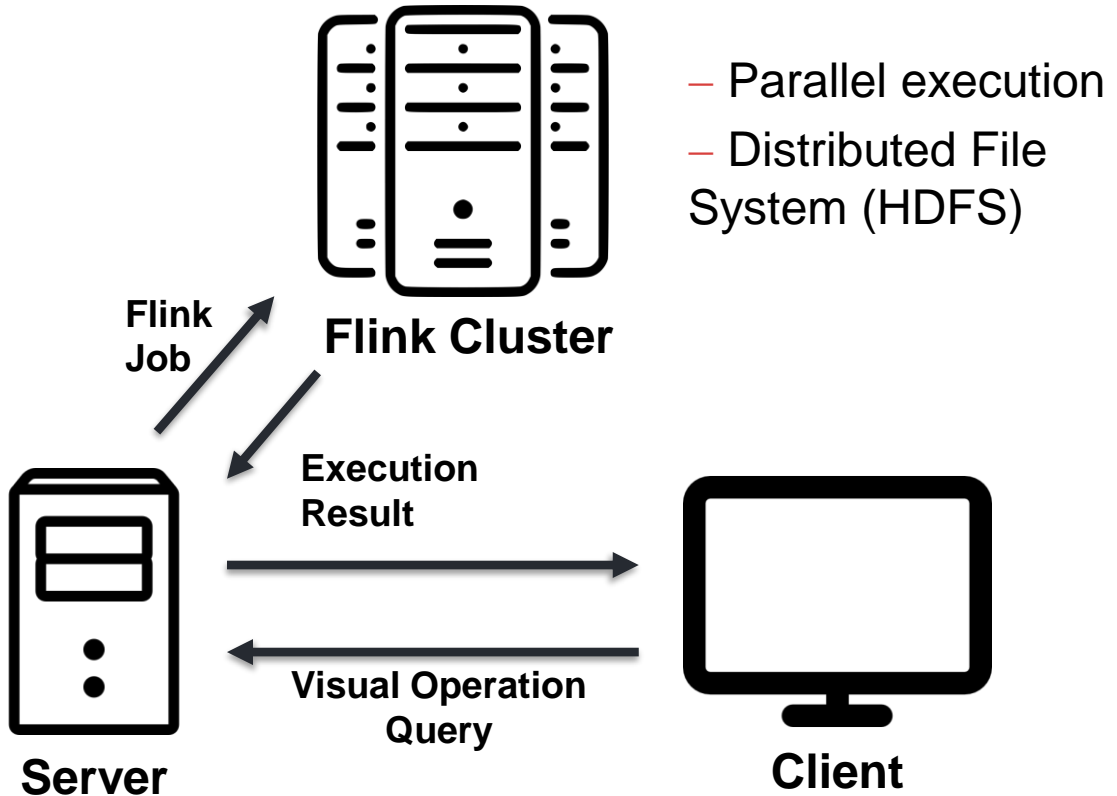


Viewport-Driven Data Reduction

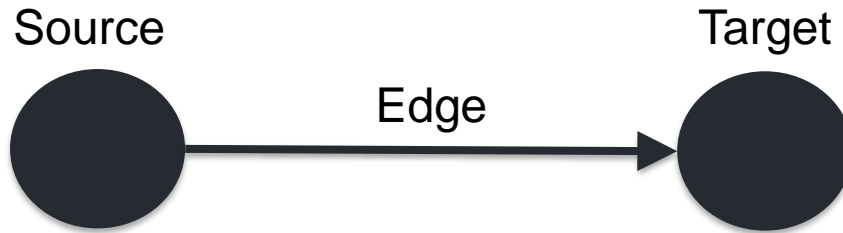
Viewport-driven data aggregation in relational data bases (Jugel, Jerzak et al. 2015)



A new approach to GraphMaps (Mondal and Nachmanson 2017)



Stream Data Object – „Wrapper“



- Vertices' label
- Edge label
- Vertices' degree
- Vertices' ID
- Edge ID
- ...



Back End Data Representation

3 Different Approaches:

Gradoop:

- edge set, vertex set
- batch baseline

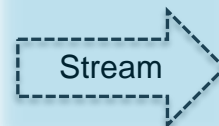
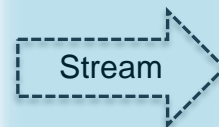
(Junghanns, Petermann et al. 2017)

Direct Wrapper Stream:

- wrapper and vertex stream source
- data sorted by degree

Adjacency Matrix:

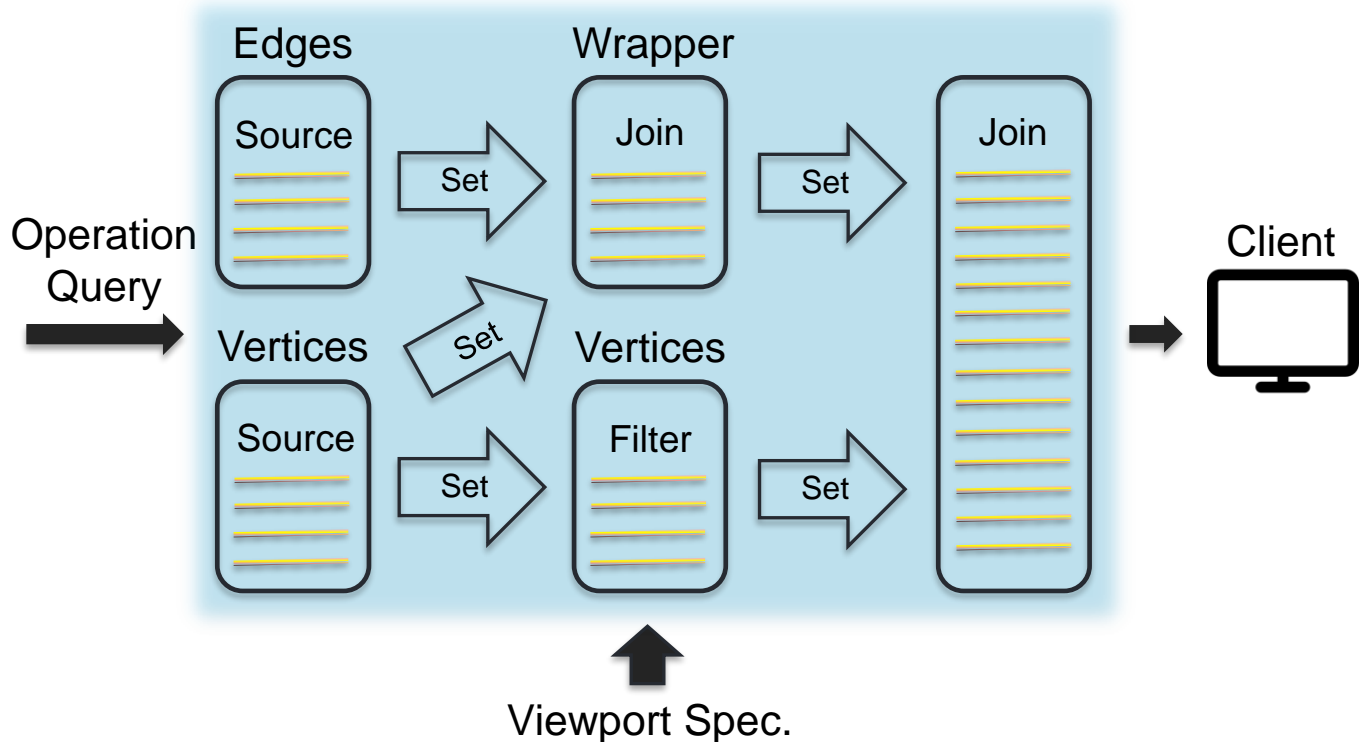
- adjacency matrix, vertex stream source, wrapper map
- data sorted by degree



Back End Data Representation



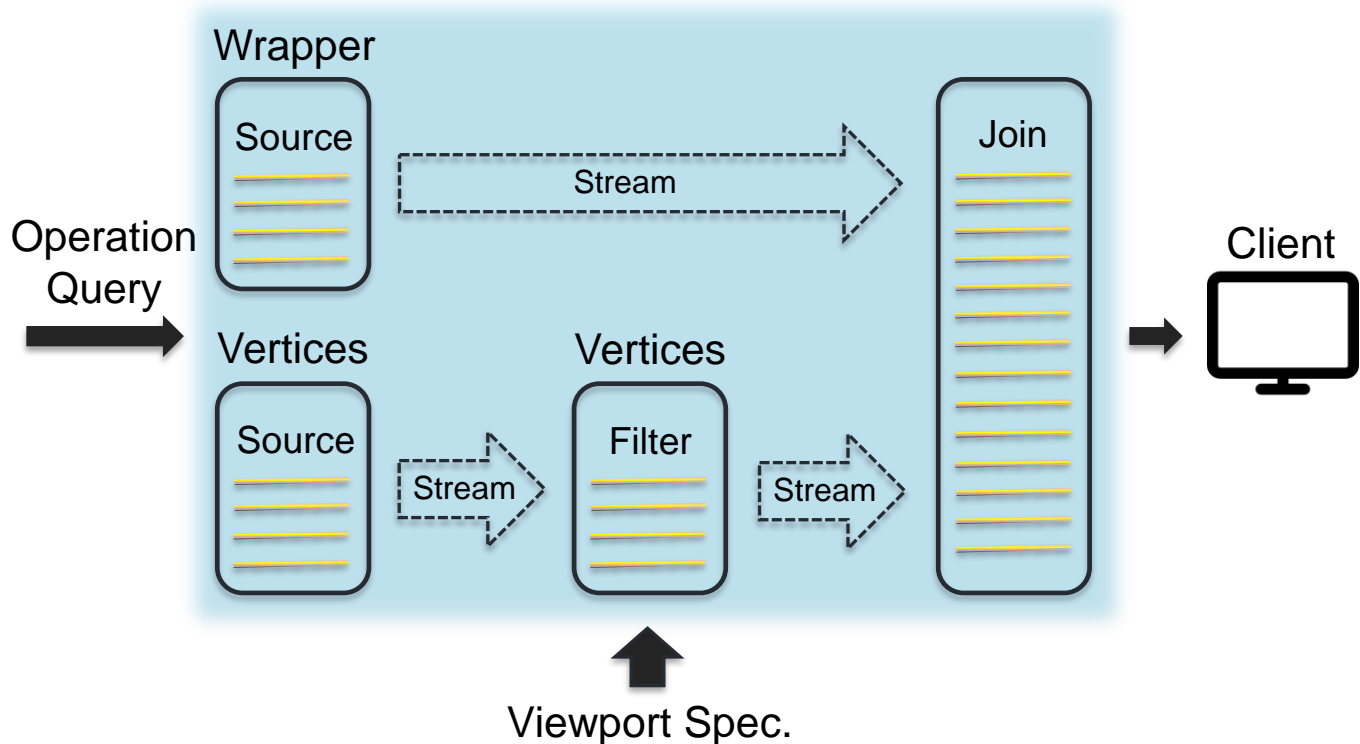
- Batch





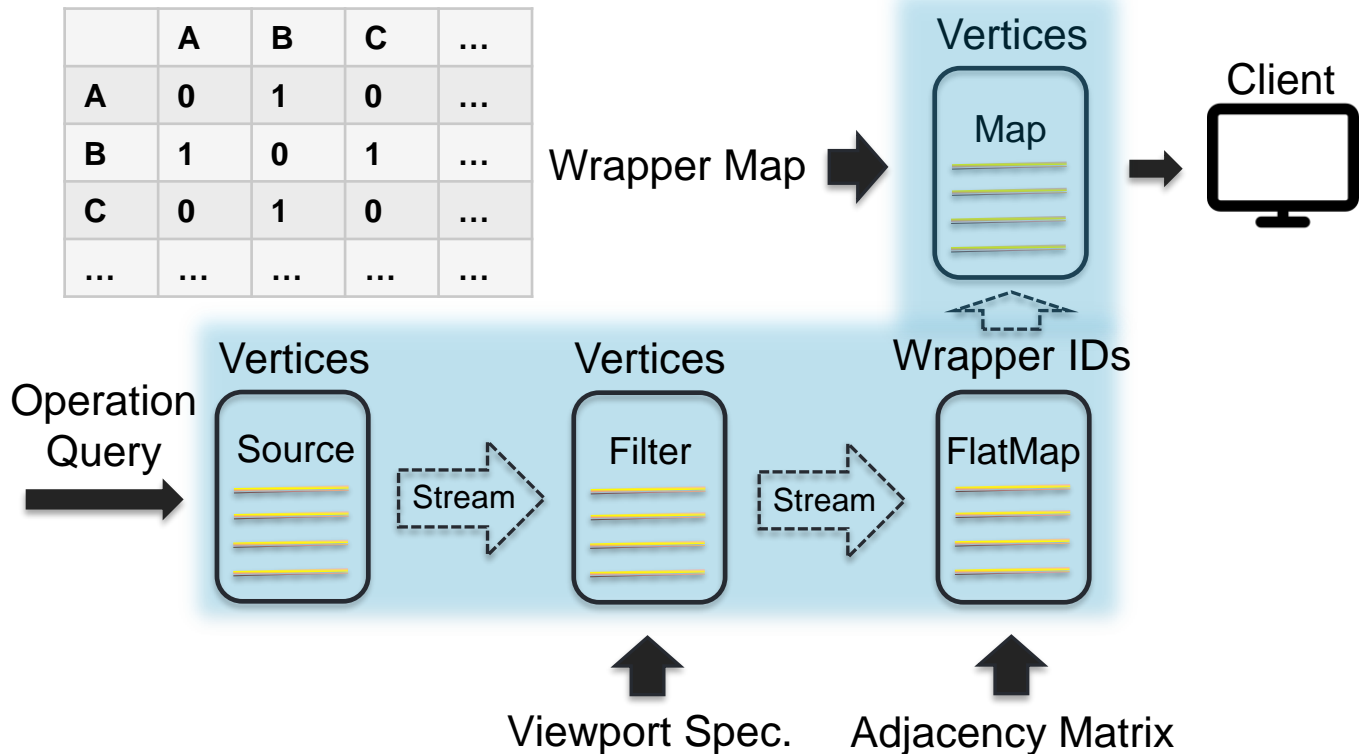
Back End Data Representation

Direct Wrapper - Stream



Back End Data Representation

Adjacency Matrix - Stream



Evaluation

- Evaluation on Galaxy Cluster of Leipzig University
- Parameters:
 - Different approaches (stream, batch)
 - Parallelism
 - Graph size (gigabyte scale)
 - Pre-layouted and non-layouted graphs
- Measured Quantities:
 - Back End procedure time scale
 - Back End memory consumption
 - Server-Client data transfer time scale
 - Layout rendering time scale



To Do List

- Client-side ad-hoc-layout
- Evaluation
- Deployment and integration into other services



Sources

- Jugel, U., et al. (2015). "VDDA: automatic visualization-driven data aggregation in relational databases." The VLDB Journal **25**(1): 53-77.
- Mondal, D. and L. Nachmanson (2017). "A new approach to GraphMaps, a system browsing large graphs as interactive maps." arXiv preprint arXiv:1705.05479.
- Junghanns, M., et al. (2017). "Distributed grouping of property graphs with GRADOOP." Datenbanksysteme für Business, Technologie und Web (BTW 2017).
- <https://dbs.uni-leipzig.de/en/research/projects/gradoop>
- <https://flink.apache.org/>



UNIVERSITÄT
LEIPZIG

Thank You!

Aljoscha Rydzyk
ar44xogu@studserv.uni-leipzig.de