Knowledge Graphs at Your Fingertips

Exploring Opportunities and Challenges of Embedded OLAP for Graph Analytics

Dmitrii Orlov   d.orlov@student.tue.nl

Nikolay Yakovets   hush@tue.nl

Daphne Miedema   d.e.miedema@uva.nl

DBDBD 2024   24/11/22

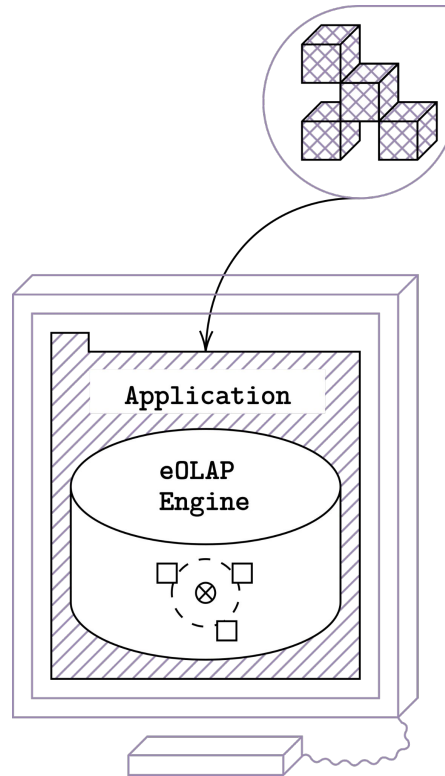# Embedded Online Analytical Processing

eOLAP

Fast analytics +

Efficient data transfer +

Uniting code and system +

Low engineering effort +

Easy integration +

Single-tenant -

Application

eOLAP
Engine

" *integrating [OLAP] data warehouses with local tools **is cumbersome and slow***

" *[eOLAP] is co-inhabitation of DBMS and analysis application code on [one] computer*

Raasveldt & Mühleisen, (2020).
Data management for data science — Towards embedded analytics.

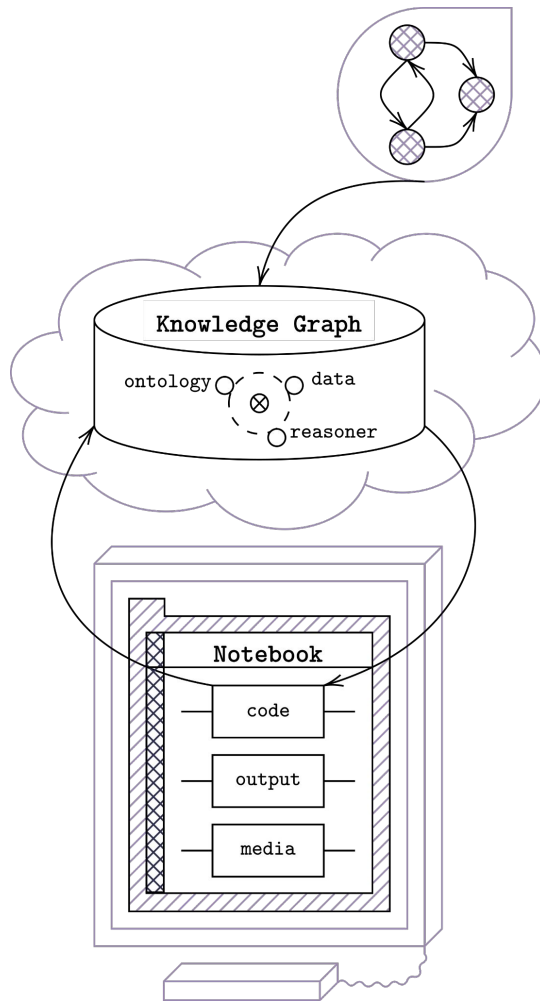## KG Systems

Efficient graph analytics +

Massive data capacity +

Multi-tenant +

Data transfer costs -

High engineering effort -

Split system and application -

" *Graph algorithms are often used for BI-ish applications [so people] call them* **graph OLAP**

Tian (2023).
The World of Graph Databases from An Industry Perspective.

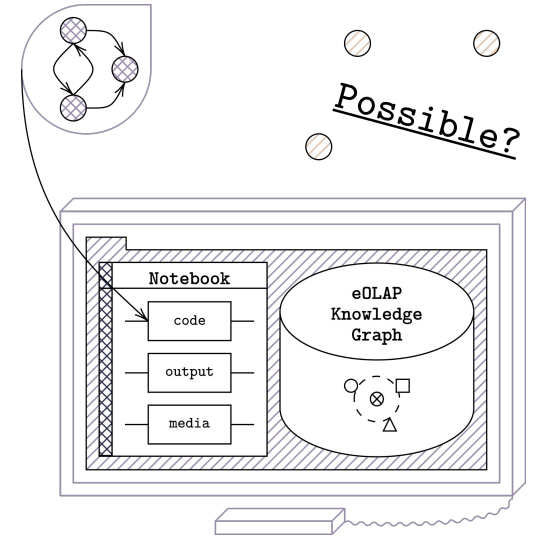" *[big] graph processing* **ecosystems match** *the common* **reference architecture** *of [OLAP]*

Sakr et al., (2021).
The Future Is Big Graphs: A Community View on Graph Processing Systems.

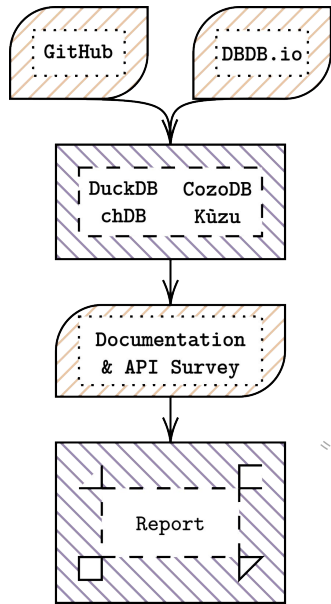RQ.1. What are the prominent **characteristics and features of** industrial-focused, open-source **eOLAP systems**?

RQ.2. What **technologies and tools** play a prominent role in **Knowledge Graph workflows** in real-world notebooks?

Possible?

Notebook

code

output

media

eOLAP
Knowledge
Graph

All four systems

➤ are **in-process Python** libraries
➤ connect **popular data science formats & tools**
➤ **adapt traditional OLAP** query and optimization

GitHub DBDB.io

DuckDB CozoDB
chDB Kùzu

Documentation & API Survey

Report

			DuckDB	CozoDB	chDB	Kùzu
⌐	Foundations	*Relational*				
		Graph				
⊏	I/O & Integration	*Scale*				
▽	Result Rendering	*Textual*				
		Visual				
▣	Custom Processing	*Flexibility*				

RQ.2. [...] technologies and tools [...] in Knowledge Graph [...] notebooks?
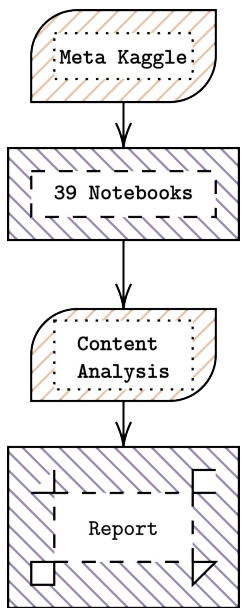
39 notebooks

➤ mostly **Pandas, NumPy, NetworkX, Matplotlib**
➤ embrace **'literate programming'** mindset
➤ flexible and **complex custom (pre)processing** steps

Meta Kaggle

39 Notebooks

Content Analysis

Report

⌐ Foundations

🔖 https://www.kaggle.com/code/kalilfagundes/knowledge-graph-of-quadrilha-a-poem-by-drummond

Pandas + NetworkX

```
#Here we use the columns from pandas dataframe to set the source and target
G=nx.from_pandas_edgelist(fd, "source", "target", ...)
```

Matplotlib

```
#define figure size
plt.figure(figsize=(20,20))
```

NetworkX

```
#define the layout type (it can be circular, flat, etc). ...
pos = nx.spring_layout(G, k=1)
#here we choose the drawing settings
nx.draw(G, with_labels=True, node_size=3000, node_color='skyblue', ...)
#set the edge labels, so the relations text will show in the plot
nx.draw_networkx_edge_labels(G, pos=pos)
#here you can change the color of the edge based in the value of "r"
nx.draw_networkx_edges(G,pos, width=1,edgelist=G.edges(), ...);
```
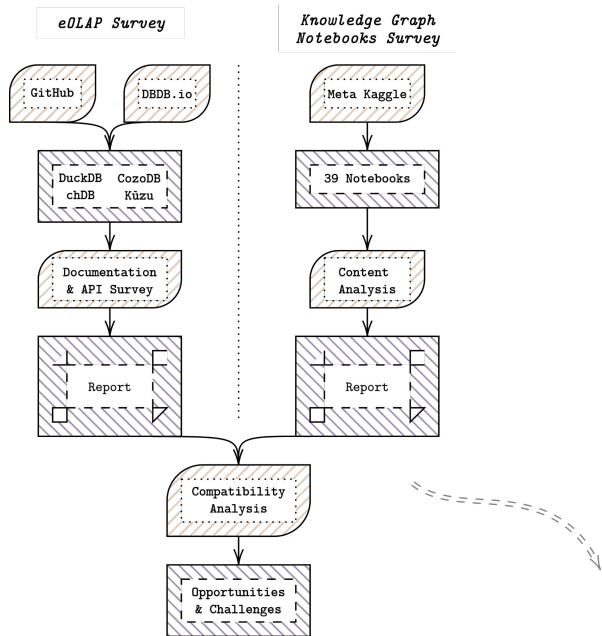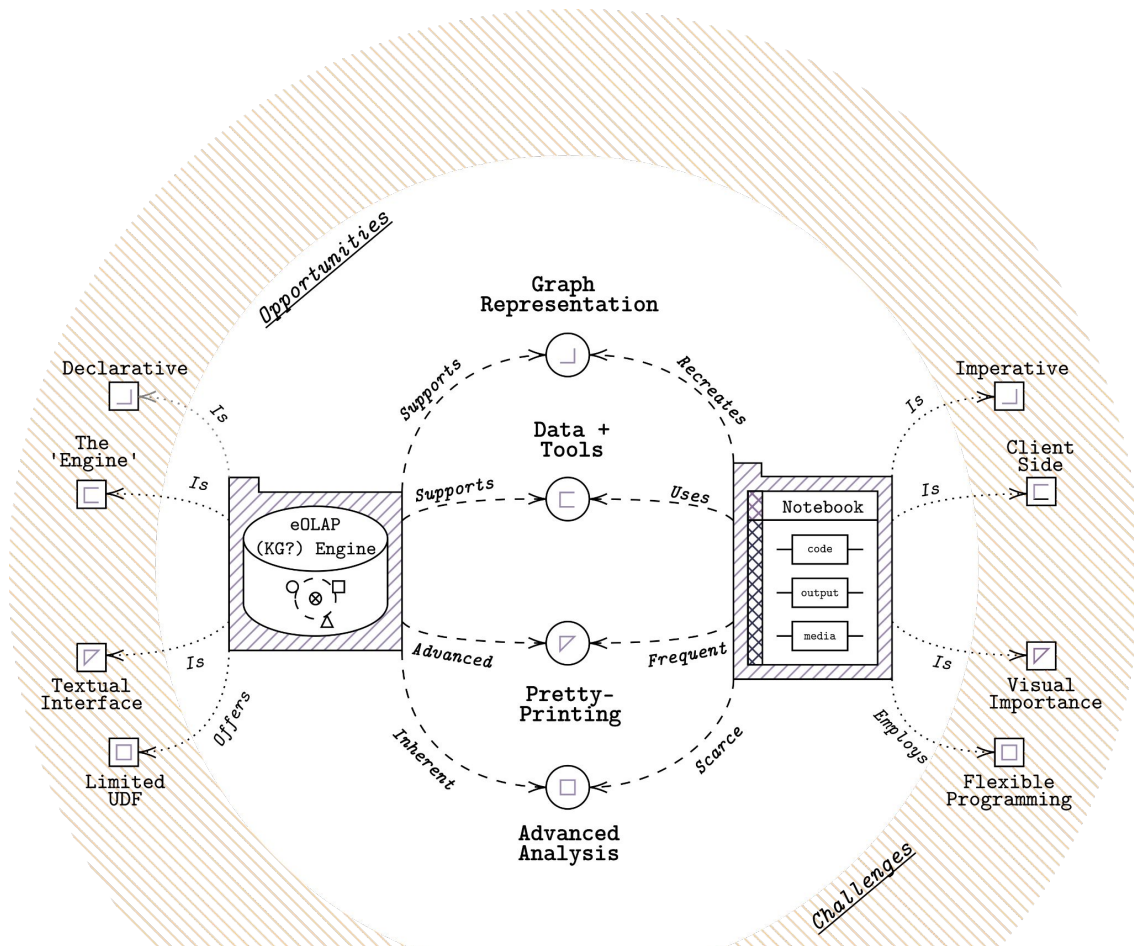
```
plt.show()
```

Matplotlib
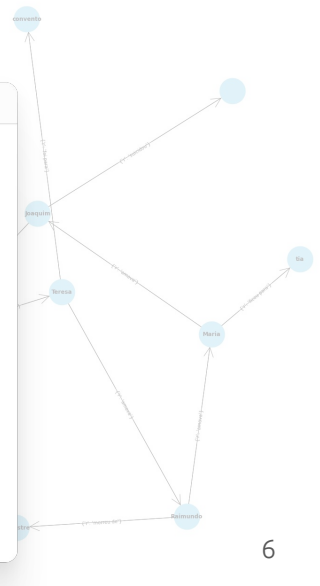
Opportunities and Challenges

Conclusion

Possible benefits
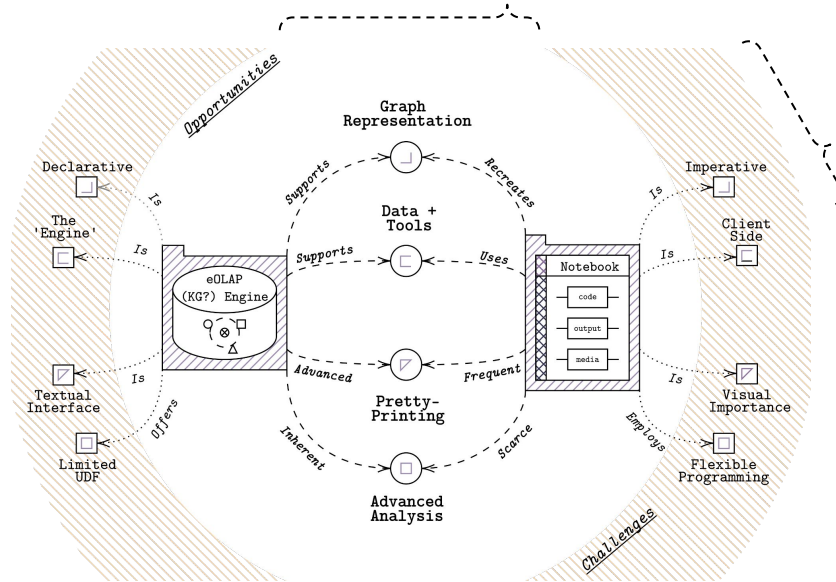warrant further investigation:
empirical, HCI & technical
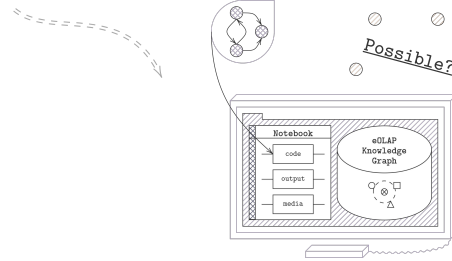
Differences require
careful design and
technical decisions

Conclusion

Thank you!

We will be happy to meet you at our poster

Possible benefits
warrant further investigation:
empirical, HCI & technical

Differences require
careful design and
technical decisions

Knowledge Graphs at Your Fingertips

Exploring Opportunities and Challenges of Embedded OLAP for Graph Analytics

Dmitrii Orlov   d.orlov@student.tue.nl
Nikolay Yakovets   hush@tue.nl
Daphne Miedema   d.e.miedema@uva.nl

Bibliography

Raasveldt, M., & Mühleisen, H. (2020). *Data management for data science—Towards embedded analytics.* https://ir.cwi.nl/pub/29372

Tian, Y. (2022). *The world of graph databases from an industry perspective* (arXiv:2211.13170). arXiv. https://doi.org/10.48550/arXiv.2211.13170

Sakr, S., Bonifati, A., Voigt, H., Iosup, A., Ammar, K., Angles, R., Aref, W., Arenas, M., Besta, M., Boncz, P. A., Daudjee, K., Valle, E. D., Dumbrava, S., Hartig, O., Haslhofer, B., Hegeman, T., Hidders, J., Hose, K., Iamnitchi, A., … Yoneki, E. (2021). The future is big graphs: A community view on graph processing systems. *Commun. ACM, 64*(9), 62–71. https://doi.org/10.1145/3434642

Appendix: Import Frequencies