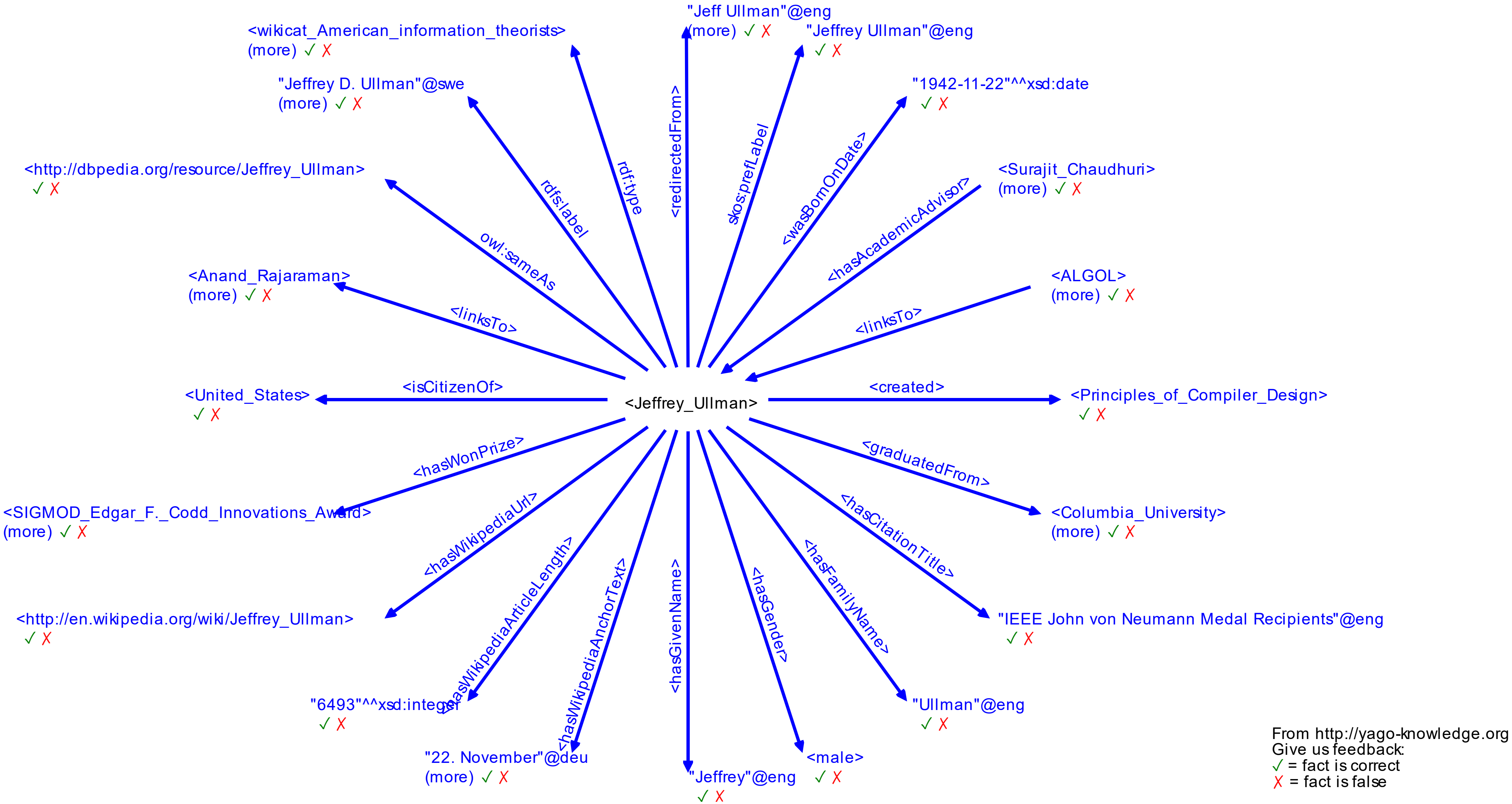
Distributed Knowledge Graph Querying on Edge, Fog and Cloud

DS 256 – Project Proposal

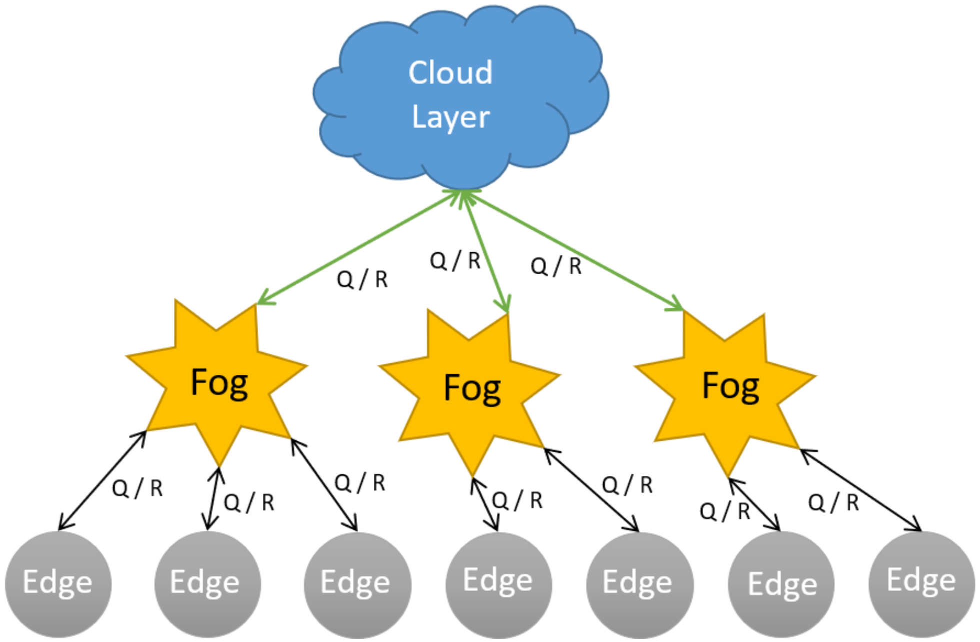
Shriram R

Knowledge Graph – YAGO [1]



Motivation

- Knowledge Graphs are huge!
- Graph mining problems are computationally hard
- IoT – Heterogenous Structure
  - Different Compute, Storage & Network Capacity
- Needs lightweight and low latency solution



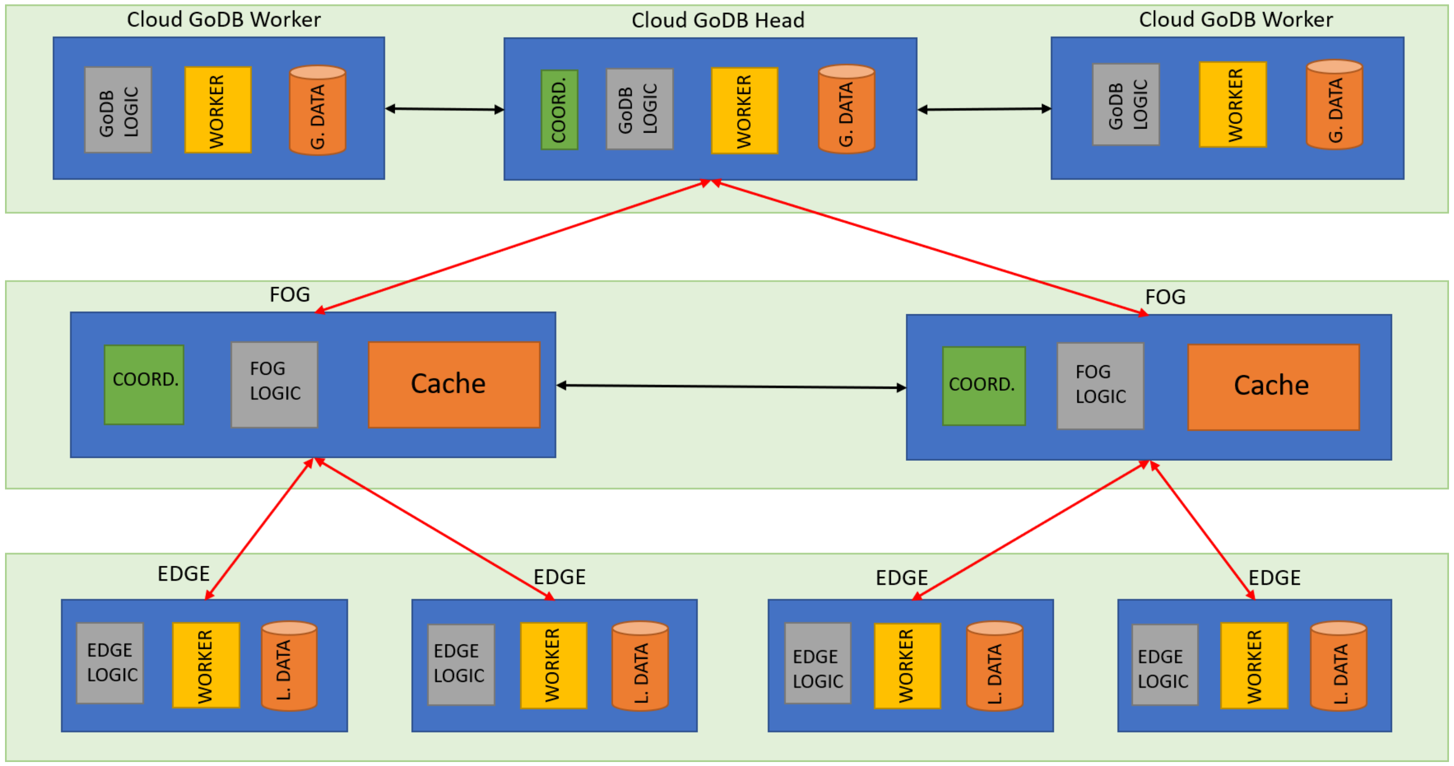
Key Contributions

- Graph Partitioning
  - Partition graph across edge devices and cloud
- Query Partitioning
  - Partition query into local and global query
- Query Caching
  - Devise caching mechanism in fog devices for global queries
- Graph Indexing
  - Indexing for local graphs in edge devices for different query types

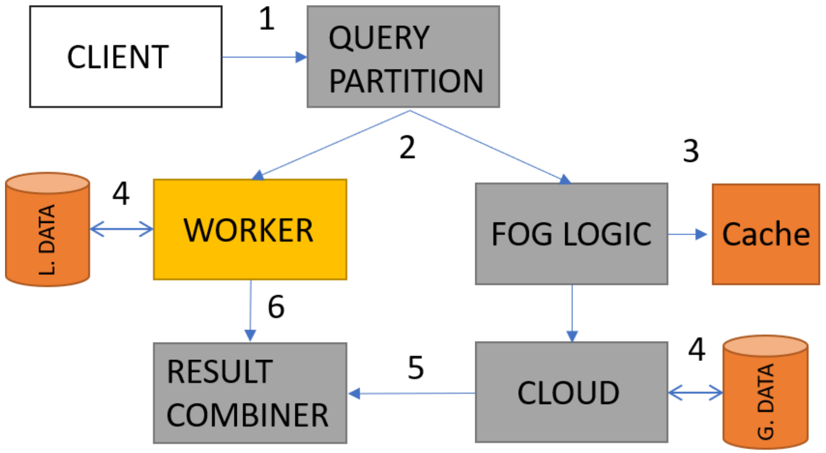
Related Work

- Graph Processing
  - Pregel [2]
  - Giraph [3]
  - GraphX [4]
  - Trinity [5]
  - GoDB [11]
  - Quegel [13]
- Graph Caching
  - GC [6]
- Graph Indexing
  - C-Tree [7]
  - Views [8]
  - FERRARI [9]
  - GraphS [10]

Problem Approach



Architecture Overview



Logical Flow

Experiment Proposal

- Microsoft Azure
- GoDB [11] on D4 VMs
- VIoLET [14] for IoT deployment
  - 3 Fogs
  - 20 Edges per Fog
- Dataset
  - YAGO [1]
  - NELL [15]
- Experiments
  - Query Partitioning
  - Query Caching
  - Graph Indexing
- Baseline – Cloud Only Graph Engine
- Local and Global Execution Times are Measured
- Random queries of all Types

Questions?

References

[1] F. M. Suchanek, G. Kasneci, and G. Weikum, “Yago: A core of semantic knowledge,” in Proceedings of the 16th International Conference on World Wide Web, ser. WWW ’07. New York, NY, USA: ACM, 2007, pp. 697–706. [Online].

[2] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, “Pregel: A system for large-scale graph processing,” in Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD ’10. New York, NY, USA: ACM, 2010, pp. 135–146.

[3] A. Ching, S. Edunov, M. Kabiljo, D. Logothetis, and S. Muthukrishnan, “One trillion edges: Graph processing at facebook-scale,” Proc. VLDB Endow., vol. 8, no. 12, pp. 1804–1815, Aug. 2015.

[4] R. S. Xin, J. E. Gonzalez, M. J. Franklin, and I. Stoica, “Graphx: A resilient distributed graph system on spark,” in First International Workshop on Graph Data Management Experiences and Systems, ser. GRADES ’13. New York, NY, USA: ACM, 2013, pp. 2:1–2:6.

[5] B. Shao, H. Wang, and Y. Li, “Trinity: A distributed graph engine on a memory cloud,” in Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD ’13. New York, NY, USA: ACM, 2013, pp. 505–516.

[6] J. Wang, N. Ntarmos, and P. Triantafillou, “Graphcache: A caching system for graph queries,” pp. 13–24, March 2017.

[7] H. He and A. K. Singh, “Closure-tree: An index structure for graph queries,” in 22nd International Conference on Data Engineering (ICDE’06), April 2006, pp. 38–38.

[8] W. Fan, X.Wang, and Y.Wu, “Answering graph pattern queries using views,” in 2014 IEEE 30th International Conference on Data Engineering, March 2014, pp. 184–195.

[9] S. Seufert, A. Anand, S. Bedathur, and G. Weikum, “Ferrari: Flexible and efficient reachability range assignment for graph indexing,” in 2013 IEEE 29th International Conference on Data Engineering (ICDE), April 2013, pp. 1009–1020.

[10] X. Qiu, W. Cen, Z. Qian, Y. Peng, Y. Zhang, X. Lin, and J. Zhou, “Real-time constrained cycle detection in large dynamic graphs,” Proc. VLDB Endow., vol. 11, no. 12, pp. 1876–1888, Aug. 2018.

[11] N. Jamadagni and Y. Simmhan, “Godb: From batch processing to distributed querying over property graphs,” in IEEE CCGRID, 2016.

[12] Y. Simmhan, A. Kumbhare, C. Wickramaarachchi, S. Nagarkar, S. Ravi, C. Raghavendra, and V. Prasanna, “Goffish: A subgraph centric framework for large-scale graph analytics,” in Euro-Par 2014 Parallel Processing, F. Silva, I. Dutra, and V. Santos Costa, Eds. Cham: Springer International Publishing, 2014, pp. 451–462.

[13] D. Yan, J. Cheng, M. T. Özsu, F. Yang, Y. Lu, J. C. S. Lui, Q. Zhang, and W. Ng, “A general-purpose query-centric framework for querying big graphs,” Proc. VLDB Endow., vol. 9, no. 7, pp. 564–575, Mar. 2016.

[14] S. Badiger, S. Baheti, and Y. Simmhan, “Violet: A largescale virtual environment for internet of things,” CoRR, vol. abs/1806.06032, 2018.

[15] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, and J. Welling, “Never-ending learning,” Commun. ACM, vol. 61, no. 5, pp. 103–115, Apr. 2018.