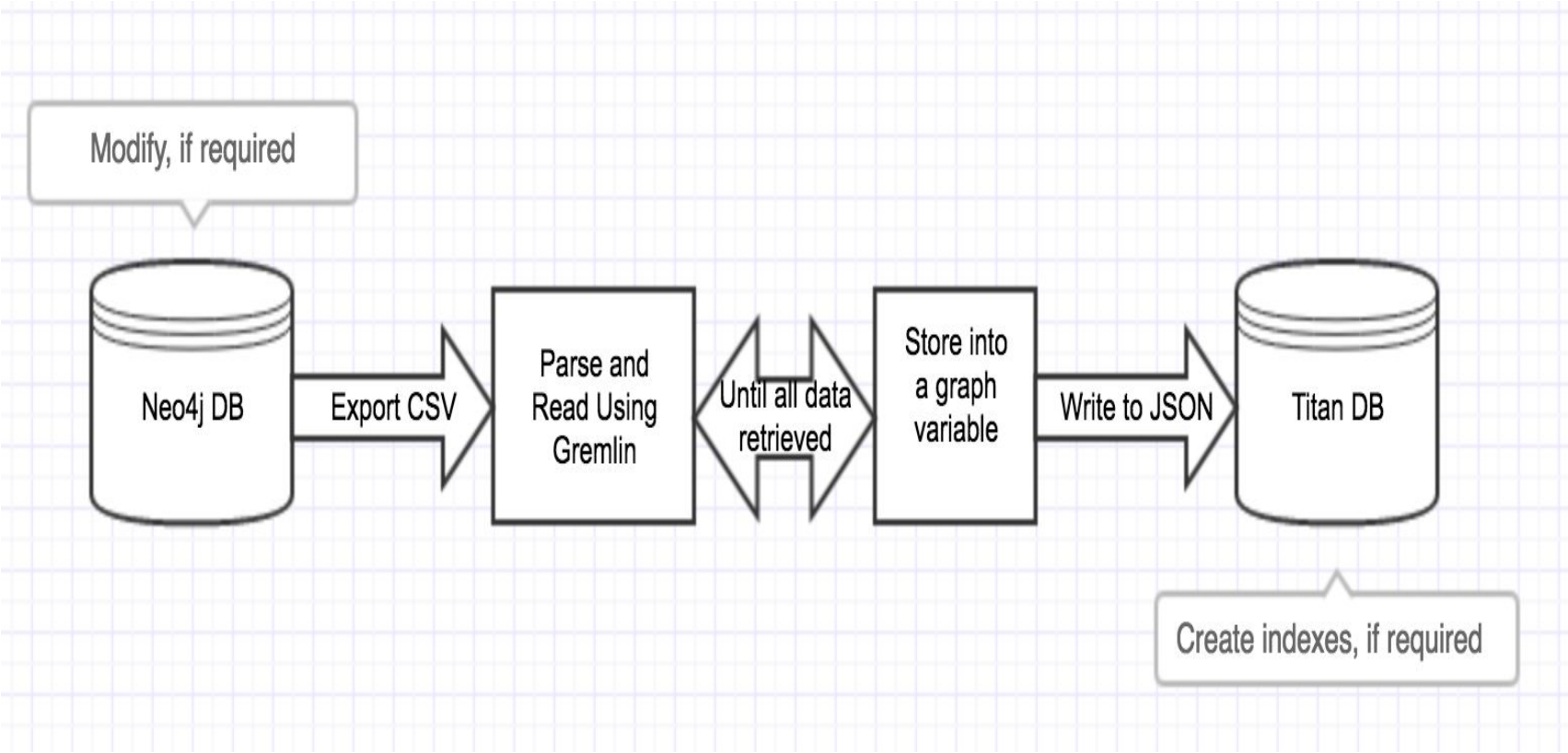
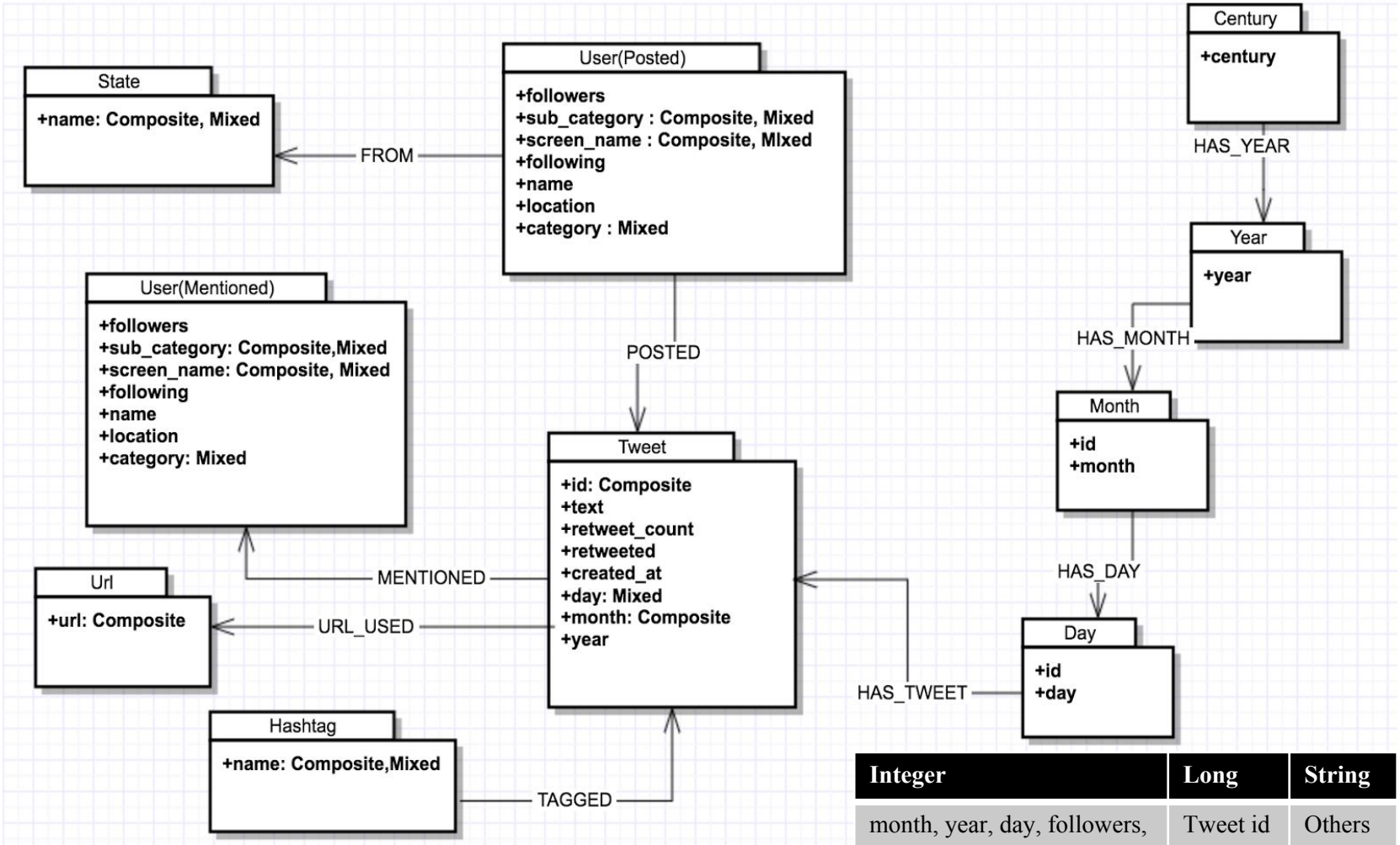
Graph Data Modeling for Political Communication on Twitter

Greeshma Reddy Padiri
Shruti Biswal
Boudhayan Banerjee
Department of Computer Science
Iowa State University

Loading Data into Titan

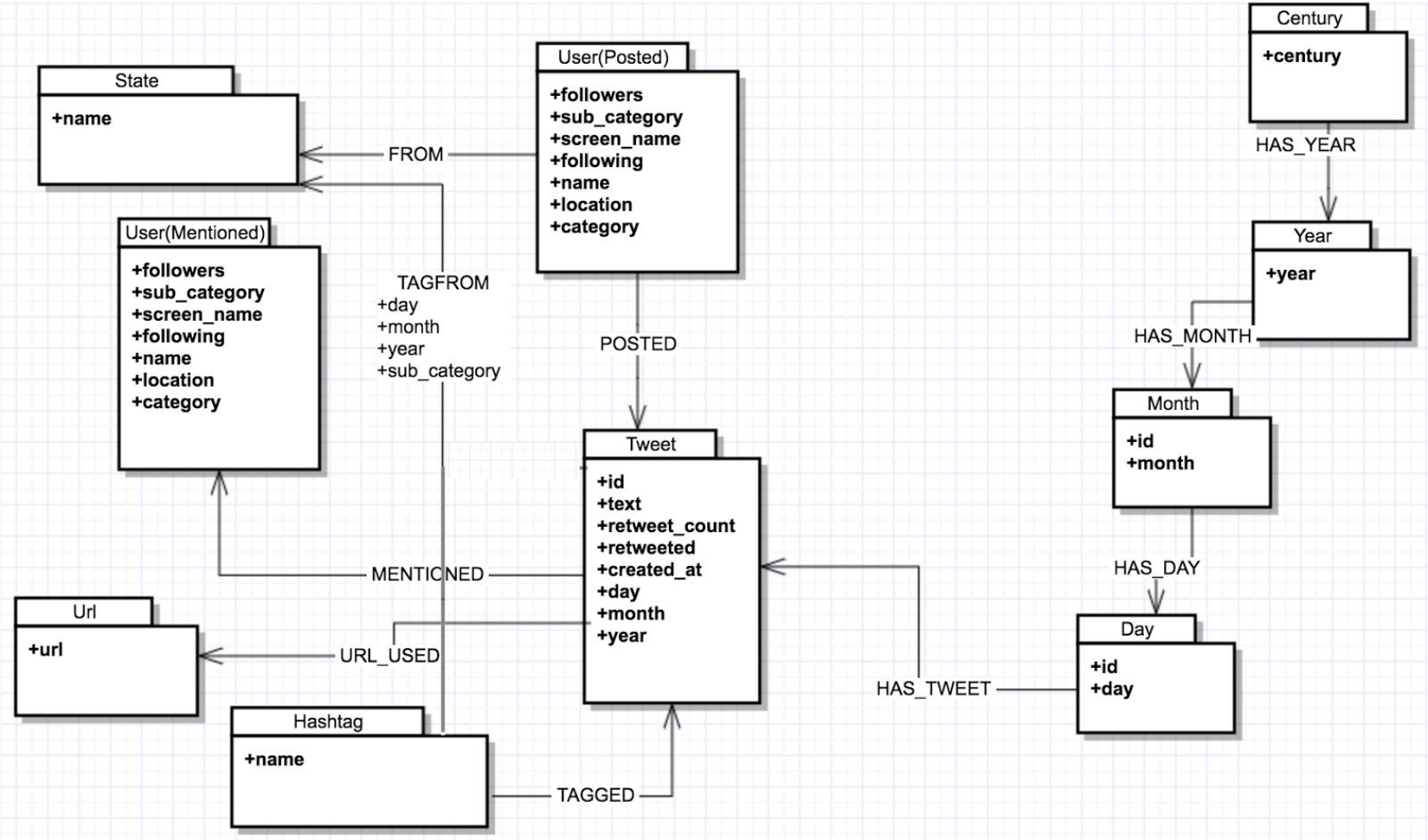


Indexed Model

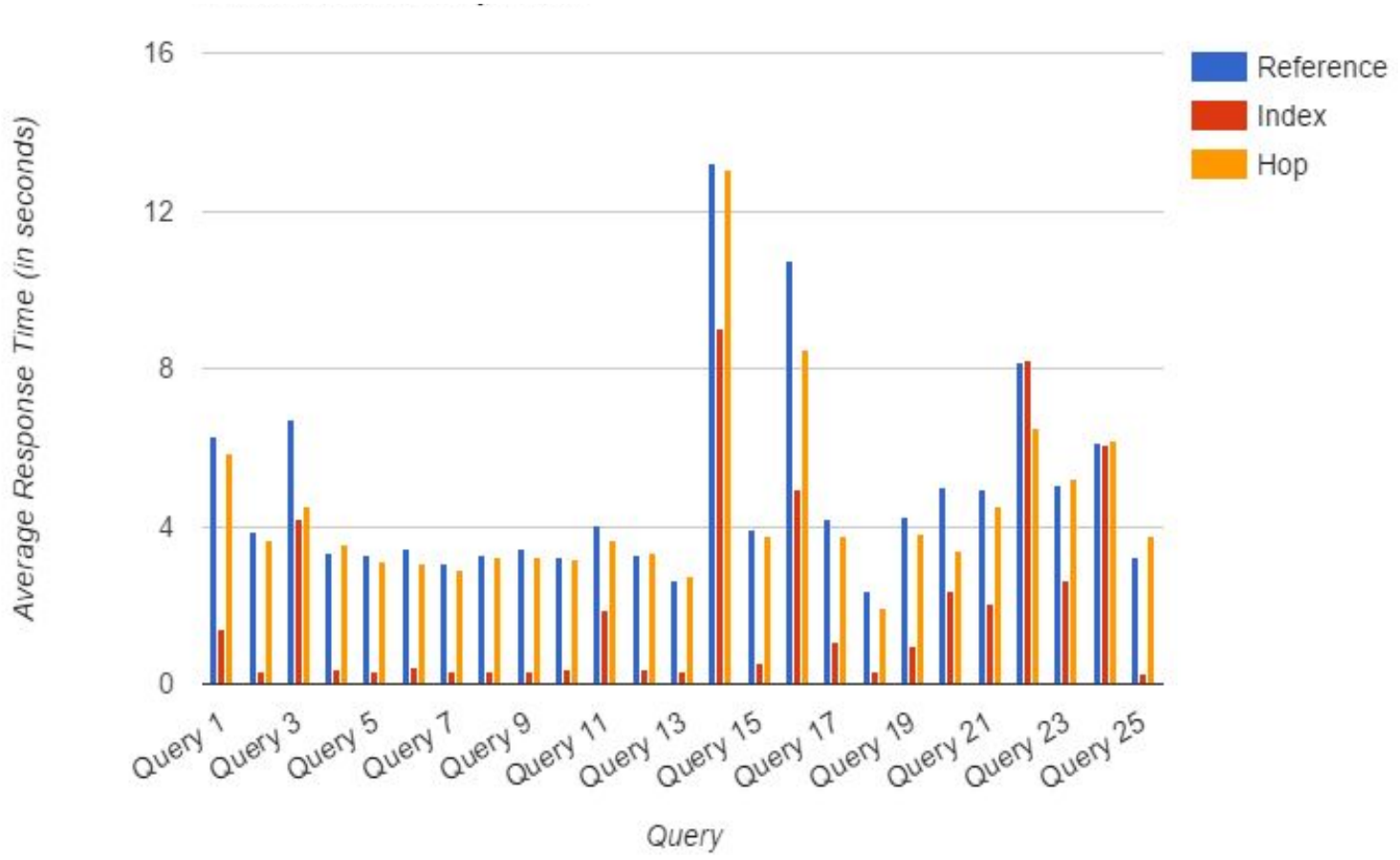


Integer	Long	String
month, year, day, followers, following, retweeted_count	Tweet id	Others

Reduced Hop Model



Performance Comparison



% Change in the Query Response Times for Best Model w.r.t. Reference Model

Query	% Change
Q1	-77.66
Q2	-91.52
Q3	-37.33
Q4	-88.94
Q5	-90.57
Q6	-87.92
Q7	-89.91
Q8	-89.58
Q9	-90.8
Q10	-88.71
Q11	-52.87
Q12	-88.29

Query	% Change
Q13	-87.19
Q14	-31.67
Q15	-85.59
Q16	-54.11
Q17	-75.06
Q18	-85.8
Q19	-77.75
Q20	-53.2
Q21	-58.52
Q22	0.53
Q23	-48.18
Q24	-1.38
Q25	-91.21

Important Findings

- Clock(int) {Query.iterate()} vs. System.currentTimemillis()
- Loading data was little time consuming but reliable and does not cause any system crashes.
- Reduced hop model with indexes performs far better than the proposed models but it is out of scope of the project!!

References

- Arijit Khan *et al.*, “Neighborhood Based Fast Graph Search in Large Networks” in ACM SIGMOD, Greece, pp. 901-912, June 12-16, 2011
- Konstantinos Xirogiannopoulos *et al.*, “GraphGen: Exploring Interesting Graphs in Relational Data” in VLDB, Hawaii, Vol 8. pp. 2032-2035, August 31- September 4, 2015.
- Florian Holzschuher,Rene Peinl, “Performance of graph query languages: Comparison of cypher, gremlin and native access in Neo4j”, in EDBT, Italy, pp. 195-204, March 18-22, 2013