

Simulation Framework for Distributed Database Query Processing in the Semantic Internet of Things

Johann Mantler

Master's Thesis
Institute of Information Systems
University of Lübeck

supervised by
Prof. Dr. Sven Groppe,
Benjamin Warnke

12. August 2021

Technology Trends

- IoT is growing
- IoT is driving Big Data
- IoT is driving Semantic Web
- computing capabilities are increasing
 - shifting computing and data storage to the edge
 - Cloud Computing → Fog Computing → Edge Computing

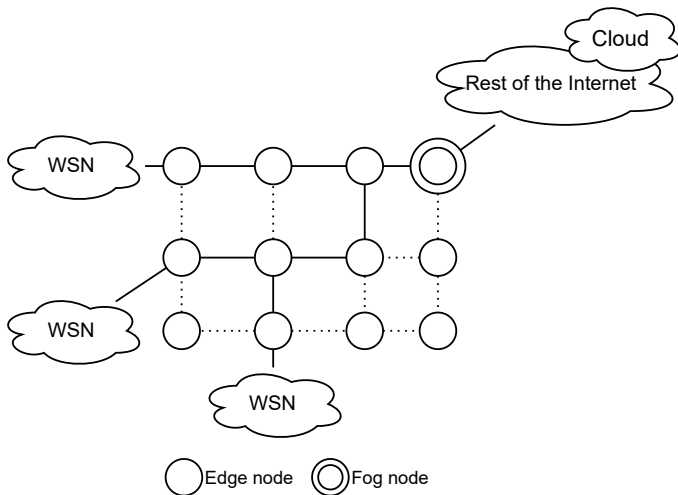
DBMS in Edge Computing

- luposdate3000: Semantic Web DBMS adapted to IoT
- instances run on edge and fog devices
 - multiplatform to support heterogeneity
 - distribution of data storage and query processing
- current work
 - simulator for integrating real DBMS into a modeled IoT environment
 - protocol for distributed query processing

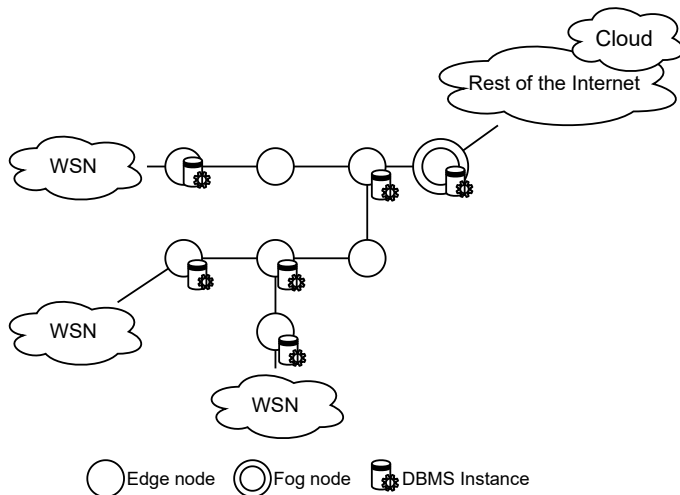
Related Work

Simulator	Features									
	Language	Purpose	Network Comm.	Network Protocols	IoT Routing	Node Performance	Real Data	Energy	External Apps	
ns-3	C++, Python	precise network simulation	✓ ¹	✓ ¹	✓ ¹		✓ ¹	✓ ¹	via file descriptor	
COOJA	Java, C	precise WSN simulation in Contiki	✓	✓	✓	✓	✓ ¹	✓		
FogNetSim++	C++	apps in fog environments	✓	✓		✓		✓		
CloudSim	Java	cloud computing	✓			✓		✓		
IoTSim-Edge	Java	app composition in edge computing	✓	✓		✓		✓		
EdgeCloudSim	Java	edge computing	✓			✓		✓		
IoTSim-Osmosis	Java	app composition	✓	✓		✓		✓		
iFogSim	Java	fog computing with data streams	✓			✓		✓		
PureEdgeSim	Java	edge computing	✓			✓		✓		
YAFS	Python	dynamic infrastructures	✓			✓		✓		
MyiFogSim	Java	mobile nodes	✓			✓		✓		
FogBed	Python	real apps in containers	✓			✓	✓		via Docker interface	
(Proposed)	Kotlin	query processing with real DBMS	✓	✓	✓	✓	✓			

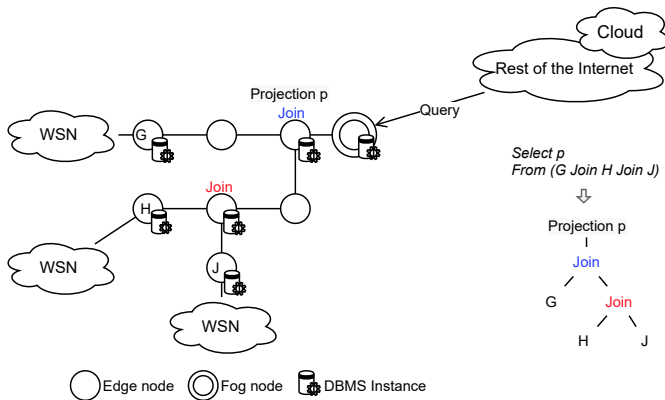
IoT Environment



DBMS Instance Distribution



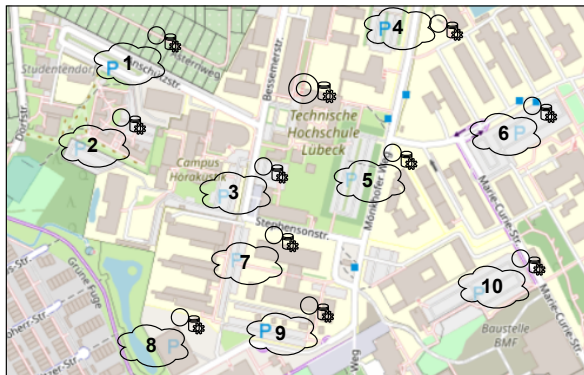
Operator Graph Mapping





Query Processing Protocol


```
function RECEIVE(pck)
  if pck is QueryPackage then
    parse, send MulticastPackage
  else if pck is MulticastPackage then
    if isLeafNode then
      calculate, send ResultPackage or EndResultPackage
    else
      prepare, send MulticastPackage
    end if
  else if pck is ResultPackage then
    calculate, send ResultPackage or EndResultPackage
  end if
end function
```


Model: Parking Space Finder Application



 WSN of parking sensors that cover parking area n

 Edge-Node and gateway of the WSN with DBMS


 Fog-Node and gateway to the Internet with DBMS


Protocol Evaluation

objective: less resource consumption than centralized processing
→ compare


- construct two topologies
 - distributed case: 10 instances as data sinks for the WSNs + 1 instance as fog node
 - centralized case: 1 instance as fog node and data sink for the WSNs
- simulate
 - insert sensor samples
 - process 8 different SPARQL queries


Visualization


 WSN of parking sensors that cover parking area n

 Root device for the DODAG and operator graphs

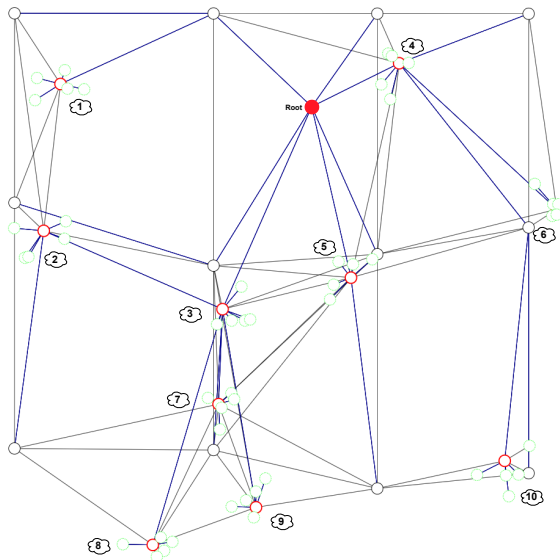
 Device

 Device with parking sensor

 Device with DBMS instance

 Available link

 RPL DODAG link



Simulation Results

	Centralized Case		Distributed Case	
	Sent Packages	Kilobytes Traffic	Sent Packages	Kilobytes Traffic
Only Sample Inserts	500	724	29060	9330
Q1 Listing A.2	500	724	29080	9451
Q2 Listing A.3	500	724	29080	9384
Q3 Listing A.4	500	724	29132	9459
Q4 Listing A.5	500	724	29080	9431
Q5 Listing A.6	500	724	29118	9413
Q6 Listing A.7	500	724	29216	9737
Q7 Listing A.8	500	724	29216	9761
Q8 Listing 3.5	500	724	29216	9726

Result Analysis

- distributed case is worse due to the current sensor data distribution
 - inserting a sample causes up to 66 messages
- without sensor data distribution: could perform better than centralized case

	Centralized Case		Distributed Case (Dummy)	
	Sent Packages	Kilobytes Traffic	Sent Packages	Kilobytes Traffic
Only Sample Inserts	500	724	500	258

- distributed query processing can have a maximum of $724 - 258 = 466$ kilobytes of traffic
- query processing traffic is always less than 466 kilobytes!

Summary and Future Work

- summary
 - new test environment for the development of IoT DBMS
 - new protocol design for distributed query processing
- future work
 - optimize sensor data distribution
 - optimize package sizes
 - compare execution times