Department of Computing and Software
McMaster University, Hamilton, ON, Canada

A Comparison of Time Series Databases for Storing Water Quality Data

Muntazir Fadhel, Emil Sekerinski, Shucai Yao
fadhelm, emil, yaos4@mcmaster.ca

February 12, 2019

Background
Water quality

- Water quality is a worldwide concern

Background
Water quality

- Water quality is a worldwide concern
- According to an analysis by Indigenous communities in 2017, the number of people in Canada who were affected because of drinking water quality was: **72,000** [MH]

Introduction
Smart Sensors

- Smart sensors change the way we see and control our environment [HS14, NS15]
  - Air or water quality
  - Soil conditions
  - Movements of wildlife

Introduction
Smart Sensors

- Smart sensors change the way we see and control our environment [HS14, NS15]
  - Air or water quality
  - Soil conditions
  - Movements of wildlife
- We focus on environmental sensing, in particular of **water quality**:
  - Dissolved Oxygen, Turbidity, Temperature, Conductivity
  - Ammonium, Nitrate, and Chloride
  - Sources of potential contamination

Purpose

As part of a Global Water Futures (GWF),

- to **empower communities** to keep track of water quality themselves through wireless sensors.
- to enable this through **high school outreach**

Purpose

As part of a Global Water Futures (GWF),

- to **empower communities** to keep track of water quality themselves through wireless sensors.
- to enable this through **high school outreach**

- Advantages:
  - Economy
  - Self-motivation
  - Trust

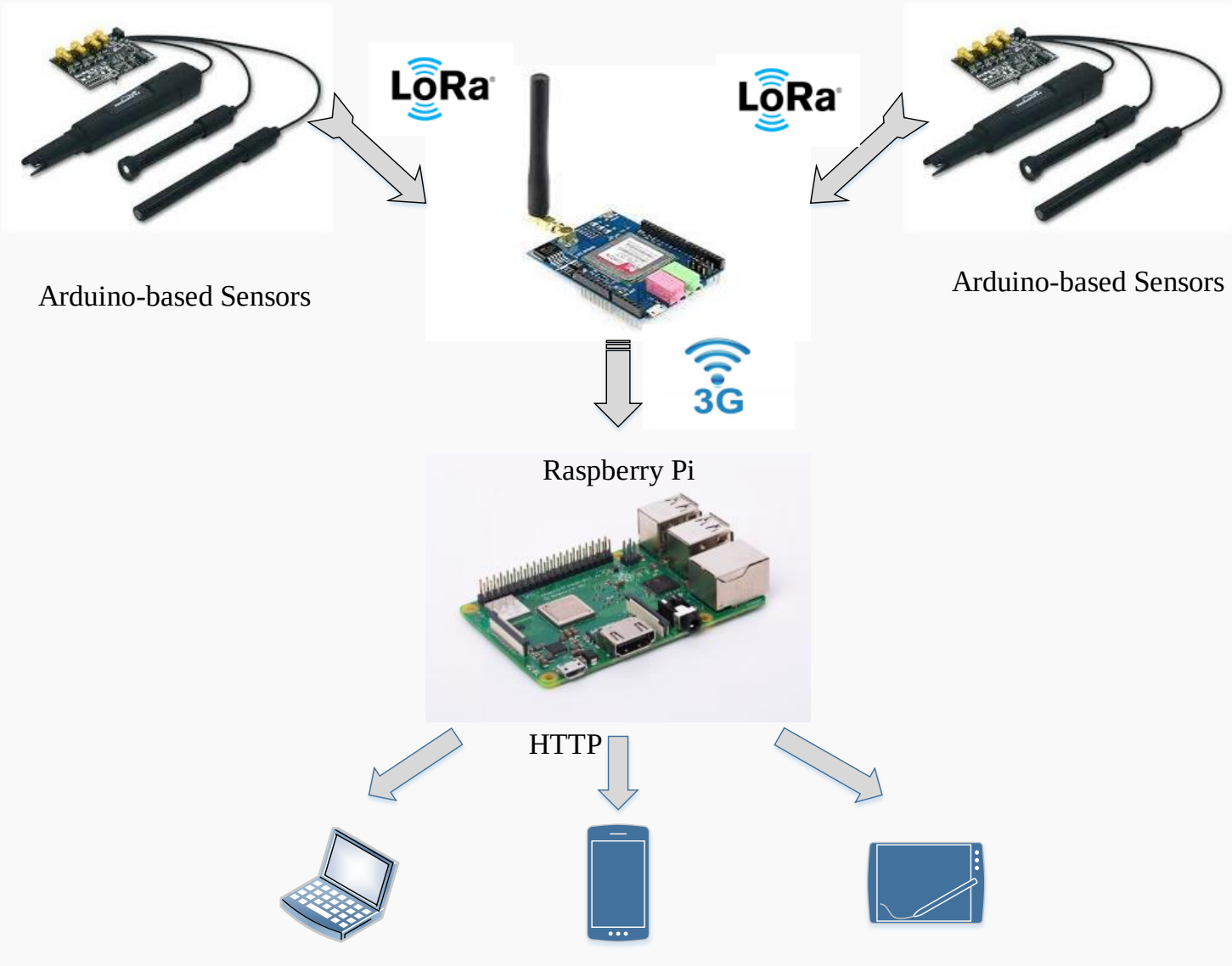
Purpose

As part of a Global Water Futures (GWF),

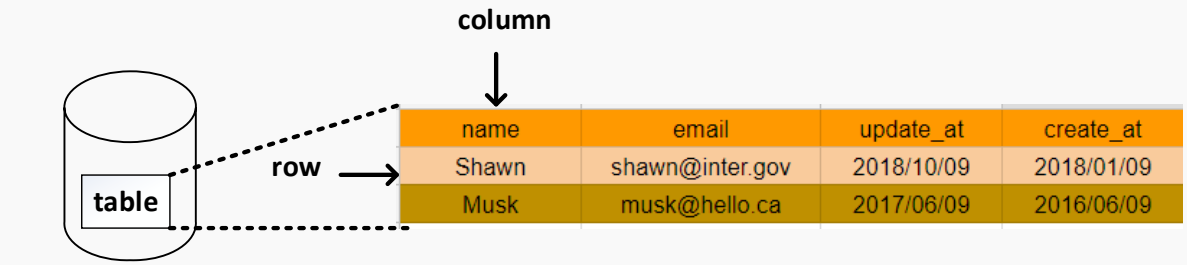
- to **empower communities** to keep track of water quality themselves through wireless sensors.
- to enable this through **high school outreach**

- Advantages:
  - Economy
  - Self-motivation
  - Trust
- Requirements:
  - **Low-cost low-power hardware**: Raspberry Pi, Arduino
  - **Open-source software**: web server, databases

Water Quality Monitoring System

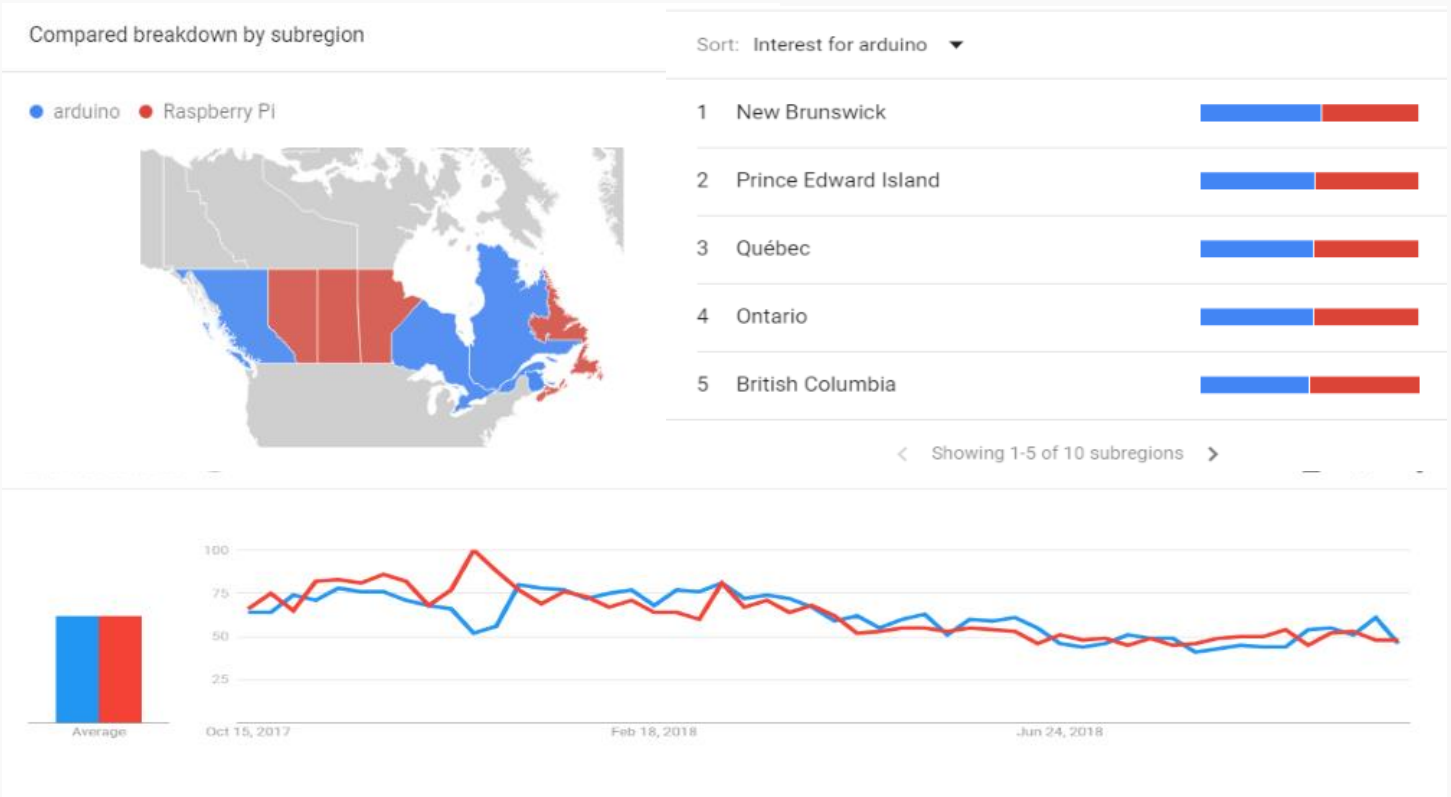


Databases



Relational Database

Time Series Database



Databases

- **Relational databases**
  - Assume that queries are frequent compared to insertions.
  - Queries can be composed.

Databases

- **Relational databases**
  - Assume that queries are frequent compared to insertions.
  - Queries can be composed.
- **Time series databases**
  - Assume that insertions are frequent compared to queries.
  - Queries do not target individual entries but summarize entries in a time interval.
  - Used for logging high-frequency events, such as requests to servers, stock trades.

Databases

- **Relational databases**
  - Assume that queries are frequent compared to insertions.
  - Queries can be composed.
- **Time series databases**
  - Assume that insertions are frequent compared to queries.
  - Queries do not target individual entries but summarize entries in a time interval.
  - Used for logging high-frequency events, such as requests to servers, stock trades.
- **Water quality data**
  - Slow rate of insertion allows the use of relational databases
  - Queries involve outlier detection and are more complex than supported by time series databases

Databases

- **Relational databases**
  - Assume that queries are frequent compared to insertions.
  - Queries can be composed.
- **Time series databases**
  - Assume that insertions are frequent compared to queries.
  - Queries do not target individual entries but summarize entries in a time interval.
  - Used for logging high-frequency events, such as requests to servers, stock trades.
- **Water quality data**
  - Slow rate of insertion allows the use of relational databases
  - Queries involve outlier detection and are more complex than supported by time series databases

Goal

To evaluate various time series databases in their ability to be incorporated into a low-cost water quality monitoring system.

Selected Databases

Database	Version	License	Compatibility	Protocols	Default Port
OpenTSDB	2.3.1	LGPL	Yes	HTTP	80
InfluxDB	0.10.0	MIT	Yes	HTTP	8086
Elasticsearch	6.4.0	Apache	No	HTTP	9200
Kdb+	3.5	Proprietary	Yes	HTTP	5001
RRDtool	1.5.5	GNU GPL.	Yes	TCP	13900
Graphite	1.1.2	Apache	Yes	TCP/UDP	2003
Prometheus	2.3.2	Apache	Yes	HTTP	80
DalmatinerDB	0.3.0	MIT	No	TCP	2003
SQLite3	3.24.0	Public domain	Yes	-	-

- Top 8 Time Series Databases from DB-Engines
- SQLite3

Comparison

- License: Kdb+

Comparison

- License: Kdb+
- Compatibility: Elasticsearch, DalmatinerDB

Comparison

- License: Kdb+
- Compatibility: Elasticsearch, DalmatinerDB
- Circular buffer-based: RRDtool, Graphite

Comparison

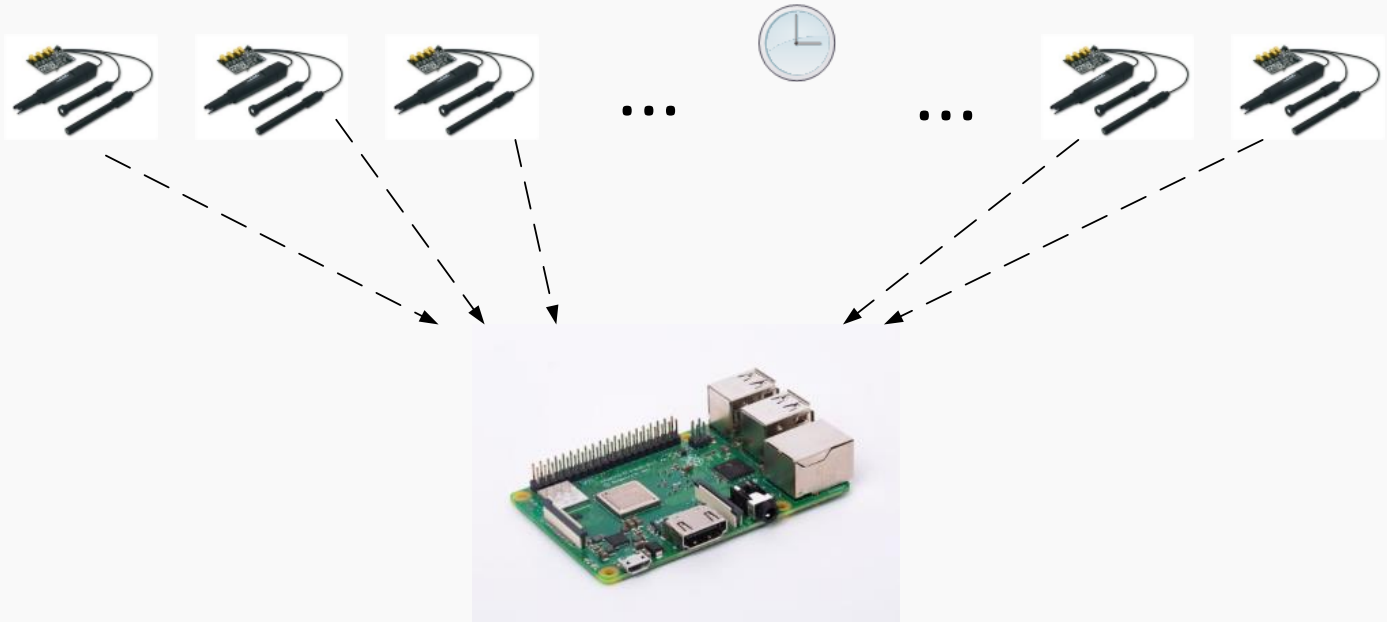
- License: Kdb+
- Compatibility: Elasticsearch, DalmatinerDB
- Circular buffer-based: RRDtool, Graphite
- Timestamp: Prometheus

Comparison

- License: Kdb+
- Compatibility: Elasticsearch, DalmatinerDB
- Circular buffer-based: RRDtool, Graphite
- Timestamp: Prometheus
- RAM: OpenTSDB

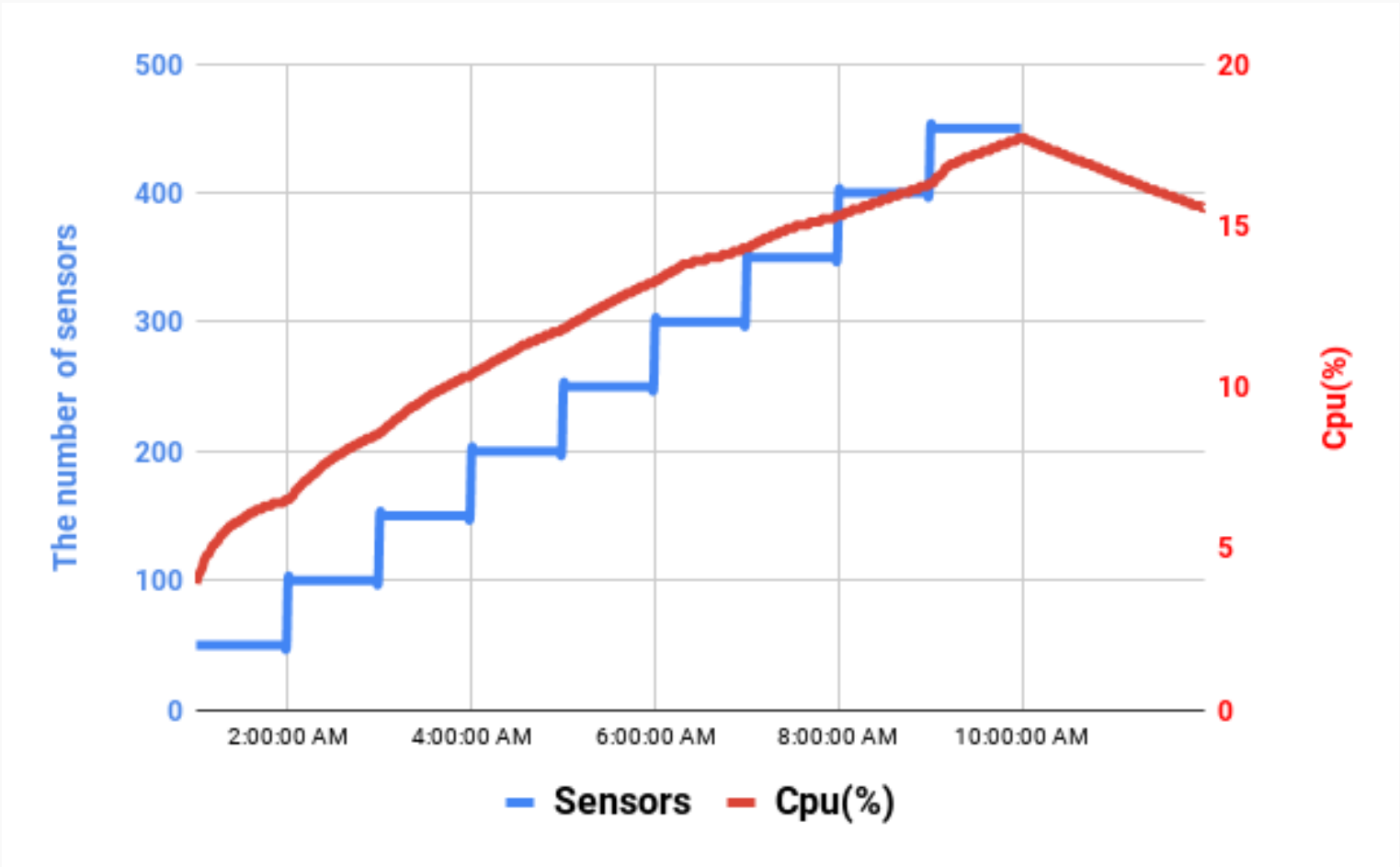
Data Points	OpenTDBS	InfluxDB(WAL/TSM)	Elasticsearch	SQLite3	RRDTool	Graphite
1	132 KB	12 KB/20 KB	6.9 KB	4.0 KB	4.0 KB	4.0 KB
1,000	132 KB	12 KB/20 KB	100.2 KB	28 KB	12 KB	12 KB
10,000	132 KB	12 KB/68 KB	866.7 KB	256 KB	80 KB	120 KB
100,000	132 KB	12 KB/564 KB	8.4 MB	2.7 MB	784 KB	1.2 MB
1,000,000	43 MB	12 KB/5.4 MB	80.2 MB	28 MB	7.7 MB	12 MB
10,000,000	-	-/-	802.1 MB	294 MB	77 MB	115 MB
100,000,000	-	-/-	8.3 GB	3.1 GB	763 MB	1.2 GB

InfluxDB on Raspberry Pi

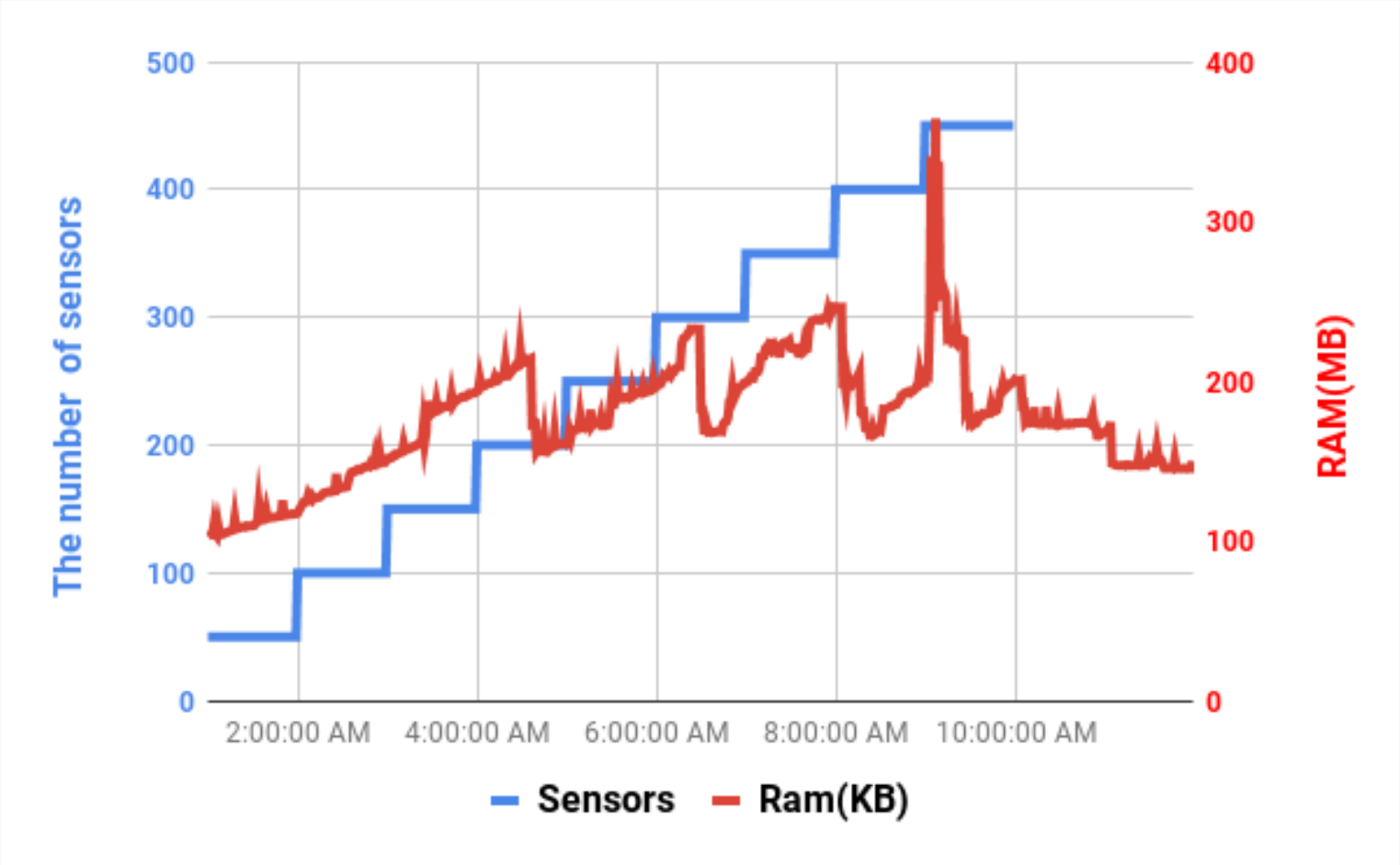


Time (UTC)	Reading
11/1/2017 2:20:00 PM	29.20630074
11/1/2017 2:21:00 PM	29.36000061
11/1/2017 2:22:00 PM	29.36050034
11/1/2017 2:23:00 PM	29.50049973
11/1/2017 2:24:00 PM	29.65399933

InfluxDB on Raspberry Pi

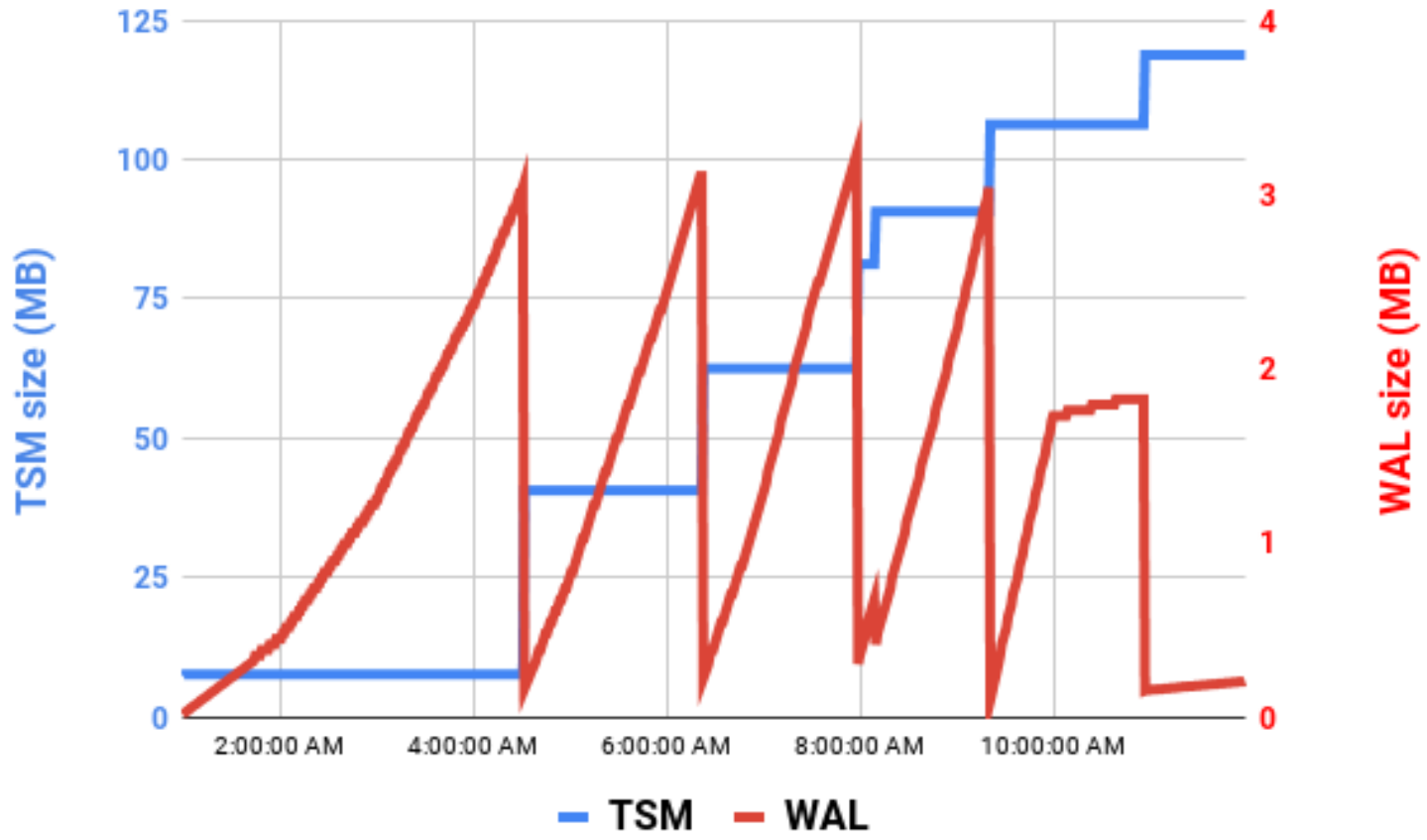


InfluxDB on Raspberry Pi



InfluxDB on Raspberry Pi

- Time-Structured Merge Tree (TSM)
- Write Ahead Log (WAL)



Conclusions

- An adequate architecture
- A carefully selected TSDB
- A low-cost low-power server
- Can handle water quality monitoring installations of substantial size: 450 sensors

References I

[HS14] Leo Hsu and P Ravi Selvaganapathy, *Stable and reusable electrochemical sensor for continuous monitoring of phosphate in water*, SENSORS, 2014 IEEE, IEEE, 2014, pp. 1423–1426.

[MH] Dawn Martin-Hill, *Co-creating of indigenous water quality tools*.

[NS15] Bojan Nokovic and Emil Sekerinski, *Automatically quantitative analysis and code generator for sensor systems: The example of great lakes water quality monitoring*, International Internet of Things Summit, Springer, 2015, pp. 313–319.

Questions

Questions?