A Scalable Database for Sensor Observations

Peter K. Kaiser[a], John N. Doe[b], Jutta Kaisaniemi[c]

[a]*Environmental Informatics Research Group, University of Zurich*
[b]*Biogeochemistry Research Group, Australian National University*
[c]*College of Engineering and Computer Science, University of Eastern Finland*

Abstract

As our main contribution, we present a scalable database capable of consuming sensor observations represented in RDF and evaluating queries for sensor observations formulated in SPARQL. We motivate the need for scalable databases for sensor observations on a case study in micrometeorology.

Keywords: Sensor Data, Data Management, Query Performance

---

1. Introduction

NoSQL systems have been utilized to manage sensor observations, specifically. Wang and Zhang (2001) present a Hadoop-based system designed to manage sensor observations. Wang and Zhang evaluate the performance of various queries.

Of particular interest here is CirrusRDF (Kaiser and Harth, 2007). It has been widely adopted in the literature (Lefort and Bobruk, 2010; Danh and Manfred, 2016; Cabral et al., 2013). The authors note that "a complete 'semantification' [...] of all data [...] seemed not feasible and promising to us, especially regarding the measurement data."

As we discuss in more details in Section 2, such systems generate large volumes of data, currently stored as files.

2. Case Study

We evaluate comparative database performance with data of a typical Sensor System for the direct measurement of CO_2, CH_4, and H_2O fluxes.

Email addresses: `p.kaiser@example.org` (Peter K. Kaiser), `j.doe@example.org` (John N. Doe), `j.kaisaniemi@example.org` (Jutta Kaisaniemi)

Table 1: A nice caption for the table.

Subset	Observations	Triples	Distinct
30 m	54 000	810 000	648 007
1 h	108 000	1 620 000	1 296 007

Large data volumes for surface-atmosphere fluxes of energy and trace gases are managed by platforms such as SOCI Portal.[1] The devices operate at 10 Hz sampling frequency and the data files include 30 min of measurement. The total number of archive files is 1 604 500. Each data file consists of a $18\,000 \times 40$ matrix. Of this matrix, we concentrate on the three columns for measured CO_2 [µmol mol^{-1}], H_2O [mmol mol^{-1}], and CH_4 [µmol mol^{-1}].

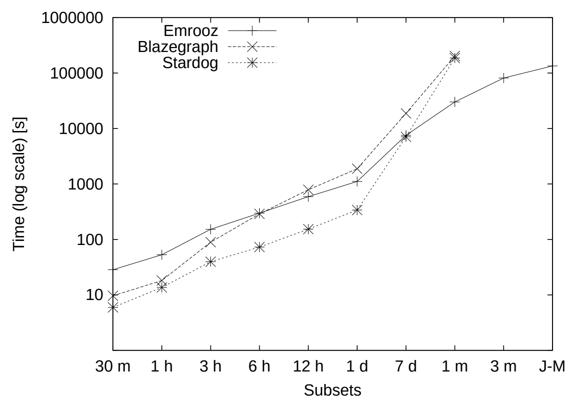

Figure 1: A nice caption for the figure.

3. Results

Table 1 summarizes subset sizes in terms of number of sensor observations, corresponding triples, and distinct triples. Figure 1 summarizes the *load* performance for the 10 subsets compared to Stardog and Blazegraph. Figure 2 summarizes the *query* performance for the 10 subsets compared to Stardog and Blazegraph.

The database is capable of evaluating SPARQL queries with a basic graph pattern with `FILTER` and `ORDER BY`. The query performance is determined by the following complex mathematical expression

[1]The Portal is online at `https://www.socip.eu/` (Accessed: May 8, 2017)

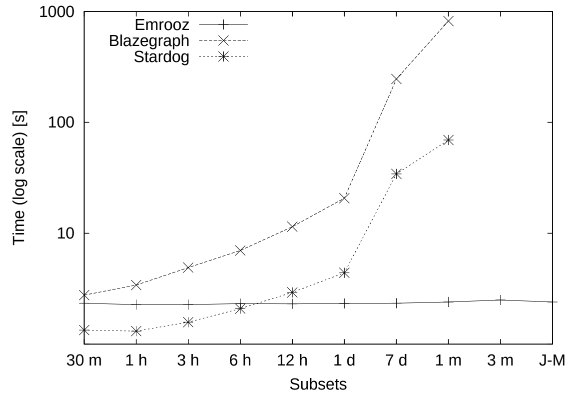

Figure 2: Another nice caption for this second figure.

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

4. Conclusion

We have presented a scalable database for sensor observations. That's it, folks! Thanks for reading.

Acknowledgements

This research is funded by the Academy of Germany project "ROTAR: High-quality Measurement Infrastructure" (Grant number 5489654).

Cabral, H., Morshed, S., Baku, Y., December 2013. From RESTful to SQL: A Case Study on Sensor Data. In: Proceedings of the 4th International Workshop on Networks. Vol. 4258. CEUR-WS, Perth, Australia, pp. 51–66.

Danh, Q., Manfred, X., January 2016. Sensor Middleware—Connecting the real world. In: Proceedings of the Web Challenge at the 5th International Web Conference. Berlin, Germany.

Kaiser, A., Harth, L., August 2007. CirrusRDF: Nested Key-Value Stores and Linked Data. In: Fokuoe, T., Liebig, Y. (Eds.), Proceedings of the 10th International Workshop on Scalable Knowledge Base Systems. Karlsruhe, Germany, pp. 30–42.

Lefort, B., Bobruk, L., August 2010. Sensor Data Cube for a 100 Year of Salinity. In: Henson, O., Corcho, K. (Eds.), Proceedings of the 3th International Workshop on Sensors. Vol. 404. CEUR-WS, Washington, DC, pp. 78–95.

Wang, X., Zhang, H., 2001. Semantic System Based on Cassandra. Great International Journal of Engineering 8 (10), 61–72.