A Scalable Database for Sensor Observations

Peter K. Kaiser[1] John N. Doe[2] Jutta Kaisaniemi[3]

[1]Environmental Informatics Research Group, University of Zurich

[2]Biogeochemistry Research Group, Australian National University

[3]College of Engineering and Computer Science, University of Eastern Finland

Introduction

NoSQL systems have been utilized to manage sensor observations, specifically. [1] present a Hadoop-based system designed to manage sensor observations.

Of particular interest here is CirrusRDF [2]. It has been widely adopted in the literature [3, 4, 5]. 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, such systems generate large volumes of data, currently stored as files.

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. Large data volumes for surface-atmosphere fluxes of energy and trace gases are managed by platforms such as SOCI Portal. The devices operate at $10\,\mathsf{Hz}$ sampling frequency and the data files include $30\,\mathsf{min}$ 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 [$\mathsf{\mu mol\,mol^{-1}}$], H_2O [$\mathsf{mmol\,mol^{-1}}$], and CH_4 [$\mathsf{\mu mol\,mol^{-1}}$].

Contact Information

- Web: http://www.university.edu/kaisaniemilab
- Email: j.kaisaniemi@example.org
- Phone: +358 (000) 111 1111

marum Center for Marine Environmental Sciences

Universität Bremen

Results

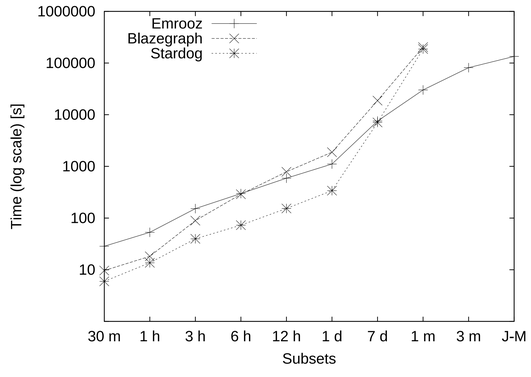

Figure 1: A nice caption for the figure.

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.

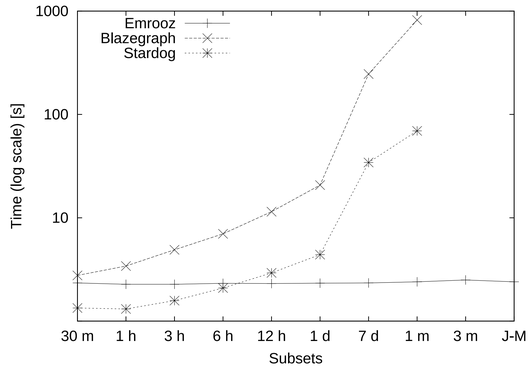

Figure 2: Another nice caption for this second figure.

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

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

Conclusion

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

References

[1] Xiaoming Wang and Hongbin Zhang. Semantic System Based on Cassandra. *Great International Journal of Engineering*, 8(10):61–72, 2001.

[2] Andreas Kaiser and Ludwig Harth. CirrusRDF: Nested Key-Value Stores and Linked Data. In Thorsten Fokuoe and Yuanbo Liebig, editors, *Proceedings of the 10th International Workshop on Scalable Knowledge Base Systems*, pages 30–42, Karlsruhe, Germany, August 2007.

[3] Bobruk Lefort and Laurent Bobruk. Sensor Data Cube for a 100 Year of Salinity. In Oscar Henson and Kerry Corcho, editors, *Proceedings of the 3th International Workshop on Sensors*, volume 404, pages 78–95, Washington, DC, August 2010. CEUR-WS.

[4] Quoc Danh and Xavier Manfred. Sensor Middleware—Connecting the real world. In *Proceedings of the Web Challenge at the 5th International Web Conference*, Berlin, Germany, January 2016.

[5] Heiko Cabral, Shu Morshed, and Yanfeng Baku. From RESTful to SQL: A Case Study on Sensor Data. In *Proceedings of the 4th International Workshop on Networks*, volume 4258, pages 51–66, Perth, Australia, December 2013. CEUR-WS.

Acknowledgements

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