

# Towards Streaming Consistency Management

Samuele Langhi

Lyon 1 University

Lyon, France

samuele.langhi@univ-lyon1.fr

Angela Bonifati

Lyon 1 University

Lyon, France

angela.bonifati@univ-lyon1.fr

Riccardo Tommasini

INSA Lyon

Lyon, France

riccardo.tommasini@insa-lyon.fr

## I. BACKGROUND & METHODS

Stream processing is designed to query unbounded and timely-ordered data flows in real-time while guaranteeing low latency and high throughput.

Despite its roots in the database community, the most recent research on data stream management has focused on performance and neglected the role of data quality. Moreover, quality management techniques are not always feasible in streaming, since the data should “keep moving” [1]. Nevertheless, data quality issues, e.g., timeliness [2] and completeness [3], become progressively more important as stream processing is applied beyond the context of analytics.

For this reason, we advocate for an approach that not just allows data quality intervention as done before [4], but also allows tracking the consistency of the records across streaming queries through provenance annotations [5]. These annotations are associated with the input and propagated to the query results, on which a degree of inconsistency is calculated.

The novelty of this method lies in not enforcing consistency [6], but making other operators aware of it through the annotations. Still, various challenges arise, including (i) defining streaming-compatible constraints, (ii) introducing annotations that preserve the semantics of classic streaming operators, and (iii) ensuring the accurate propagation of these annotations with the lowest overhead possible.

## II. RESULTS

Our extensive experiments on real-world streaming data underscore the effectiveness and efficiency of our approach in comparison to baselines that lack quality-related annotations or are adapted from static settings.

In particular, we demonstrate the efficiency of our methods concerning different windowing policies, with results showing our solutions as the best option in terms of overhead. Moreover, we demonstrate how our solution scales concerning the amount of inconsistency, showing a sub-linear decrease in performance with respect to the number of inconsistent records within the stream.

Our work introduces the first method for tracking inconsistencies over data streams and continuous query results. This method can propel further studies in identifying a class of streaming-compatible constraints. Furthermore, the influence of various windowing policies on consistency analysis remains an open direction for further studies, which may lay the ground for introducing a novel form of consistency-driven windowing.

## REFERENCES

- [1] M. Stonebraker, U. Çetintemel, and S. B. Zdonik, “The 8 requirements of real-time stream processing,” *SIGMOD Rec.*, vol. 34, no. 4, pp. 42–47, 2005. [Online]. Available: <https://doi.org/10.1145/1107499.1107504>
- [2] J. Traub, P. M. Grulich, A. R. Cuellar, S. Breß, A. Katsifodimos, T. Rabl, and V. Markl, “Scotty: Efficient window aggregation for out-of-order stream processing,” in *34th IEEE International Conference on Data Engineering, ICDE 2018, Paris, France, April 16-19, 2018*. IEEE Computer Society, 2018, pp. 1300–1303. [Online]. Available: <https://doi.org/10.1109/ICDE.2018.00135>
- [3] D. Palyvos-Giannas, K. Tzompanaki, M. Papatriantafidou, and V. Gulisano, “Erebus: Explaining the outputs of data streaming queries,” *Proc. VLDB Endow.*, vol. 16, no. 2, pp. 230–242, 2022. [Online]. Available: <https://www.vldb.org/pvldb/vol16/p230-palyvos-giannas.pdf>
- [4] O. Issa, A. Bonifati, and F. Toumani, “Evaluating top-k queries with inconsistency degrees,” *Proc. VLDB Endow.*, vol. 13, no. 11, pp. 2146–2158, 2020. [Online]. Available: <http://www.vldb.org/pvldb/vol13/p2146-issa.pdf>
- [5] T. J. Green, G. Karvounarakis, and V. Tannen, “Provenance semirings,” in *Proceedings of the Twenty-Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 11-13, 2007, Beijing, China*, L. Libkin, Ed. ACM, 2007, pp. 31–40. [Online]. Available: <https://doi.org/10.1145/1265530.1265535>
- [6] S. Song, A. Zhang, J. Wang, and P. S. Yu, “SCREEN: stream data cleaning under speed constraints,” in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, T. K. Sellis, S. B. Davidson, and Z. G. Ives, Eds. ACM, 2015, pp. 827–841. [Online]. Available: <https://doi.org/10.1145/2723372.2723730>